"Near Data" and "Far Data" for Urban Sustainability: How Do Community Advocates Envision Data Intermediaries?

HAN QIAO, University of Toronto, Canada
SIYI WU, University of Toronto, Canada
CHRISTOPH BECKER, University of Toronto, Canada

In the densifying data ecosystem of today's cities, data intermediaries are crucial stakeholders in facilitating data access and use. Community advocates live in these sites of social injustices and opportunities for change. Highly experienced in working with data to enact change, they offer distinctive insights on data practices and tools. This paper examines the unique perspectives that community advocates offer on data intermediaries. Based on interviews with 17 advocates working with 23 grassroots and nonprofit organizations, we propose the quality of "near" and "far" to be seriously considered in data intermediaries' works and articulate advocates' vision of connecting "near data" and "far data." To pursue this vision, we identified three pathways for data intermediaries: align data exploration with ways of storytelling, communicate context and uncertainties, and decenter artifacts for relationship building. These pathways help data intermediaries to put data feminism into practice, surface design opportunities and tensions, and raise key questions for supporting the pursuit of the Right to the City.

CCS Concepts: • **Human-centered computing → Empirical studies in collaborative and social computing**.

Additional Key Words and Phrases: urban informatics, civic tech, data intermediary, advocacy, data feminism, design, urban sustainability

1 INTRODUCTION

Cities have long been sites of social injustice and opportunities for radical changes [35]. Many urban areas are experiencing two converging forces: rapid urbanization and datafication [23]. Private companies gather sales and cellular data to predict consumer behavior [45], governments establish open data programs to promote transparency and encourage civic engagement [30, 58, 67], and grassroot community efforts organize to collect and use data to address issues of shared concern [13, 33].

Urban data ecosystems encompass diverse stakeholders, data sources, various data practices, infrastructures, policies, and social and cultural factors that shape these data practices [61]. Among the various components forming this urban data ecosystem, **data intermediaries** – individuals or groups providing services and tools to support others to understand and use data – play a key role

Authors' addresses: Han Qiao, h.qiao@mail.utoronto.ca, University of Toronto, , Toronto, Ontario, Canada, ; Siyi Wu, reyna. wu@mail.utoronto.ca, University of Toronto, , Toronto, Ontario, Canada, ; Christoph Becker, christoph.becker@utoronto.ca, University of Toronto, , Toronto, Ontario, Canada,

[41]. Almost all stakeholders in this urban data ecosystem rely on some sort of data intermediary. Many also act as data intermediaries for others, either by sharing data in structured forms or by providing access through other means such as data visualizations, and sometimes both. As critical geographer Rob Kitchin [60] argues, urban data does not only describe urban systems, but also shapes how a city should be governed as well as what a city should become in the future. However, whose views are represented by these data, who is empowered by the datafication process, and who has a say? The future of cities is unevenly distributed [14, 46]. The work of data intermediaries is central to this process and should be critically examined.

Nonprofit and grassroots organizations are major forces of advocating for marginalized voices in urban areas and enacting change. Even though members of these groups have to navigate through disempowerment from datafication [14, 97], they overcome all kinds of hurdles so that they can continuously develop data practices that center their values and commitments. For instance, advocates focus on the affective and narrative nature of data through agonistic data practices [22], and they prioritize values of care through data tinkering [13, 93]. With these unique perspectives, community advocates are experts in working with and as data intermediaries to enact change.

As Darian et al. [24] argue, today's advocacy work is, in many regards, data work. D'Ignazio [27] describes community advocates as "real-world people who are using data science in the service of real-world struggles for justice". Community advocates then have much to offer those of us researching design methods and theories for data practices and exploration tools. They show us how data intermediaries can design data exploration tools that empower marginalized voices . Anchoring on the concept of "the Right to the City" [62], where every inhabitant has the right to participate and shape urban futures, our work is guided by the question **what are community advocates' perspectives, insights and visions for data intermediaries and data tools built by data intermediaries?**

In this paper, we examine the perspectives of community advocates working with grassroots and nonprofit organizations through semi-structured interviews that incorporate a technology probe. We interviewed 17 advocates working with 23 grassroots and nonprofit organizations focusing on different areas of urban sustainability in Toronto, Canada. During the interviews, we incorporated *Curbcut Toronto*, a tool for exploring and visualizing urban sustainability data in Toronto designed by a group of data intermediaries, as a technology probe. Through these conversations, we learned how advocates work with a wide spectrum of data, ranging from their own lived experience to large scale open data. We recognize the distance between advocates and the data they use is important to their work, and we propose the terms **"near data"** and **"far data"** to distinguish the two extremes to help people think through this important aspect of urban data seriously considered by community advocates. Near data refers to data that is close to the everyday lives of the community advocates, including self-collected data, lived experiences, and conversations. Far data refers to heavily processed data that are often passed through many hands from other locations in the urban data ecosystem, such as open data from government or academic institutions. Through understanding advocates' ways of using data, challenges of using data and their reflections around their experience interacting with *Curbcut Toronto*, we identified one overarching design vision for data intermediaries, **connecting near data and far data**, as well as three pathways to work towards this vision: align data exploration with ways of storytelling, communicate context and uncertainties, and decenter artifacts for relationship building. Based on the vision and the pathways, we examine how advocates' visions of data intermediaries reflect commitments to data feminism. Finally, we discuss tensions that arise when connecting near data and far data, data intermediaries' roles in pursuing "the Right to the City", and how the concept of near and far data should be emphasized as part of a dynamic yet important context for working with data. Through this study, we contribute to the growing literature on data practices in community advocacy work and develop

design considerations for HCI communities and civic tech groups who research with and work as data intermediaries in support for more democratic and inclusive urban data ecosystems and urban futures.

2 RELATED WORK

The foundation of our work builds on the concept of "the Right to the City," which emphasizes that the right to re-imagine future urban spaces should belong to its residents and highlights the importance of collective action and advocacy. It has found new relevance and interpretations as cities going through rapid datafication. We discuss how our work is grounded in previous work on understanding data practices in grassroots and nonprofit organizations and on exploring roles of data intermediaries in the larger data ecosystem. Our work aims to speak to both bodies of literature in the HCI community, connecting them to open up spaces for new discussions.

2.1 "The Right to the City" Today

The concept of "the Right to the City" (RTTC) emerged in the 1960s, when many parts of the world were going through intensified urbanization. In Lefebvre's famous essay "the Right to the City," he highlighted the collective struggle of urban inhabitants under accelerated urbanization pushed forward by industrial capitalism. He argued for a collective re-imagination of urban life that embodies democratization and recenters social relations [62]. For Lefebvre, urban space goes beyond the material space to include social relationships and lived experiences of everyday life. The right to such space goes beyond citizen rights defined by the state. It encompasses a demand for and commitment to change stemming out of collective struggles [50, 62, 81]. RTTC has a spatial highlight in cities and urban areas, sites that carry political, social and cultural significance – the city is a site for developing capitalism but also a site for the destruction of capitalism; it is a site of struggle and conflict but also the most common site where alternative visions for our societies have emerged.

As urban areas experience rapid datafication in recent years, many scholars have been expanding and examining the meaning of RTTC in newer contexts [7]. A growing body of literature examines how urban data and digital tools, like urban dashboards, prioritize reductionist and objectivist ways of knowing and governing cities that impact collective efforts towards RTTC [59, 69]. An inseparable topic is recognizing the socio-digital inequalities posed by the digitization and datafication process. Past research has examined the adverse impact of datafication in urban spaces, such as security of personal data, gaps in digital literacy among different social groups, as well as lack of equal opportunities and tools to influence important matters [51]. Socio-economic factors such as educational inequality, age, gender, ethnicity and residence all contribute to socio-digital inequalities brought by digitization and datafication in cities [101]. When the idea of data-driven urban governance becomes widespread, resource-constrained mission-driven organizations, such as nonprofits and social enterprises, are disempowered as they invest time and sacrifice expertise to respond to the pressure of becoming data-driven at the expense of their mission [14].

Beyond presenting a foundation for us to understand the struggles brought by datafication in urban areas, RTTC also continues to serve as a call to build solidarity in the face of difference and to imagine urban futures that center democracy, diversity, justice, and sustainability. Anastasiu [7] elaborates RTTC in the context of the smart or digital city to argue that this right goes beyond access to data or technology, but should include "the right to both produce, manage and own all of these as part of an act of political and economic empowerment that is geared primarily towards the collective benefit and the strengthening of social relations." Heitlinger et al. [52] further explored how design and co-design can participate in this call for RTTC through focusing on the commons, care and biocultural diversity. Emerging work around data justice sought to foreground

the political nature of datafication, challenge existing power-structure manifested in data and empower the impacted communities to frame problems and address issues related to datafication [26, 34, 48, 73, 75, 91]. Currie et al. [23], bridging the concept of data justice and RTTC in the age of 'Big Data', algorithms, artificial intelligence and other data-intensive technologies, highlighted the central role of people, lived experiences, and bottom-up efforts in leading radical changes.

Shaw and Graham [89] argue that data as representation of cities are "often as important as their bricks and mortar" (p908) and that those who have more control over data have more power over the city. Thus data intermediaries act as an important stakeholder in the urban data ecosystem affecting the right to access and use data, to form narratives through data, to argue for what are important urban issues, and to shape possible urban futures. HCI researchers are especially well-positioned to contribute to this intersection between design, computing, and social and political implications in the urban context.

2.2 Data intermediaries in HCI

Data intermediaries are individuals or groups providing services and tools for supporting others to understand and use data [41]. Given the limitations and challenges surrounding the access and effective use of data, particularly for marginalized communities, data intermediaries become a critical stakeholder in this space. As Shaharudin et al. [88] discussed, the data intermediary role can be played by a variety of actors such as civil society bodies, journalists, entrepreneurs and researchers [65, 85] making data more useful in various ways such as compiling data from various sources [88], breaking down technical barriers [32], curating datasets based on smaller geographies such as neighborhoods [102], or offering consultation and partnership services to empower local communities in data utilization [96]. Data intermediaries serve as vital facilitators, bridging the divide between data availability and its impactful utilization [32].

The HCI community has explored various topics relevant to the role of data intermediaries and designs of data intermediary tools. One key research area for data intermediaries is access and use of open data. Past researchers have examined what it means for open data to be accessible, developing evaluation metrics and identifying challenges and barriers for effective use [32, 37, 47]. They have also suggested the important role of data intermediaries in facilitating meaningful engagement with open data. Corbett et al. [20] studied open data programs in the city of Edmonton, Canada and highlighted that data intermediaries are key to the success of such efforts. Dove et al. [32], through interviewing data intermediaries, identified pathways to successfully nurturing a community of data intermediaries. In the context of community advocacy works, other researchers highlighted the pivotal role of data intermediaries as they connect various stakeholders and bridge gaps in existing data [82, 88].

The HCI community has also explored the design of data intermediary tools. For example, Data:In Place was an open-source platform co-designed with communities and charities to simplify access to open data for civic advocacy [82]; Governor was an open-source web app that streamlines searching and integrating data tables in Open Government Data Platforms, with provenance tracking validated through user studies [63]. Data intermediaries' work is never just restricted to research settings. Close to the location of our study, the City of Toronto has many data intermediary projects led by local residents: OpenWaterData gathers various sources of water quality information and presents it through a web application for informing local open water swimmers and paddle-boarders [3], BikeSpace is a crowdsourced data platform where cyclists anonymously report bicycle parking issues across Toronto [1], and No More Noise Toronto combines crowdsourced urban noise data with the city's open information to raise awareness of urban noise pollution and advocate for policy change [92].

The work of data intermediaries might also carry risks such as deteriorating data quality, reinforcing bias, oppressing marginalized voices, and thus leading to unethical or harmful uses of data. For instance, as data is processed by numerous layers of data intermediaries, White [99] illustrated how data can be used and interpreted in various ways not anticipated with respect to their original generation. Bates et al. [9] proposed the concept of data journey to emphasize the importance of recognizing the physical sites and socio-material conditions of data practices as layers of data intermediaries pass down the data. When analyzing data practices through the lens of data journeys, Jensen et al. [56] also revealed challenges, such as assessing data validity through standardized data algorithms versus competencies shared socially. While the work done by data intermediaries can bridge access and use of various data, they might not necessarily be inclusive of local community members [18].

Therefore data intermediaries play key roles in making data better or worse, and their work should be examined carefully. CSCW and HCI researchers exploring design methods and data practices have called for collective sense-making and co-design in data work [72, 87]. Community-oriented citizen science researchers also emphasized the importance of co-design with local community, highlighting crucial values in democratic participation, local knowledge, and equitable power relationships in enacting social change [17, 53, 54, 77, 80]. Therefore, community advocates' perspectives on data intermediaries hold much value as they represent and amplify lived experiences and marginalized voices.

2.3 Data Practices in Grassroots and Nonprofit Organizations

Many scholars have examined inequalities caused by datafication [51, 101] and challenges experienced by grassroots and nonprofit organizations [14]. A growing body of research in HCI also highlighted the creative and attentive ways of using data when these groups navigate through the constraints and tensions posed by datafication [13, 27, 93]. As D'Ignazio [27, p.10] wrote: "grassroots data activists at the margins – real-world people who are using data science in the service of real-world struggles for justice – have much more expertise to offer those of us trying to craft data frameworks and practices in the service of justice."

Grassroots and nonprofit organizations are leveraging data in diverse ways. These organizations use open data for storytelling to achieve goals such as grant applications, internal evaluations, and obtaining support from local government officials [33]. Community advocacy groups leverage data for legitimizing the organization's expertise and mission, mobilizing audiences, promoting innovative work, and amplifying the visibility of marginalized beneficiaries [24]. Under the overarching goal of legitimization, community organizers leverage data in diverse forms of legitimation towards different audiences, such as "legitimizing the need for change within the community," "legitimizing claims against the state," and "legitimizing lived experience" [79].

Grassroots and nonprofit organizations are exemplary role models for pushing against the norms of a data economy conventionally organized around science, surveillance and selling [28]. These organizations, on the other hand, organize around social good and goals of co-liberation that strive for the values, goals and voices of their beneficiaries [24, 28]. Their work is also usually an enactment of care, a focus beyond productivity to prioritization of relationships, social and cultural community context, and attentiveness to ways of nurturing and sustaining one another [13, 93].

Researchers also highlighted how grassroots and nonprofit organizations navigated the challenge of double bind, contradictory situations and tensions when using data [22, 36]. For example, feminist activists adopt a practice of "refusing while using data and using while refusing data" when documenting stories related to feminicides as they navigate tensions carried by quantitative data science [27]. Community organizers also navigated the tension between protecting their communities from surveillance through data and securing funds through reporting data to funders

[79]. Crooks and Currie [22] proposed the concept of agonistic data practices, which recognizes the limitations of using quantification and offers a liberating framework that centers affective and narrative ways of using data. Darian et al.[24] showed how advocates enact the seven principles of data feminism [28], listed in Table 1, which we will refer to in our analysis.

Data Feminism Principle	This principle means...
Examine Power (DF1)	Recognize forces of oppression shaping urban lives and reflected in urban data and data tools.
Challenge Power (DF2)	Push back against unequal power structures to work towards alternative urban futures.
Elevate Emotion and Embodiment (DF3)	Recognize that knowledge about urban spaces comes from inhabitants' lived and bodily experiences.
Rethink Binaries and Hierarchies (DF4)	Challenge systems of counting and classification often found in data describing cities.
Embrace Pluralism (DF5)	Value multiple perspectives and diverse ways of knowing a city and using urban data.
Consider Context (DF6)	Attend to the provenance and environment from which the data was collected and processed, recognizing that data are never objective representations of urban lives.
Make Labor Visible (DF7)	Value and make visible the work of those who collect, process, annotate, and curate urban data.

Table 1. Data feminism principles and their explanations in the context of urban data.

Ultimately, data advocacy work goes far beyond data. A successful campaign or mobilization requires social and political capital [22, 27]. As grassroots and nonprofit organizations fight in the frontline of social injustice and adopt powerful ways of leveraging data in challenging situations, they form a unique lens towards data for advocacy that has much to offer those of us working in the HCI research community, hoping to understand design and democratization of data support tools and practices. Thus, we talk to and learn from community advocates working with grassroots and nonprofit organizations to understand their insights on working with and as data intermediaries.

3 METHOD

3.1 Participants and Sampling Method

We interviewed 17 participants working with 23 grassroots and nonprofit organizations focusing on urban sustainability issues in Toronto neighborhoods. Their focal topics range from affordable housing, transit access, road safety, mental health, climate, park and food access etc. For instance, P1's focus on urban noise addresses an urban environmental health hazard that impacts residents' quality of life, and P5's work on youth engagement in urban planning promotes inclusivity and long-term community resilience, both essential to urban sustainability. We reference all participants by number P1 to P17 and document detailed information about their topics of work in Table 2.

Participants were initially recruited from the first author's existing network stemming out of participation in a local civic tech organization for two years, then through snowball sampling as well as cold emailing through organizations' websites. Contact information was collected through a scan of Toronto's urban sustainability focused community groups, including environment, housing,

transportation and health, which are not only key issues in urban sustainability, but also the four data topics included in the current version of *Curbcut Toronto*, the technology probe we included in the interviews. When recruiting participants, we specifically focused on recruiting participants who think data already influences their approach to community advocacy work.

Among the participants, eight identified as female, two identified as non-binary and seven identified as male. Ten of the 17 participants are involved in more than one community group focusing on different yet interrelated topics around urban sustainability. P7 explained that all the different kinds of community advocacy work *"synergize"* with each other and then elaborated that *"many advocates are wearing multiple hats with multiple lenses, but all working towards hopefully the same goal of creating safer spaces and safer means to get to those spaces and places throughout the city"* and that *"they all work together in a way that looks at it with an equitable lens."* To categorize the type of advocacy work they do, we adopted Frumkin's framework [38], which was also used in [24]: seven of the participants have a focus on civic and political engagement, five have a focus on service delivery and the other five cover both types of work. The participants self-identified the type of organization that they work with as either more informal community organizing, which we refer to as grassroots organizations, or nonprofit organizations. Previous works in HCI have discussed both types of organizations together [93] and focused on each type separately [33, 79]. In our project, both types of organizations share more similarities than differences when working with data for advocacy. Among our participants, all are users of data intermediaries' work and some act as data intermediaries. Initially, we recruited participants with a focus on community advocates as potential users of data tools developed by data intermediaries. However, through our interviews, we discovered that many participants not only work with data provided by data intermediaries but also contribute to the work of data intermediaries in various ways. For example, P3 uses data from intermediaries and combines it with their own data for storytelling through local magazines, while P1, in addition to using works done by data intermediaries, also acts as a data intermediary by collecting counter-data [27] and providing data visualizations to the broader public.

3.2 Data Collection

We collected data through a series of semi-structured interviews remotely through a video conferencing platform, Zoom. All interviews were conducted by the first author, recorded and transcribed by both the first and the second author for further analysis. The interviews lasted between 28 and 96 minutes, accumulating a total duration of 16 hours and 29 minutes, with a median length of 53 minutes and an average of 58 minutes. All interviews were conducted between April and November 2023. Two of the interviews were conducted in two sittings within a week due to participants' time availability and eagerness to chat more than the expected time.

The interviews were conducted based on a protocol that consisted of three main parts: 1) the community advocacy work that the participant is involved in 2) their experiences of using or not using data in their work 3) interacting with *Curbcut Toronto*, a data exploration and visualization tool, and sharing their thoughts and reflections around using data intermediary tools. During each interview, follow-up questions were asked based on participants' specific experiences.

For the third part, when *Curbcut Toronto* came in as a technology probe, participants were asked to share their screen through the video conferencing tool while interacting with the tool. Before sharing the tool to participants, the interviewer explained the think-aloud method [6, 42] and then asked the participants to freely explore the tool while thinking aloud. After the participants felt that they had fully explored the tool, the interviewer continued the semi-structured interview, asking follow up questions based on the interview protocol to prompt users to reflect on their experiences using the tool.

Participant	Advocacy topic	Type of advocacy work	Type of organization
P1	Urban noise, biking and bike safety	Civic & Political Engagement	Grassroots organization
P2	Road safety, park access	Civic & Political Engagement	Grassroots organization
P3	Mental health, affordable housing, public transit	Civic & Political Engagement, Service Delivery	Grassroots organization, Nonprofit organization
P4	Climate	Civic & Political Engagement, Service Delivery	Grassroots organization
P5	Road safety, youth engagement in urban planning	Civic & Political Engagement	Grassroots organization
P6	Public space activation	Service Delivery	Nonprofit organization
P7	Bike safety, library, public space activation	Civic & Political Engagement, Service Delivery	Grassroots organization, Nonprofit organization
P8	Public transit, accessibility	Civic & Political Engagement	Grassroots organization
P9	Park access, environment	Service Delivery	Nonprofit organization
P10	Youth engagement in urban planning	Service Delivery	Nonprofit organization
P11	Public health, civic engagement	Civic & Political Engagement	Grassroots organization
P12	Park access, food access	Civic & Political Engagement, Service Delivery	Grassroots organization, Nonprofit organization
P13	Employment opportunity access	Service Delivery	Nonprofit organization
P14	Shelter services	Service Delivery	Nonprofit organization
P15	Cultural resilience	Service Delivery	Nonprofit organization
P16	Climate	Civic & Political Engagement	Grassroots organization
P17	Climate, food sovereignty	Civic & Political Engagement, Service Delivery	Nonprofit organization

Table 2. Summary of Participants and Their Advocacy Work

3.3 *Curbcut Toronto* as a Data Intermediary Technology Probe

To better understand community advocates' insights and perspectives around data intermediaries, we incorporated *Curbcut Toronto* as a technology probe during our interviews to provoke thoughts and help participants articulate their insights around data intermediary tools.

Curbcut Toronto is an open-source urban data exploration and visualization tool, aiming to support urban data access and use. The original version of *Curbcut* was developed for Montreal, Canada by the McGill University Sustainability Systems Initiative, aiming to support flexible and

comprehensive exploration of urban sustainability data. A Toronto version of the tool was launched in partnership with the University of Toronto Digital Curation Institute.

Curbcut Toronto offers map-based visualizations for data related to various urban sustainability topics, including environment, health, housing, and transportation. All data comes from open data sources including City of Toronto's Open Data Portal [74], Canadian census data [44], Canadian Mortgage and Housing Corporation [19], the Canadian Urban Environmental Health Research Consortium [16] as well as OpenStreetMap [4]. The tool offers visualizations for one variable in different geographical scales as shown in Fig 1, bivariable relationships between two variables as shown in Fig 2, and visualizations for temporal changes as shown in Fig 3. *Curbcut Toronto* represents in many ways a common type of data intermediary tool: it combines open data from various sources in new ways, making them more accessible through visualizations and explanatory texts. It differs from other map-based visualizations tools by providing more flexible ways of visualization and analysis.
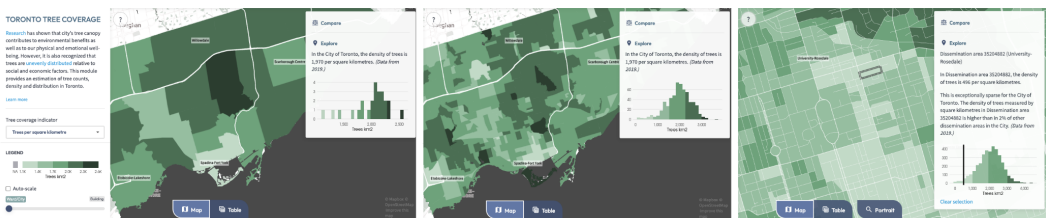

Fig. 1. Screenshots of *Curbcut Toronto* visualizing tree coverage in Toronto in three different geographic scales. Left: an example of visualization in ward level. Middle: an example of visualization in census tract level. Right: an example of visualization in dissemination area level.

Like other systems of this kind [2, 5, 78], *Curbcut Toronto* relies on layers of data intermediaries curating, preparing and processing data from a wide range of sources (Fig 4). For instance, the visualization module on Toronto tree coverage is prepared through incorporating the Canadian census data [44] and Topographic Mapping of Trees data from Toronto Open Data Portal [74]. For these steps to be possible, many layers of data intermediaries already worked to collect, process and curate these data. The data presented on *Curbcut* will also be passed on to future layers of data intermediaries through the export features that allow users to reuse the platform's datasets.

We chose to incorporate *Curbcut Toronto* in our semi-structured interviews as a technology probe because it offers a lens for provoking reflections and insights. The HCI and design community have had rich experiences adopting variations of cultural probes since its introduction by Gaver et al [40]. Since then, technology probe has become one of the most popular variations of the original method [12], where a technology artifact is used to collect information about the use and user of a technology and inspire new designs to support users' needs [55]. Boehner et al. [12] summarized that, within the HCI community, technology probes are used for various purposes from collecting information around usability for a single design to providing inspiration for opening up new design spaces. In our study, *Curbcut Toronto* serves more towards the latter purpose, as we hope to examine the general design space and visions for data intermediary tools as opposed to formulating a rigid design guideline for data intermediary tools to fulfill. Echoing previous works [39, 57], we chose to use *Curbcut Toronto* as a lens to provoke reflections and reactions, as well as to support participants to articulate thoughts, feedback and questions.
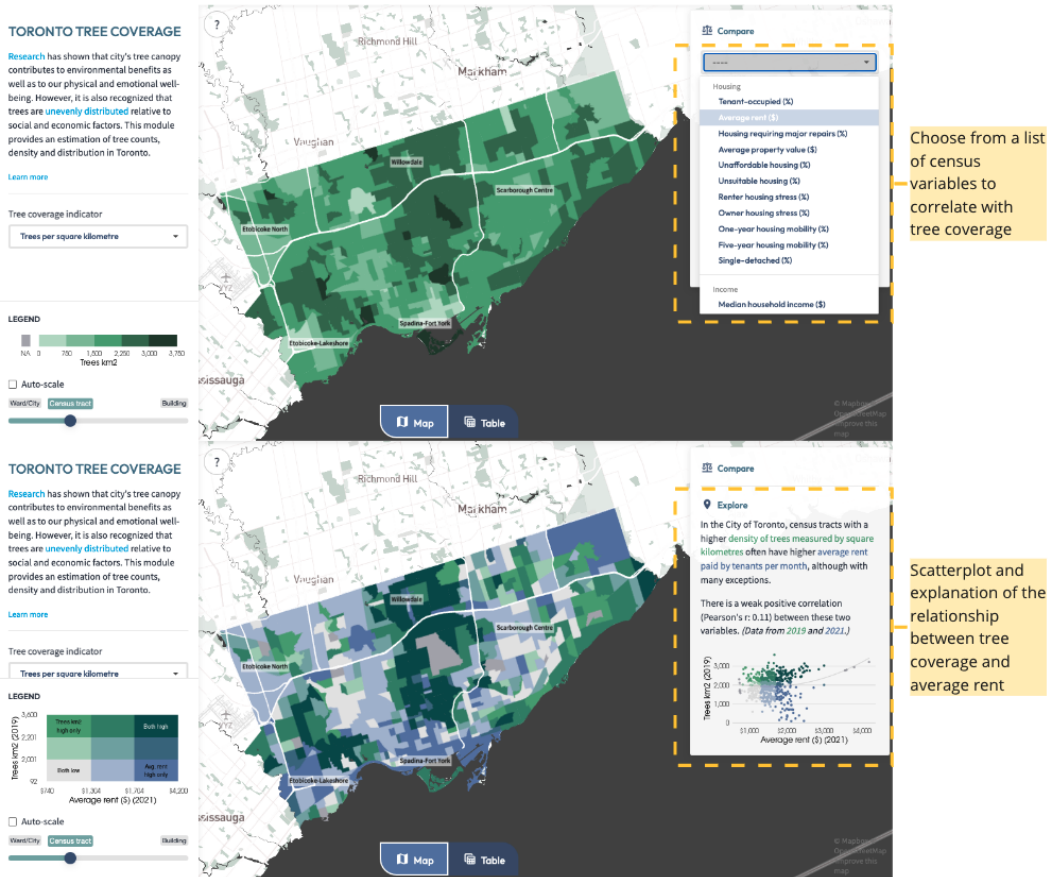
Fig. 2. *Curbcut Toronto* allows users to correlate the featured variable on the map with a census variable. In the top screenshot, the users can select a variable from a list of census variables. In the bottom screenshot, the user is visualizing the correlation between Toronto tree coverage and average rent. A description text box is showing on the right indicating that there is a weak positive correlation between the two variables.

3.4   Data Analysis

We adopted an inductive thematic analysis approach [15] to analyze the interview videos and transcripts, to iteratively form codes for understanding community advocacy groups' existing ways of using data and their visions for data intermediary tools.

The first and second author watched the interview recordings to transcribe interviews with notes and screenshots of how participants interacted with *Curbcut Toronto*. For each interview, the first and second author would first watch recordings, read and re-read the transcripts to generate initial codes and then come together in open coding sessions to read through the transcripts again, watch recordings when necessary, discuss codes and finalize the initial rounds of codes that were recorded using a coding software, ATLAS.ti.

After we conducted seven interviews, we documented five emergent themes, including the nature of advocacy work, ways of using data, feelings towards data, ways of using data intermediaries, and challenges of using data intermediaries. As we continued conducting more interviews, we also returned to literature around the topics of theoretical work of understanding urban data and
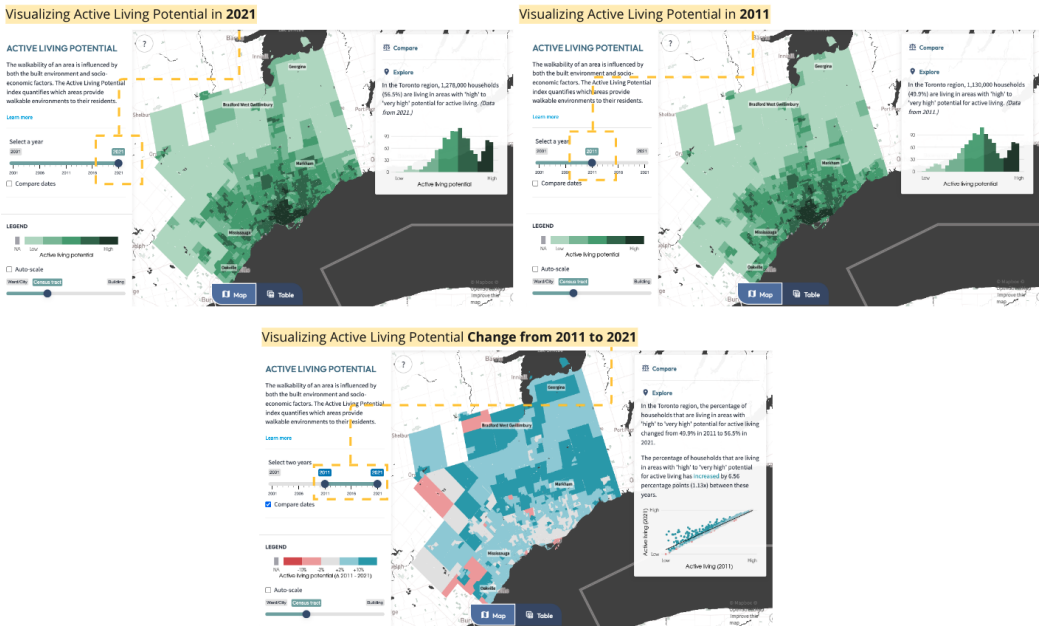
Fig. 3. Temporal exploration features of *Curbcut Toronto*. The tool allows users to look through available historic data and compare data at two different dates. Top left: visualizing active living potential in 2021. Top right: visualizing active living potential in 2011. Bottom: visualizing active living potential change from 2011 to 2021.

open government data [59, 68, 82], data practices in community works [13, 14, 27], and design practices used in data intermediary tools [27, 33, 88]. We iteratively went back to refine our codes and as we conducted 11 interviews, we transferred all codes to a Miro board, an online collaborative visualization platform, as sticky notes, and conducted iterations of affinity diagramming to summarize emergent themes around community groups' ongoing ways of working with data and their visions for data intermediaries. Informed by existing literature, emerging codes and themes from interviews, all authors came together to discuss four key insights in the result section below that illustrate the visions for data intermediaries from the perspectives of community advocates. As the relevance of data feminism principles emerged, the first author mapped each of the principles to all participants' quotes relevant to the four key themes and reviewed the mapping and codes with the last author.

The final result was shared with participants through a paper draft with each participant's quotes and insights derived from their interview highlighted as well as a recorded video presentation. Participants were provided opportunity for comments and corrections. We received feedback like "*so nice to hear from you and read this report and watch your video (very helpful!)*", "*It looks and sounds great. Congratulations on getting it to this point. I'm happy to have played a very small part in it*", "*Thank you so much! I really appreciate the video and the time you took to highlight my quotes*", and no corrections were made.

4 THE VISION AND THE PATHWAYS FOR DATA INTERMEDIARIES

From these interviews with community advocates, we identified four key themes around advocates' perspectives for data intermediary: an overarching vision of connecting *near data* and *far data*
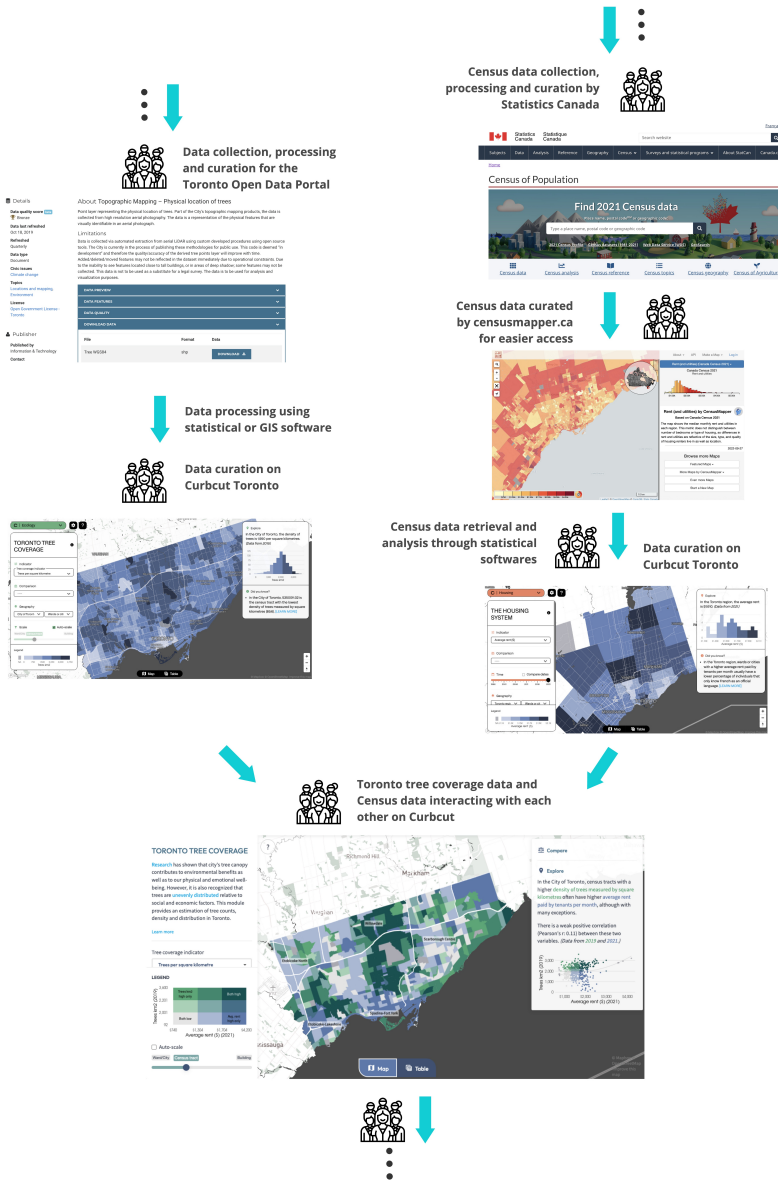
Fig. 4. An illustration of a segment within the urban data ecosystem focusing on how data passes through data intermediaries to get to *Curbcut Toronto*. Data intermediaries processing and curating data before it is presented on *Curbcut Toronto* include people working on the Toronto Open Data Portal, Statistics Canada, CensusMapper and so on. Designers of *Curbcut Toronto* act as another data intermediary. Data from Curbcut Toronto can then passed on to other data workers for future uses.

and three pathways for pursuing the vision. The vision and the pathways arise from community advocates' existing ways of working with data, tying closely to community advocates' objectives of challenging the status quo and re-imagining alternative urban futures. In the next sections, we

discuss advocates' current ways of working with data, how that connects with their vision for data intermediaries, and how the pathways for pursuing this vision support data intermediaries to put data feminism principles into practice. We labeled the emphasis on each of the seven data feminism principles throughout our result section as DF1 to DF7, corresponding to the list in related works section 2.3.

4.1 The Vision: Connect "near data" and "far data"

Community advocates shared their complex consideration of using and interacting with data they have direct experience with and know very well, as well as data obtained from distant platforms and stakeholders. We propose the concept of "near data" to refer to data that is close to community advocates' everyday lives, including self-collected data, lived experiences, and conversations. We use "far data" to refer to heavily processed data from other people, such as open data from government or academic institutions. We recognize that community advocates use a variety of data on the spectrum of near and far and that they view the distance of data as an important context for their advocacy work. Our concept of near and far builds on feminist critiques on how oftentimes data is presented as facts collected from a distance through what Haraway calls "the god trick," when in fact it represents perspectives from dominant groups [27, 49]. This concept of near and far data is also constructed as a response to D'Ignazio and Klein's data feminism call to examine and challenge power when working with data [27]. While community advocates actively use both near data and far data, they treat these data differently and have different ways of working with different types of data. Advocates emphasize the importance of near data to make sense of, complement, question, and challenge far data. From their views emerges a vision for data intermediaries to facilitate engagement with both near data and far data, connecting the two to balance both, empower near, and better use all data.

4.1.1 How advocates engage with far data. Advocates use far data in their daily operations, for example when they create reports or write grant applications, to furnish a broader statistical context and substantiate their claims or support their arguments. Almost all participants mentioned their various ways of using census data, City of Toronto's open data portal, and research data from academic institutions. Resonating with findings from Pei et al. [79], advocates use far data to legitimize their advocacy works to a variety of audiences. For example, P9 shared using census data to provide contexts for legitimizing their service delivery work to funders: *"I used Statistics Canada a lot. So in the reports I just have a general introduction to the park context, where I provide some context around the neighborhood that the parks are located in."* Similarly, P12 mentioned their use of open data in grant applications *"when our grant writer applies for funding for our project, they use open data to find information about park usage."* P17 shared that when communicating with their community members, they would leverage research data from both the city as well as journal articles from the global South, where a lot of experts on food sovereignty are located. P6 directly acknowledges that *"we rely heavily on open public data and we're very thankful for it."*

However, engaging with far data has presented several shared challenges among advocates. One of the most recurrent concerns is access. P1 shared their frustration, commenting that *"how does somebody like me who's fighting for safe streets get access to this data? And the answer was, well, you can't."* This frustration is also shared by P3, who finds it puzzling that after a third-party initiated environmental study on a local subway line, *"they won't let us see that information."* Financial constraints act as another substantial barrier. P4 mentioned *"Esri makes it free for students. But let me tell you, it is not free, even with a nonprofit. It's really expensive."* Many organizations are compelled to explore free alternatives, which may not offer the same quality as their paid counterparts, hindering advocates from effectively making use of open data for their work. In

addition to these concerns, both P9 and P16 pointed out that scientific measurements often fail to resonate with communities and that an important role of advocates is to translate results from statistical modeling to information that their community members can easily interpret and understand. These stories collectively illustrate concepts proposed by previous researchers on minoritization, disempowerment and inequity brought by datafication and over-reliance on far data [10, 14, 21, 59, 79].

4.1.2   Connect near data and far data. While advocates rely on far data, they see more value in near data to help them make sense of, enrich, question and challenge far data. Almost all participants mentioned relying on their embodied experience to understand the more abstract far data. P16 mentioned how community members made sense of the usually referenced two degrees in climate change relying on their lived experiences (DF3): *"I don't understand what two degrees means. I understand it being so hot that I can't get out of my apartment and all I can do is lie down."* Advocates also highlighted how near data enriches far data through contextualizing the abstract, making data more tangible and relatable to wider audiences (DF3, DF6). P1 argues that when near data is weaved in, *"the story is not just flat,"* but extends to *"how people feel, how people see, how people approach their day."* P7 connected near data and far data through filming hours of videos of themselves and their children riding on bike lanes that are in need of improvement alongside analysis of existing traffic collisions collected by the Toronto Police Service to advocate for safer biking infrastructure and street design for marginalized road users (DF3, DF5).

Advocates also connect near data and far data to question and challenge dominant views. Through P1's own traffic collision accident while biking, they found out that their data is missing in the publicly available data set and started to question and challenge the far data, leading to their initiative on a *counterdata* [27] project called "All Road Crash Victims Count," to allow victims of road accidents to self-report their experiences that are not captured by official sources (DF1, DF2, DF4). Similarly, P2, together with other residents in their neighborhood, identified missing speed camera data of their neighborhood in open datasets and advocated for speed reduction through referencing their years of knowledge living in the neighborhood (DF2, DF5). P11 connected their own experience navigating the healthcare system with public data on municipal political donations records to initiate projects on analyzing inequity and injustice in the public health system (DF1, DF2, DF3).

Beyond the direct usage of near data to complement or challenge far data, advocates engage with near data through more affective and embodied ways. As P1 expressed, *"it's still my data,"* emphasizing the personal and emotive attachment advocates have to their self-collected data and lived experience (DF3). They explained further, *"it's very rewarding to have that relationship and then to enrich that data and be able to weave friends and that qualitative piece into it."* P3 relies on their lived experience to tell their own stories. They remember every digit of their near data: *"I received a supplement every month for 158.25 with my disability payment to cover in those days...I didn't have to think about should I spend 7.50, 3.25 one way and 3.25 to go to therapy, or am I going to eat today."* When using near data, advocates feel their stories are more compelling, *"I can talk about those numbers. I don't have to look at a chart. The charts are in my head because I lived it."* P10 emphasized the need for more near data, as they offer depth, rooted in lived experience, but still lacking, *"there are real, lived experiences and there are real people who live in these areas and whose stories are not documented."*

With the ongoing effort in connecting near data and far data, community advocates share a vision for data intermediaries to support such endeavors. When advocates interact with *Curbcut Toronto*, a tool that only provides far data, this vision of linking near data and far data was repeatedly highlighted. Many advocates discussed how data intermediaries could provide technical support

for connecting near data and far data. Advocates also highlighted that data intermediaries would add much value in making advocates aware of available far data so that community advocates could better use or examine problems within these data with their near data (DF1). Furthermore, advocates emphasized the value of data intermediaries fostering local networks of community groups collecting near data to multiply existing efforts (DF7, DF1, DF2). Thus, from our conversations with community advocates, we summarize and discuss three potential pathways for data intermediaries to pursue the vision of connecting near data and far data – (1) align data exploration with diverse ways of storytelling, (2) communicate context and uncertainties, and (3) decenter artifacts for relationship building.

4.2 Pathway 1: Align data exploration with diverse ways of storytelling

Storytelling in advocacy serves as a potent way to weave together data and lived experiences to create compelling narratives [33, 79]. Community advocates from our interviews shared how they leverage data to form narratives in diverse ways for different audiences. Storytelling remains a center piece for engaging civic participation and influencing social change. Advocates' creative and thoughtful ways of using data in their storytelling also shed light on how data intermediaries could better align data exploration with diverse ways of storytelling.

4.2.1 Support access and exploration of far data. By walking us through the diverse ways of considering and using data, community advocates offered suggestions for data intermediaries to bridge technical gaps in data access and exploration for storytelling so that far data could be brought closer to advocates' data works. Many advocates (P5, P6, P10, P11) mentioned that when crafting their stories, they hope to know what data is available, but oftentimes they might not be aware of relevant far data that have been made available. As P11 emphasized, they hope to approach far data in an exploratory and flexible approach, not to *"prove [their] theory"* but to *"poke around and find a comprehensive picture"* and to be *"always open to find other things in the data"* (DF1, DF5). Thus, data intermediaries can play important roles in making far data more accessible and pointing advocates to possible data related to their work.

Almost all advocates communicate their insights and stories through reports and presentations that benefit from incorporating visualizations and maps. For example, P6 talked about storytelling through a *"final report in which we have a bunch of visualized data,"* where a lot of effort is put into resonating with their audiences. P7 shared that mapping brings a *"more street-level understanding and a more micro understanding of the work that I do"*. By connecting abstract data to concrete contexts, mapping makes the narratives tangible and relatable for their audiences (DF3). Community advocates found *Curbcut Toronto* to be useful when they could leverage visual ways of understanding the city and telling their stories. P9 shared that *"a lot of nonprofits don't have people who know how to use GIS, or if they do, they don't have the licensing because it's expensive. And so having maps like this that we could use for reports would be super helpful."* Similarly, P8 commented *"[By] adding colors on a map and knowing what areas are especially affected by this issue a lot, it really helps bring in the point and leads my curiosity into knowing more topics about it."*

Quantitative analysis support such as temporal, geographical comparison and correlation was also highlighted to be useful for supporting advocates in their storytelling. P13 commented *"I think my favorite part is the ability to compare changes across time in a way that is easy."* With a longitudinal perspective offered in data intermediary tools, advocates could easily enrich their narratives through examining how urban spaces have evolved over time, as P14 commented, *"if we know, for instance, trends around homelessness, maybe one year homelessness, increased or decreased. And then you can take a look at something like what was the housing cost at that time."* Beyond temporal comparisons, advocates (P5, P8, P14, P15) also make geographical comparisons between

one neighborhood and another to add new dimensions to their storytelling. P5 noted, *"so if I saw this bike comfort data, I could maybe compare it to another town."* P12 also talked about comparing infrastructure spending in affluent neighborhoods with less privileged areas to advocate for more funding for their local neighborhoods (DF1, DF2). P2 pointed out the value of having access to various geographic scales of data, because it allows users to ask critical questions like *"why does this disparity exist"* and *"how can we address it"* (DF1, DF2). Offering flexibility in exploring data in different temporal and geographic scales enhances advocates' existing ways of telling local stories. Another technical support from data intermediaries could be correlations of different data. P9 was advocating for more greenspace by connecting it to health benefits in their reports: *"there's this whole movement around understanding parks and green spaces' role in social prescribing."* P11 also leveraged correlation between household income and voting patterns to reveal underlying stories of their local provincial election. These technical ways of support align with community advocates' existing ways of storytelling.

4.2.2   Center marginalized ways of storytelling. Beyond technical support for accessing and using far data, advocates see more importance in recognizing diverse ways of understanding what counts as data and what counts as meaningful data. For instance, even though *Curbcut Toronto* is heavily focused on presenting quantitative data, many advocates highlighted the value of the *Toronto Stories* module (Fig 5). P14 appreciated how the module adds "a human dimension" to the sometimes "cold" and "not inviting" data (DF3). Presenting stories with writings, photographs and art works beyond quantitative far data empowers community advocates to embrace diverse ways of knowing and storytelling. P17 also emphasized including and acknowledging data that are often ignored by institutionally supported open data programs. They pointed to the data and maps on the lost rivers of Toronto [84] which *"influence a lot of things in our neighborhoods"*. Mapping these waterways centers Indigenous ways of knowing and understanding the city as a part of nature rather than apart from it (DF5). Although most urban data are associated with scientific ways of knowing associated with the census system, community advocates see much value in presenting and prioritizing local and Indigenous ways of knowing.
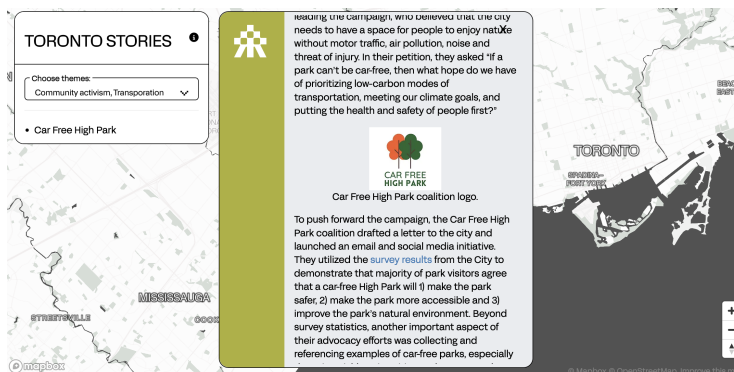

Fig. 5. Toronto Stories module on *Curbcut Toronto*. This module shows a map of place-based stories and advocacy campaigns happening around Toronto.

The pathway of aligning data exploration with diverse ways of storytelling emphasizes both making far data nearer and bringing various near data from the periphery to the center. Bridging technical gaps help bring far data nearer to community advocates, making data more accessible and usable and offering more opportunities for community advocates to examine issues or biases

embedded in far data. Through bringing far data nearer, data intermediaries can work together with community advocates to examine power (DF1) and challenge power (DF2) around the shared concerns that are dominated by unequal power structures. At the same time, obtaining more space for marginalized ways of knowing is especially important for advocates. They push for incorporating historically marginalized near data used by Indigenous communities for understanding cities as well as storytelling techniques other than scientific analysis. Advocates' existing ways of working with data are reflected in data feminism principles of elevating emotion and embodiment (DF3) and embracing pluralism (DF5). They envision data intermediaries that align with and empower their diverse ways of storytelling, valuing multiple forms of knowledge, synthesizing varied perspectives through affording diverse ways of exploring and visualizing data.

4.3 Pathway 2: Communicate context and uncertainties

While urban data is usually being presented in an authoritative narrative of accurate depictions of cities [60, 68], community advocates are highly experienced in recognizing the limitations of urban data. Therefore, advocates emphasized that it is important for data intermediaries to communicate contexts and uncertainties behind data in order to open up spaces for dynamic discussion to imagine urban futures.

4.3.1 Communicate context of far data. Nearly all participants recognized limitations of quantitative urban data through their various experiences of working with data for advocacy. To work with and around these limitations, participants emphasized the importance of questioning context when making sense of data. P11 highlighted the importance of questioning power and diversity behind data: *"I think that data explorations are really important to talk about power and diversity of areas… It's often the most powerful group of people doing something that impacts others"* (DF6, DF1). P13 recognized that quantitative data only offers partial views, and they discussed the importance of incorporating qualitative data to provide contexts: *"It's really important to have both [quantitative and qualitative data] for context, because numbers can't capture the entire impact that we're having…There are a lot of things that are not quantifiable"* (DF6, DF5). P5 and P12 emphasized the importance of *"deliberately"* designing their survey questions for data collection to go beyond numbers by including questions focused on the *"why"* to provide better context. P5 stated, *"we chose this [survey question] because we really want to know why, what prevented people from biking."*

Urban indicators, usually found in maps created by data intermediaries, add another layer of complexity to community advocacy work. Nearly all advocates were aware of the limitations and critiques related to urban indices. For example, when interacting with *Curbcut Toronto*, many participants (P1, P6, P7, P9, P11) critiqued the lack of context associated with the bikeway comfort and safety index. For example, P9 challenged the index through referring to their own lived experience, *"this area probably does have the most by quantity bike lanes and bike infrastructure. But I also feel that this is an area that I don't like to bike to work because it doesn't feel safe, because it's so dense and there is so much traffic"* (DF3). P7 also commented on the nuances in feeling comfortable, *"I wonder what their context of feeling comfortable is. There are individuals that think just because there's paint on the ground that you've got a bike lane. That's perfectly fine, but are you gonna let an 8 year old ride down a painted line or would you rather be in an actual protected piece of infrastructure"* (DF6, DF4). P9 posed several questions on the contextual information that they hoped to understand, *"how is this data collected? Is this a survey? I wonder if this is about the literal number of infrastructure? Especially with a metric like this, if it's based on people's feelings of safety or some sort of audit or something."*

These experiences and expertise when sifting through data, questioning context, asking questions to delve into the context beyond data are transferred to their visions for data intermediaries'

work. Before P11 opened up *Curbcut Toronto*, they commented: *"before I even click into this, my biggest thought is: does it give folks data with enough context, which is a hard thing to automate. But that's always my thing because anyone can throw out a number, but sometimes they have ridiculous numbers"* (DF6). P11 added that when they use open data from different departments of the government, the discrepancies and various ways in presenting the same information, such as budget, constantly remind them that *"data isn't just bulletproof infallible information."* Other participants also recognized the dynamic nature of data, commenting that *"nothing is perfect and everything is ever evolving"* (P15) and *"we have to view it [data] as something that's living because it's changing all the time. People's experiences are changing all the time, and so that is never going to be stagnant, even though we might want it desperately to be"* (P16), and so data intermediaries should be mindful in acknowledging the dynamic nature and partiality of data.

4.3.2   Reflect on intentions and values. Beyond acknowledging uncertainties behind data, community advocates also emphasize the importance of communicating the intentions and motivations behind data intermediaries' work. When presenting data through visualizations, P11 was aware of how different people can easily have different interpretations (DF6). They emphasized that providing context through complementary texts or blog posts is crucial for communicating their intent and stories with clarity, and explained: *"so the thing with data visualization is that I'm very careful not to ever present it on its own. Everyone learns in different types of ways and takes information in different types of ways. So I would never just send you this link and not tell you anything about it...I would embed this in a post about why you would want to look this up."* P17 urges data intermediaries to be explicit about their positionalities, values, motivations behind their data works and why they think presenting a certain type of data is important to their values (DF6, DF7). P17 shared how different people might have different definitions on what urban sustainability means to them including *"financial sustainability, environmental sustainability, or working towards a sustainable earth, a sustainable earth with humans."* Therefore, they appreciate when data intermediaries *"name very clearly what [they] are aiming for."*

Expressing uncertainties in data and processes of data analysis might be daunting for data professionals because the field has been conventionally driven by ideals of objectivity and neutrality. However, data are never raw or neutral [43], and community advocates are very well aware of that. Pursuing the seemingly objective approaches of data science undermines the quality and effectiveness of both far data as well as near data. Being explicit of the limitations of the usually 'objective' far data is a commitment to data feminism's principle of rethinking binaries and hierarchies (DF4). Clarifying the positionalities and intentions of data intermediaries support the recognition and use of near data. When data intermediaries surface motivation of far data work, they not only provide future data users an important aspect of the context, but also acknowledge that far data, similarly to near data, are always coming from feeling bodies in this world, reflecting their values and situated knowledges [49]. This supports working towards data feminism's principle of elevating emotion and embodiment (DF3), making labor visible (DF7), and embracing pluralism (DF5). Advocates' envisioned pathway for communicating uncertainties and providing context is a crucial step towards the commitment to co-liberation [29] for both data intermediaries as data facilitators as well as other stakeholders relying on their work as voices that challenge dominating narratives and reimagine future cities.

4.4   Pathway 3: Decenter artifacts for building relationships

While community advocates see value in data intermediaries designing tools that support advocacy work, they highlighted the relational nature of their work and emphasized the importance of relationships beyond transactional interactions through technology artifacts. This theme echoes

previous works looking into civic tech and community groups' roles of capacities building beyond the mere production of end products [66, 70, 98]. Through fostering long term connections, data intermediaries and community advocates can collectively gain more civic capacity to connect near and far data for understanding urban issues and to build solidarity among different values in shaping urban futures.

4.4.1 Synergize data work through relationship building. Most participants (P1, P3, P4, P5, P6, P7, P9, P10, P11, P12, P14, P16, P17) shed light on the relational nature of advocacy and how the values of prioritizing relationship building shaped their visions for data intermediaries. For some, building networks with and for community members is a key focus of their work (P1, P9, P10, P12). Advocates also shared that it is common to see people involved in multiple community groups, which opens up opportunities for working together and influencing each other. P11 emphasized that *"I don't work in isolation and this ends up dictating what I'm working on."* P7 summarized that *"many advocates are wearing multiple hats with multiple lenses"* and work together towards the goals that *"synergize with each other."*

Advocates recognize that data is only a starting point of their work. This recognition often allows them to build more local capacity for better access to far data and for discovering valuable near data. When obtaining and analyzing urban data, community groups prioritize values of care [13, 14], building symbiotic relationships that support each other in resource-constrained situations. P6, P8, P13 and P14 all mentioned the importance of relying on their existing relationships to access and make sense of far data. P6 explained that they were able to access data related to how people use a local helpline for social services *"only because we had that connection"* and *"their GIS app helped us create the picture we want."* Similarly, P11 shared that as they became familiarized with a data space, they were happy to help other groups to navigate. Advocates also described how they support other advocates and community groups through what P1 called a *"symbiotic relationship."* P1 described how they got into advocacy work through the help of other advocates' networks and commented that *"my connections with all those others have been strengthened and my goal is always to support the organization and then together it's a symbiotic relationship. I get better from their activities, they benefit from my activity."*

Advocates' focus on relationship building also fostered a more collective effort in empowering the use of near data. P1 started their initiative on collecting urban noise data and was able to make visualizations and build a website for the project through connecting with other people and sharing their experiences at local events, where collaborations emerged. P2 shared their experience working with a group of people to push forward a campaign on local park access, where they together collected thousands of local survey responses about people's experiences in parks. P3 collaborated with journalists to write articles that make visible their lived experiences navigating urban life after a brain injury. This amplified voices of those experiencing hurdles related to physical recovery, mental health, and access to affordable housing. These advocacy efforts in collecting and telling stories leveraging near data were achieved through advocates' constant effort in building long term and trusting relationships through local events, sharing back with their communities and embracing new relationships that arise through their work (DF2, DF5, DF7). The dynamic, organic relationship among advocates was summarized by P7 as *"it [advocacy work] may start with me, but I'm only there for a little bit before I'm bringing in a data visualization person, or a teacher from a school, or a community member that's been in the neighborhood for 25 years, or the politician, and maybe I get to step back and I let others go forward."*

4.4.2 Build relationships with and for community advocacy groups. Community advocates' attention to relationships beyond mere transactions were again reflected when they interacted with *Curbcut Toronto*. They emphasized the vision of building long term relationships with data intermediaries

and expanding their network through data intermediaries' work. Several participants mentioned that they would not trust a data intermediary tool without knowing the team behind it (DF7). P6 commented that *"I would trust it if it was offered to my organization through outreach, or I heard about it through a workshop or webinar."* Similarly P7 mentioned that data intermediary tools need ambassadors and advocates *"who can speak to the clarity of this tool"* or *"launch an AMA, ask me anything."* Many participants (P10, P14, P15, P17) also shared that the information provided by a data intermediary might not be the most important aspect, but a starting point for conversations and discussions for building connections among different stakeholders. P10 commented that *"we have team members who are still in school and we have team members who are seasoned professionals"* and the tool could help *"open up a lot of ways of understanding for the team so we can share our thoughts and have a mutual connection"* (DF5). P17 stressed the importance of data intermediaries to present other group's ongoing effort in leveraging near data, so that groups will be able to learn from each other and work in solidarity (DF7): *"what I would do personally is if I saw a case study dealing with the same issue over here, I would give them a call…[Data intermediary tool] doesn't need to provide all the information, it just needs to provide a story so that I can go and talk to them myself."*

Advocates found data intermediary's work to be more valuable if they collaborate with existing local social infrastructures and civic groups. To better support ongoing community work, P1 suggested data intermediaries to connect with communities such as civic tech groups and local schools, explaining that *"it is not necessarily presenting a tool for people to use for their own purposes, but also educating people on their surroundings."* P5 discussed how data intermediaries could bridge the gap for high school students who are starting to learn about urban information. P7 discussed connecting with an even broader audience and emphasized that it is important to *"get the regular public to be aware of the things around them and have easy access to all of these data pieces…don't shy away from being able to have elementary schools gain access to something like this."* P3 explained how they benefited from affordable classes offered by their local libraries and are advocating for more: *"So let's say if I was taking a class on Google tools, or Photoshop or Windows 11, we should have the same thing with the information where we teach our community members at a minimum cost to learn how to get to the data."* P7, giving back to their community through creating an interactive map on community resources for their local library, commented *"the library is being really active in being a community hub as a piece of social infrastructure. So that's my little give back with all of these maps that I'm creating."*

When working with data, advocates pointed out that data is not the purpose but relationship and capacity building is much more important. P1's eventual goal of their data work is going beyond data *"to push it out to the public to educate, to engage, to make aware,"* and so they are learning press releases and also writing a blog on the Toronto Open Data Portal. For P4's group, the primary goal of their data collection project *"was not the data, it was the conversation. So we were trying to get people to think about their home energy use."* Similarly, P11 involved data in their work because *"it empowers people to speak with their friends, their family and their colleagues."* P17 emphasized that the crux of advocacy work lays beyond data in finding a strategy that can move forward even when there are divergent views. Pursuing this pathway allows data intermediaries to follow data feminism's call for looking at data from an embodied perspective, putting more attention to the people who work on the data and use the data, and making that labor visible (DF3, DF7). Advocates share an understanding that data tools are only starting points of supporting ongoing efforts to shape urban futures beyond dominant imaginaries. Only when local networks are formed will there be more capacity for connecting near data and far data, embracing diverse perspectives, and finding solidarity in differences to enact change.

5 DISCUSSION

5.1 Actions for putting data feminism into practice

Data intermediaries are never neutral. Previous works in HCI and CSCW have illustrated how community advocates enact data feminism in their data advocacy work. For instance, Darian et al. [24] presented how nonprofit advocacy groups enact data feminism principles when reappropriating data across various functions from the margins of the data economy. Paudel and Soden [76] used the data feminism principles to evaluate open data platforms and highlighted the gaps in how these platforms lack consideration in values from data feminism, such as making labor visible, engaging with affective aspects of disasters, and challenging power. Tran et al. [94] studied local housing activist organization's data work through the lens of data feminism and proposed design implications when situating data for grassroot organizers. The work of counterdata is also surfaced by D'Ignazio [27] as a central practice for data activism, especially when data intermediaries fail to produce marginalized information and reproduce oppressive narratives. Extending this growing body of work, we reflect on how community advocates' vision for data intermediaries surface data feminism principles, and we provide concrete actions for data intermediaries to put data feminism into practice.
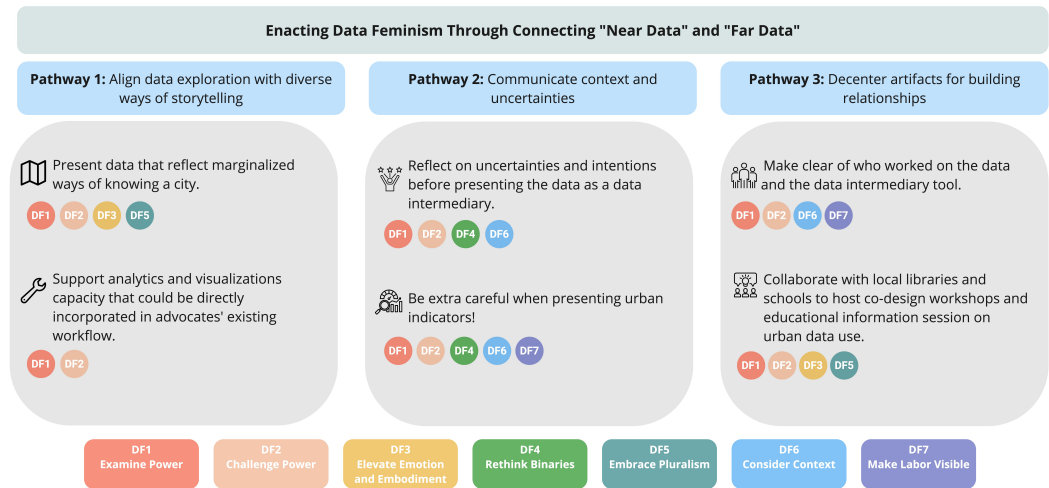

Fig. 6. Starting points and actionable items suggested by community advocates for data intermediaries to connect near data and far data. Each recommendation was raised by multiple community advocates and speaks to data feminism principles.

Throughout the results section, we illustrated how our interviewees' ways of working with data and their visions for data intermediaries reflect principles of data feminism, labeling their quotes with DF1 to DF7. We distill advocates' insights and perspectives into Figure 6 to offer data intermediaries several actionable next steps for connecting near data and far data along each of the three pathways. Each recommendation is tagged with the related data feminism principles. We argue that following these recommendations can help data intermediaries enact data feminism and support those who rely on their work to further enact data feminism. For instance, intermediaries should present data that reflect marginalized ways of knowing a city, such as the lost rivers of Toronto [84]. By incorporating marginalized experiential ways of knowing, data intermediaries can examine power (DF1) to challenge power (DF2). Their work in turn can elevate emotion and

embodiment often neglected by data-centric tools (DF3) and can serve as a medium for others to embrace and create plural narratives of urban spaces (DF5). They should be careful when presenting urban indicators, decomposing the messy meaning and work behind the numbers. These actions help data intermediaries to rethink binaries (DF4) often introduced in urban indicators, consider the context (DF6) of why and how these indicators were created, and make those who worked on defining and using these indicators visible to the users (DF7), whether the users engage with these indicators or refuse to be constrained by them. These actionable items serve as concrete starting points for those data intermediaries who are ready to enact data feminism to shape urban futures that center justice, diversity and inclusivity.

5.2 Data intermediaries' roles in pursuing "The Right to the City"

As Currie et al. [23] put forward, data work is central for pursuing the Right to the City in the datafied city today. We build on this argument and argue that data intermediaries play crucial roles in pursuing the Right to the City.

Data intermediaries work is never just making it easy for more people to access data. Every design decision – what data to include and exclude, how to frame explanations, and how the interface guides user interaction – is political. As Loukissas argues, "interfaces recontextualize data" [64, p.125]. Data intermediaries create new contexts for data that shape how users engage with the information, what ways of knowing are made available, how users connect data into their own knowledge systems, and how they will navigate their right to the city. When data intermediaries recognize the distance between their work and the data they work with, they become more attuned to the data's original contexts and how their processing and design choices reshape it.

Data intermediaries' work is also inextricably linked to the Right to the City through addressing data justice concerns [23]. There are different approaches to data justice [26, 34, 48, 73, 75, 91]. Here, we use Taylor's framework [91] to illustrate data intermediaries roles, as it clearly recognizes both the potential and limitations of data technology. Under the framework, there are three pillars of data justice: visibility refers to the right to be visible and invisible; engagement refers to the freedom of accessing and the freedom to not use particular technologies; non-discrimination emphasizes challenging bias and the freedom to be not discriminated against. The vision of connecting near data and far data is closely linked to all pillars of data justice. As discussed in Section 4.1, advocates pointed out how connecting near data and far data allowed them to make their lived experience more visible to the public. At the same time, advocates emphasized how data intermediaries' work of aggregating and merging data serves as key process to protect privacy for those who wish to remain anonymous. Connecting near data and far data also ease the daunting process of making sense of far data, supporting those who hope to engage. Decentering artifact for relationship building allows data intermediaries to foster relationships beyond mere transactional interaction [70]. When conversation is opened up and local networks are fostered, community advocates will be empowered to decide when not to use a technology because they can achieve their advocacy goals through relational ways of skills and data sharing. Lastly, connecting near and far data supports users to notice what is being discriminated against and what is missing. Rather than elevating far data to the point where it marginalizes and invalidates lived experience, the vision guides data intermediaries to support users in appreciating diverse types of data, examining whose data is represented and how data is presented. For many advocates, lived experiences are in fact crucial both for making sense of that far data and for pinpointing the biases and uncertainties embedded in those data. When near data, such as shared experiences, bring advocates together, these collective efforts will have a greater impact on pursuing the Right to the City.

5.3 Locating "near data" and "far data"

Near and far are not static qualities of data but dynamic, fluid and relational qualities, dependent on who is working with the data, how it is presented and passed down by data intermediaries across the urban data ecosystem, and to whom it is presented. When interacting with data, do we know the people who created the data? Do we know how the data are collected, processed, and presented? What are the values and intentions of the people who selected the data and worked on them? How were these values baked into the data when being collected and processed? [90]

The quality of near and far can be influenced by geographical and temporal, but most importantly social and socio-technical dimensions. Drawing from social and environmental psychology research, Bhardwaj et al. [11] argued that psychological distance can be understood in four dimensions (social, temporal, geographical, and hypothetical) and that these dimensions present important factors for data visualization designers to consider as they affect how people perceive, negotiate, reason and decide [83, 95]. Similarly, we argue that the quality of "near" and "far" is especially important for data intermediaries to consider when working with data and presenting data. Data are considered near when they arise in close proximity: for example, from our own experience of biking in our own city in the recent past. In contrast, projections about future urban heat islands and their effects on bikeway comfort may be geographically and socially close but are temporally removed and hypothetical. How advocates relate to a data set presenting such information, how they work with it, and how they decide to trust it is contingent on this distance. Anticipating that relationship can help data intermediaries design the appropriate context documentation, including heat models and evaluations underpinning a comfort index.

Many advocates collect near data, including the urban noise dataset for P1, community garden mapping for P4, as well as biking experience documentation for P7. These are near data because the advocates know who is collecting the data, how the data is collected and why the data is collected. To someone who comes across these datasets online, they may not be: it depends on how the data intermediaries re-present these data. With the right documentation and explanation, the data can become closer to those encountering them. That is a central role and value of data intermediaries. When data intermediaries provide data access and analytics support to other stakeholders in the urban data ecosystem, they should reflect on their distance to the data. Their design decisions will impact how others experience and understand the data as near or far.

As Loukissas argues in "All Data are Local" [64], data workers and users should learn to analyze data settings rather than data sets, attending to the places, institutions, processes and people that shape the data. This is why near data is extremely valuable. Designs of community-oriented advocacy platforms, such as *Smell Pittsburgh*, are often most effective when they seamlessly embed individually collected data, near to the users who submit them, within the geographic and temporal context of community advocacy [53]. By allowing community members to recognize others' data instances within the community context, the platform as a whole becomes an effective advocacy tool and is able to present the sum of the near data as a far-reaching dataset. We contend that reflecting on the nearness and farness of data offers a concrete approach for data intermediaries to attend to data settings.

When we emphasize the importance of near data, we are not dismissing far data, but in fact, far data offer fresh perspectives especially for those engaging from afar. Historical data, for instance, allow us to uncover stories our ancestors experienced and reveal temporal trends. Spatially distant data help us recognize differences and facilitate comparisons. However, we need to be careful with abstractions associated with far data. Far data risks being regarded as singular fact presented in a "view from nowhere" that, as feminist scholars point out, is an illusion [29, 49]. In Seeing Like a State [86], Scott examines how scientific rationalism is applied to govern people through simplification,

raising concerns about decision-making processes that reduce everyday experiences to calculable abstractions. In CSCW and HCI, Dourish [31] argues that the development of information systems is strongly related to systems rationalism underlying the phenomena that Scott analyzes. More recently, Wong et al. [100] also shows how Scott's concept of simplification to govern is reflected in AI ethics toolkits. In the context of data intermediaries, presenting far data through abstraction and aggregation of geographic or temporal scale is not necessarily bad: data users are not expecting to see every tree in a city throughout history labeled on a map. However, data intermediaries should reflect on their distance to the data, their processes of abstraction, the locally situated practices and near data that are abstracted away, and prioritize design decisions that enable users to also reflect, examine and challenge the data provided.

5.4 Tensions of connecting "near data" and "far data"

We recognize that connecting near data and far data is a complicated and messy process. Here we illustrate two tensions that could arise and call for research into how data intermediaries and community advocates can work together to navigate these tensions.

We recognize that urban governance prioritizes conventional data science and technical analysis, which leads to the tension between codified statistical analysis and advocates' experiential ways of knowing and storytelling. As P9 noted, high-level statistical analysis, while invaluable, may not always be the most effective tool for advocacy work: *"I simplified it [regression model] into like 'there is an association between these variables'... because, for our audience, it's not really that easy to interpret."* Neff et al. [72] articulated that "stories occur before data production, during production, and are used in exchange to give data meaning across communities with different expertise, cultures, and practices." This perspective highlights the significance of storytelling at every phase of data work and emphasizes the value of embracing diverse cultures and practices when working with data. This tension has been pointed out by many scholars in feminist HCI, critical data studies and adjacent fields [8, 28, 33, 43]. It intensifies in the context of designing data exploration tools for urban sustainability advocacy. We invite researchers to explore in the context of data intermediaries: how can data intermediaries work with advocates to prioritize experiential ways of knowing over codified statistical analysis in practice? How would this differ across the spectrum of data intermediaries work and various types of advocacy work?

We also recognize the tension between community advocates' hope for an *"impartial third party"* to provide data and the impossibility for any data intermediary to be genuinely impartial. During our interview, P1 explained the push back they got when advocating for bike lanes in local neighborhoods, *"car people tend to look at our information with a grain of salt. They're like, you guys are biased in this, in that. So having impartial third party data to show is hugely important."* P3, as someone who is distrustful of many sources of data, emphasized that *"we need access to accurate data where a third party is collecting the data."* Even though there is a need and tendency for data projects to present themselves as objective and authoritative *"third parties"* [28], it is crucial for data intermediaries to make clear their positionalities, intentions and values that influenced the work that they are doing, to work towards co-liberation, the underlying commitment of data feminism [28]. To continuously work towards liberating data science from the dominant view of objectivity and authority, key questions arise for the HCI and CSCW community: how can data intermediaries better communicate intentions and uncertainties in data and data visualization? How can advocates and data intermediaries work together to reflect on contexts and uncertainties from multiple perspectives? How can reflexivity become a key component of data work? [71] We see value in keep asking these questions and invite more HCI researchers joining us to work towards datafied urban futures that center social justice and inclusivity.

6 CONCLUSION

To quote Neff et al., data intermediaries should strive to make data "less terrible" [72, p.88], not more, and this requires data intermediaries to "think with, not for or against" [72, p.93] community advocates. Community advocates have rich lived experience and unique perspectives to offer those of us working with and studying data. This research explores community advocates' visions for data intermediaries, a crucial stakeholder in the urban data ecosystem facilitating interactions amongst almost all stakeholders. Through conducting interviews with 17 community advocates working with 23 different community groups, we surface the important quality of near and far data to be considered in data intermediaries' work. We contribute a description and reflection on community advocates' vision for data intermediaries, connecting near data and far data, as well as three pathways for pursuing this vision: align ways of data exploration with diverse ways of storytelling, communicate context and uncertainties, and decenter artifacts for relationship building. We illustrate how following the pathways allows data intermediaries to bring data feminism principles into the specific context of their work. To conclude, we propose several actionable next steps for data intermediaries, discuss on how these pathways contribute to the pursuit of "the Right to City" in the datafied city today, and reflect on tensions of working within the urban data ecosystem. We reinforce the importance of questions around the role, intention, and impact of data intermediaries. With relational ways of working that center justice and pluralism, data intermediaries can collectively work together with other stakeholders to pursue the Right to the City: a change that is more than legal, technical or procedural, but a "political capability – a configuration of civic life that can enable and nurture solidarity as grounds for struggle." [25, p.XVi]

7 ACKNOWLEDGEMENTS

REFERENCES

[1] 2023. *BikeSpace - Toronto's Bike Parking App.* https://bikespace.ca/
[2] 2023. HealthyPlan. https://healthyplan.city/en
[3] 2023. *Open Water Data.* https://www.openwaterdata.com/
[4] 2023. *OpenStreetMap.* https://www.openstreetmap.org/
[5] 2024. Manhattan Population Explorer. http://manpopex.us/
[6] Obead Alhadreti and Pam Mayhew. 2018. Rethinking Thinking Aloud: A Comparison of Three Think-Aloud Protocols. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. ACM, Montreal QC Canada, 1–12. https://doi.org/10.1145/3173574.3173618
[7] Irina Anastasiu. 2019. Unpacking the Smart City Through the Lens of the Right to the City: A Taxonomy as a Way Forward in Participatory City-Making. In *The Hackable City: Digital Media and Collaborative City-Making in the Network Society*, Michiel de Lange and Martijn de Waal (Eds.). Springer, Singapore, 239–260. https://doi.org/10.1007/978-981-13-2694-3_13
[8] Cecilia Aragon, Clayton Hutto, Andy Echenique, Brittany Fiore-Gartland, Yun Huang, Jinyoung Kim, Gina Neff, Wanli Xing, and Joseph Bayer. 2016. Developing a Research Agenda for Human-Centered Data Science. In *Proceedings of the 19th ACM Conference on Computer Supported Cooperative Work and Social Computing Companion (CSCW '16 Companion)*. ACM, New York, NY, USA, 529–535. https://doi.org/10.1145/2818052.2855518
[9] Jo Bates, Yu-Wei Lin, and Paula Goodale. 2016. Data journeys: Capturing the socio-material constitution of data objects and flows. *Big Data & Society* 3, 2 (Dec. 2016), 2053951716654502. https://doi.org/10.1177/2053951716654502 Publisher: SAGE Publications Ltd.
[10] Ruha Benjamin. 2019. *Race after technology: abolitionist tools for the New Jim Code.* Polity Press, Cambridge, UK.
[11] Eshta Bhardwaj, Han Qiao, and Christoph Becker. 2024. Limits at a Distance: Design Directions to Address Psychological Distance in Policy Decisions Affecting Planetary Boundaries.

[12] Kirsten Boehner, Janet Vertesi, Phoebe Sengers, and Paul Dourish. 2007. How HCI interprets the probes. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, San Jose California USA, 1077–1086. https://doi.org/10.1145/1240624.1240789

[13] Ashley Boone, Carl Disalvo, and Christopher A Le Dantec. 2023. Data Practice for a Politics of Care: Food Assistance as a Site of Careful Data Work. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 152, 13 pages. https://doi.org/10.1145/3544548.3580831

[14] Chris Bopp, Ellie Harmon, and Amy Voida. 2017. Disempowered by Data: Nonprofits, Social Enterprises, and the Consequences of Data-Driven Work. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (Denver, Colorado, USA) *(CHI '17)*. Association for Computing Machinery, New York, NY, USA, 3608–3619. https://doi.org/10.1145/3025453.3025694

[15] Virginia Braun and Victoria Clarke. 2006. Using thematic analysis in psychology. *Qualitative Research in Psychology* 3, 2 (Jan. 2006), 77–101. https://doi.org/10.1191/1478088706qp063oa Publisher: Routledge _eprint: https://www.tandfonline.com/doi/pdf/10.1191/1478088706qp063oa.

[16] CANUE. 2023. The Canadian Urban Environmental Health Research Consortium. https://canue.ca/

[17] Ramya Chari, Luke J. Matthews, Marjory S. Blumenthal, Amanda F. Edelman, and Therese Marie Jones. 2017. *The Promise of Community Citizen Science.* Technical Report. RAND Corporation. https://www.rand.org/pubs/perspectives/PE256.html

[18] 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*. ACM, Portland Oregon USA, 835–846. https://doi.org/10.1145/2998181.2998265

[19] CMHC. [n. d.]. Canada Mortgage and Housing Corporation. https://www.cmhc-schl.gc.ca/

[20] Jacqueline Corbett, Mathieu Templier, and Hirotoshi Takeda. 2018. The Making of a 'Top' Open Data City: A Case Study of Edmonton's Open Data Initiative. https://doi.org/10.24251/HICSS.2018.308

[21] Roderic Crooks. 2019. Accesso Libre: Equity of Access to Information through the Lens of Neoliberal Responsiblization. *Journal of Critical Library and Information Studies* 2, 1 (March 2019). https://doi.org/10.24242/jclis.v2i1.91 Number: 1.

[22] Roderic Crooks and Morgan Currie. 2021. Numbers will not save us: Agonistic data practices. *The Information Society* 37, 4 (2021), 201–213. https://doi.org/10.1080/01972243.2021.1920081 arXiv:https://doi.org/10.1080/01972243.2021.1920081

[23] Morgan Currie (Ed.). 2022. *Data Justice and the Right to the City.* Edinburgh University Press, Edinburgh. https://doi.org/doi:10.1515/9781474492973

[24] Shiva Darian, Aarjav Chauhan, Ricky Marton, Janet Ruppert, Kathleen Anderson, Ryan Clune, Madeline Cupchak, Max Gannett, Joel Holton, Elizabeth Kamas, Jason Kibozi-Yocka, Devin Mauro-Gallegos, Simon Naylor, Meghan O'Malley, Mehul Patel, Jack Sandberg, Troy Siegler, Ryan Tate, Abigil Temtim, Samantha Whaley, and Amy Voida. 2023. Enacting Data Feminism in Advocacy Data Work. *Proc. ACM Hum.-Comput. Interact.* 7, CSCW1, Article 47 (apr 2023), 28 pages. https://doi.org/10.1145/3579480

[25] Lina Dencik. 2022. Foreword. In *Data Justice and the Right to the City*. Edinburgh University Press, XV – XVii.

[26] Lina Dencik, Fieke Jansen, and Philippa Metcalfe. 2018. A conceptual framework for approaching social justice in an age of datafication. (2018).

[27] Catherine D'Ignazio. 2022. *Counting Feminicide: Data Feminism in Action.* PubPub. https://mitpressonpubpub.mitpress.mit.edu/counting-feminicide Book Title: Counting Feminicide: Data Feminism in Action.

[28] Catherine D'Ignazio and Lauren Klein. 2021. Who Collects the Data? A Tale of Three Maps. *MIT Case Studies in Social and Ethical Responsibilities of Computing* Winter 2021 (Feb. 2021). https://doi.org/10.21428/2c646de5.fc6a97cc Publisher: MIT Schwarzman College of Computing.

[29] Catherine D'Ignazio and Lauren F. Klein. 2020. *Data Feminism.* The MIT Press, Cambridge, Massachusetts.

[30] Gayatri Doctor and Prajakta Joshi. 2022. Empowering Cities through Open Data - Open Government Data Initiatives in India. In *Proceedings of the 14th International Conference on Theory and Practice of Electronic Governance (ICEGOV '21)*. Association for Computing Machinery, New York, NY, USA, 352–361. https://doi.org/10.1145/3494193.3494241

[31] Paul Dourish. 2007. Seeing like an interface. In *Proceedings of the 19th Australasian conference on Computer-Human Interaction: Entertaining User Interfaces*. ACM, Adelaide Australia, 1–8. https://doi.org/10.1145/1324892.1324894

[32] Graham Dove, Jack Shanley, Camillia Matuk, and Oded Nov. 2023. Open Data Intermediaries: Motivations, Barriers and Facilitators to Engagement. *Proc. ACM Hum.-Comput. Interact.* 7, CSCW1, Article 78 (apr 2023), 22 pages. https://doi.org/10.1145/3579511

[33] 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*. ACM, San Francisco California USA, 1273–1283. https://doi.org/10.1145/2818048.2820068

[34] Virginia Eubanks. 2018. *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor.* St. Martin's Press, Inc., USA.

[35] Susan S. Fainstein. 2014. The just city. *International Journal of Urban Sciences* 18, 1 (2014), 1–18. https://doi.org/10.1080/12265934.2013.834643 arXiv:https://doi.org/10.1080/12265934.2013.834643

[36] Kim Fortun. 2001. *Advocacy after Bhopal: Environmentalism, Disaster, New Global Orders.* University of Chicago Press, Chicago, IL. https://press.uchicago.edu/ucp/books/book/chicago/A/bo3641096.html

[37] Mark Frank and Johanna Walker. 2016. User Centred Methods for Measuring the Value of Open Data. *The Journal of Community Informatics* 12, 2 (June 2016). https://doi.org/10.15353/joci.v12i2.3221 Number: 2.

[38] Peter Frumkin. 2005. *On Being Nonprofit: A Conceptual and Policy Primer.* Harvard University Press, Cambridge, MA.

[39] Julie M. Funk, Matthew Lakier, Marcel O'Gorman, and Daniel Vogel. 2021. Exploring Smartphone Relationships through Roland Barthes using an Instrumented Pillow Technology Probe. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (CHI '21)*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3411764.3445548

[40] Bill Gaver, Tony Dunne, and Elena Pacenti. 1999. Design: Cultural probes. *Interactions* 6, 1 (Jan. 1999), 21–29. https://doi.org/10.1145/291224.291235

[41] Elisabeth Gebka, Jonathan Crusoe, and Karin Ahlin. 2021. Methods to Capture User Information Needs: Design Principles for Open Data Intermediaries and Data Providers. *Data Science Journal* 20, 1 (Dec. 2021), 37. https://doi.org/10.5334/dsj-2021-037

[42] Amy M. Gill and Blair Nonnecke. 2012. Think aloud: effects and validity. In *Proceedings of the 30th ACM international conference on Design of communication*. ACM, Seattle Washington USA, 31–36. https://doi.org/10.1145/2379057.2379065

[43] Lisa Gitelman. 2013. *"Raw Data" Is an Oxymoron.* The MIT Press.

[44] Statistics Canada Government of Canada. 2001. Census of Population. https://www12.statcan.gc.ca/census-recensement/index-eng.cfm Last Modified: 2023-09-12.

[45] Ben Green. 2019. *The smart enough city: putting technology in its place to reclaim our urban future.* The MIT Press, Cambridge, Massachusetts.

[46] Daniel Greene. 2021. *The Promise of Access: Technology, Inequality, and the Political Economy of Hope.* The MIT Press, Cambridge, Massachusetts.

[47] Michael B. Gurstein. 2011. Open data: Empowering the empowered or effective data use for everyone? *First Monday* (Jan. 2011). https://doi.org/10.5210/fm.v16i2.3316

[48] Byung-Chul Han. 2017. *Psychopolitics: Neoliberalism and New Technologies of Power.* Verso, London ; New York.

[49] Donna Haraway. 1988. Situated Knowledges: The Science Question in Feminism and the Privilege of Partial Perspective. *Feminist Studies* 14, 3 (1988), 575–599. https://doi.org/10.2307/3178066 Publisher: Feminist Studies, Inc..

[50] David Harvey. 2013. *Rebel Cities: From the Right to the City to the Urban Revolution* (reprint edition ed.). Verso, London.

[51] Tali Hatuka, Eran Toch, Michael Birnhack, and Hadas Zur. 2020. The Digital City: Critical Dimensions in Implementing the Smart City. https://doi.org/10.2139/ssrn.3766782

[52] Sara Heitlinger, Nick Bryan-Kinns, and Rob Comber. 2019. The Right to the Sustainable Smart City. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. ACM, Glasgow Scotland Uk, 1–13. https://doi.org/10.1145/3290605.3300517

[53] Yen-Chia Hsu, Jennifer Cross, Paul Dille, Michael Tasota, Beatrice Dias, Randy Sargent, Ting-Hao (Kenneth) Huang, and Illah Nourbakhsh. 2019. Smell Pittsburgh: community-empowered mobile smell reporting system. In *Proceedings of the 24th International Conference on Intelligent User Interfaces*. ACM, Marina del Ray California, 65–79. https://doi.org/10.1145/3301275.3302293

[54] Yen-Chia Hsu, Paul Dille, Jennifer Cross, Beatrice Dias, Randy Sargent, and Illah Nourbakhsh. 2018. Community-Empowered Air Quality Monitoring System. arXiv:1804.03293 [cs.HC]

[55] Hilary Hutchinson, Wendy Mackay, Bosse Westerlund, Benjamin B Bederson, Allison Druin, Catherine Plaisant, Michel Beaudouin-Lafon, Stéphane Conversy, Helen Evans, Heiko Hansen, Nicolas Roussel, Björn Eiderbäck, Sinna Lindquist, and Yngve Sundblad. 2003. Technology Probes: Inspiring Design for and with Families. *NEW HORIZONS* 5 (2003).

[56] Rikke Hagensby Jensen, Stine S. Johansen, Nicolai Brodersen Hansen, Ashna Mahmood Zada, and Peter Axel Nielsen. 2023. Tensions in Data Journey Activities: Mobilising, Processing, Producing, and Re-purposing Data in Environmental Assessment Practice. *Proceedings of the ACM on Human-Computer Interaction* 7, CSCW2 (Oct. 2023), 363:1–363:22. https://doi.org/10.1145/3610212

[57] Matthew Jörke, Yasaman S. Sefidgar, Talie Massachi, Jina Suh, and Gonzalo Ramos. 2023. Pearl: A Technology Probe for Machine-Assisted Reflection on Personal Data. In *Proceedings of the 28th International Conference on Intelligent User Interfaces (IUI '23)*. Association for Computing Machinery, New York, NY, USA, 902–918. https:

//doi.org/10.1145/3581641.3584054

[58] Maxat Kassen. 2013. A promising phenomenon of open data: A case study of the Chicago open data project. *Government Information Quarterly* 30, 4 (Oct. 2013), 508–513. https://doi.org/10.1016/j.giq.2013.05.012

[59] Rob Kitchin. 2014. *The Data Revolution: Big Data, Open Data, Data Infrastructures &amp; Their Consequences.* SAGE Publications Ltd. https://doi.org/10.4135/9781473909472

[60] Rob Kitchin, Tracey P. Lauriault, and Gavin McArdle. 2015. Knowing and governing cities through urban indicators, city benchmarking and real-time dashboards. *Regional Studies, Regional Science* 2, 1 (Jan. 2015), 6–28. https://doi.org/10.1080/21681376.2014.983149 Publisher: Routledge _eprint: https://doi.org/10.1080/21681376.2014.983149.

[61] Rob Kitchin and Niamh Moore-Cherry. 2021. Fragmented governance, the urban data ecosystem and smart city-regions: the case of Metropolitan Boston. *Regional Studies* 55, 12 (Dec. 2021), 1913–1923. https://doi.org/10.1080/00343404.2020.1735627 Publisher: RSA Website _eprint: https://doi.org/10.1080/00343404.2020.1735627.

[62] Henri Lefebvre. 1968. *The Right to the City.* https://theanarchistlibrary.org/library/henri-lefebvre-right-to-the-city

[63] Chang Liu, Arif Usta, Jian Zhao, and Semih Salihoglu. 2023. Governor: Turning Open Government Data Portals into Interactive Databases. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI '23).* Association for Computing Machinery, New York, NY, USA, Article 415, 16 pages. https://doi.org/10.1145/3544548.3580868

[64] Yanni Alexander Loukissas. 2019. *All Data Are Local: Thinking Critically in a Data-Driven Society* (1 ed.). MIT Press, Cambridge, MA, USA. https://doi.org/10.7551/mitpress/11543.001.0001

[65] 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.* ACM, Seoul Republic of Korea, 330–333. https://doi.org/10.1145/2591888.2591947

[66] Samuel Mann, Oliver Emile Glaves Bates, and Raymond Maher. 2018. Shifting the maturity needle of ICT for Sustainability, Birgit Penzenstadler, Steve Easterbrook, Colin Venters, and Syed Ishtiaque Ahmed (Eds.). EasyChair, CAN, 209–226. https://doi.org/10.29007/d6g3 Meeting Name: ICT4S 2018 Num Pages: 28.

[67] Ricardo Matheus, José Carlos Vaz, and Manuella Maia Ribeiro. 2014. Open government data and the data usage for improvement of public services in the Rio de Janeiro City. In *Proceedings of the 8th International Conference on Theory and Practice of Electronic Governance (ICEGOV '14).* Association for Computing Machinery, New York, NY, USA, 338–341. https://doi.org/10.1145/2691195.2691240

[68] Shannon Mattern. 2015. History of the Urban Dashboard. https://placesjournal.org/article/mission-control-a-history-of-the-urban-dashboard]/#ref_38

[69] Shannon Mattern. 2017. A City Is Not a Computer. *Places Journal* (Feb. 2017). https://doi.org/10.22269/170207

[70] Curtis W. McCord and Christoph Becker. 2023. Beyond Transactional Democracy: A Study of Civic Tech in Canada. *Proceedings of the ACM on Human-Computer Interaction* 7, CSCW1 (April 2023), 1–37. https://doi.org/10.1145/3579462

[71] Milagros Miceli, Tianling Yang, Laurens Naudts, Martin Schuessler, Diana Serbanescu, and Alex Hanna. 2021. Documenting Computer Vision Datasets: An Invitation to Reflexive Data Practices. In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency (FAccT '21).* Association for Computing Machinery, New York, NY, USA, 161–172. https://doi.org/10.1145/3442188.3445880

[72] Gina Neff, Anissa Tanweer, Brittany Fiore-Gartland, and Laura Osburn. 2017. Critique and Contribute: A Practice-Based Framework for Improving Critical Data Studies and Data Science. *Big Data* 5, 2 (June 2017), 85–97. https://doi.org/10.1089/big.2016.0050

[73] Safiya Umoja Noble. 2018. *Algorithms of Oppression: How Search Engines Reinforce Racism* (illustrated edition ed.). NYU Press, New York.

[74] City of Toronto. 2017. Toronto Open Data. https://www.toronto.ca/city-government/data-research-maps/open-data/ Archive Location: Toronto, Ontario, Canada Publisher: City of Toronto Scroll: yes.

[75] Cathy O'Neil. 2016. *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy.* Crown, New York.

[76] Shreyasha Paudel and Robert Soden. 2023. Reimagining Open Data during Disaster Response: Applying a Feminist Lens to Three Open Data Projects in Post-Earthquake Nepal. *Proceedings of the ACM on Human-Computer Interaction* 7, CSCW1 (April 2023), 1–25. https://doi.org/10.1145/3579519

[77] E. Paulos, R. Honicky, and B. Hooker. 2008. Citizen science: Enabling participatory urbanism. *Handbook of Research on Urban Informatics* (Jan. 2008), 414–436. https://doi.org/10.4018/978-1-60566-152-0.ch028

[78] Rosamund Pearce. 2017. Mapped: How the US generates electricity. https://www.carbonbrief.org/mapped-how-the-us-generates-electricity/

[79] Lucy Pei, Benedict Salazar Olgado, and Roderic Crooks. 2022. Narrativity, Audience, Legitimacy: Data Practices of Community Organizers. In *CHI Conference on Human Factors in Computing Systems Extended Abstracts.* ACM, New Orleans LA USA, 1–6. https://doi.org/10.1145/3491101.3519673

[80] Jennifer Preece, Daniel Pauw, and Tamara Clegg. 2019. Interaction design of community-driven environmental projects (CDEPs): A case study from the Anacostia Watershed. *Proceedings of the National Academy of Sciences* 116, 6 (2019), 1886–1893. https://doi.org/10.1073/pnas.1808635115 arXiv:https://www.pnas.org/doi/pdf/10.1073/pnas.1808635115

[81] Mark Purcell. 2014. Possible Worlds: Henri Lefebvre and the Right to the City. *Journal of Urban Affairs* 36, 1 (Feb. 2014), 141–154. https://doi.org/10.1111/juaf.12034 Publisher: Routledge _eprint: https://doi.org/10.1111/juaf.12034.

[82] Aare Puussaar, Ian G. Johnson, Kyle Montague, Philip James, and Peter Wright. 2018. Making Open Data Work for Civic Advocacy. *Proc. ACM Hum.-Comput. Interact.* 2, CSCW, Article 143 (nov 2018), 20 pages. https://doi.org/10.1145/3274412

[83] SoYon Rim, Jochim Hansen, and Yaacov Trope. 2013. What happens why? Psychological distance and focusing on causes versus consequences of events. 104, 3 (2013), 457–472. https://doi.org/10.1037/a0031024

[84] Lost Rivers. 2024. Lost Rivers of Toronto Map. https://www.lostrivers.ca/disappearing.html

[85] 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 (Dec. 1996), 512–523. https://doi.org/10.1080/01944369608975715

[86] James C. Scott. 1999. *Seeing Like a State : How Certain Schemes to Improve the Human Condition Have Failed - University of Toronto.* Yale University Press. https://librarysearch.library.utoronto.ca/discovery/fulldisplay?docid= alma991106970410406196&context=L&vid=01UTORONTO_INST:UTORONTO&lang=en&search_scope=UTL_ AND_CI&adaptor=Local%20Search%20Engine&tab=Everything&query=any,contains,seeing%20like%20a%20state& offset=0

[87] Cathrine Seidelin, Yvonne Dittrich, and Erik Grönvall. 2020. Co-designing Data Experiments: Domain Experts' Exploration and Experimentation with self-selected Data Sources. In *Proceedings of the 11th Nordic Conference on Human-Computer Interaction: Shaping Experiences, Shaping Society.* ACM, Tallinn Estonia, 1–11. https://doi.org/10.1145/3419249.3420152

[88] Ashraf Shaharudin, Bastiaan van Loenen, and Marijn Janssen. 2023. Towards a Common Definition of Open Data Intermediaries. *Digit. Gov.: Res. Pract.* 4, 2, Article 6 (jun 2023), 21 pages. https://doi.org/10.1145/3585537

[89] Joe Shaw and Mark Graham. 2017. An Informational Right to the City? Code, Content, Control, and the Urbanization of Information. *Antipode* 49, 4 (2017), 907–927. https://doi.org/10.1111/anti.12312 _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/anti.12312.

[90] Katie Shilton, Jes A. Koepfler, and Kenneth R. Fleischmann. 2014. How to see values in social computing: methods for studying values dimensions. In *Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing.* ACM, 426–435. http://dl.acm.org/citation.cfm?id=2531625

[91] Linnet Taylor. 2017. What is data justice? The case for connecting digital rights and freedoms globally. *Big Data & Society* 4, 2 (Dec. 2017), 2053951717736335. https://doi.org/10.1177/2053951717736335 Publisher: SAGE Publications Ltd.

[92] No More Noise Toronto. 2024. No More Noise Toronto. https://www.nomorenoisetoronto.com/

[93] Anh-Ton Tran, Ashley Boone, Christopher A. Le Dantec, and Carl DiSalvo. 2022. Careful Data Tinkering. *Proc. ACM Hum.-Comput. Interact.* 6, CSCW2, Article 431 (nov 2022), 29 pages. https://doi.org/10.1145/3555532

[94] Anh-Ton Tran, Grace Guo, Jordan Taylor, Katsuki Andrew Chan, Elora Lee Raymond, and Carl Disalvo. 2024. Situating Datasets: Making Public Eviction Data Actionable for Housing Justice. In *Proceedings of the CHI Conference on Human Factors in Computing Systems.* ACM, Honolulu HI USA, 1–16. https://doi.org/10.1145/3613904.3642452

[95] Yaacov Trope and Nira Liberman. 2010-04. Construal-Level Theory of Psychological Distance. 117, 2 (2010-04), 440–463. https://doi.org/10.1037/a0018963

[96] Vasillis Vlachokyriakos, Clara Crivellaro, Christopher A. Le Dantec, Eric Gordon, Pete Wright, and Patrick Olivier. 2016. Digital Civics: Citizen Empowerment With and Through Technology. In *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems.* ACM, San Jose California USA, 1096–1099. https://doi.org/10.1145/2851581.2886436

[97] Amy Voida, Ellie Harmon, and Ban Al-Ani. 2011. Homebrew Databases: Complexities of Everyday Information Management in Nonprofit Organizations. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Vancouver, BC, Canada) *(CHI '11).* Association for Computing Machinery, New York, NY, USA, 915–924. https://doi.org/10.1145/1978942.1979078

[98] Dawn Walker and Christoph Becker. 2016. Sustainability Design: Lessons from Designing A "Green Map". Atlantis Press, 167–176. https://doi.org/10.2991/ict4s-16.2016.20 ISSN: 2352-538X.

[99] James Merricks White. 2017. Following data threads. In *Data and the City.* Routledge. Num Pages: 13.

[100] Richmond Y. Wong, Michael A. Madaio, and Nick Merrill. 2023. Seeing Like a Toolkit: How Toolkits Envision the Work of AI Ethics. *Proceedings of the ACM on Human-Computer Interaction* 7, CSCW1 (April 2023), 1–27. https://doi.org/10.1145/3579621

[101] Johanna Ylipulli and João Hämäläinen. 2023. Towards Practice-oriented Framework for Digital Inequality in Smart Cities. In *The 11th International Conference on Communities and Technologies (C&T).* ACM, Lahti Finland, 180–190.

          https://doi.org/10.1145/3593743.3593781

[102] Ayoung Yoon, Andrea Copeland, and Paula Jo McNally. 2018. Empowering communities with data: Role of data
          intermediaries for communities' data utilization. *Proceedings of the Association for Information Science and Technology*
          55, 1 (Jan. 2018), 583–592.   https://doi.org/10.1002/pra2.2018.14505501063