

Storytelling with Data: Examining the Use of Data by Non-Profit Organizations

Sheena Erete[†], Emily Ryou[†], Geoff Smith^{*}, Khristina Fassett^{*}, Sarah Duda^{*}

[†]College of Computing and Digital Media

^{*}Institute for Housing Studies

DePaul University Chicago, IL USA

serete@cdm.depaul.edu, emily.ryou@gmail.com, [gsmith33, kfassett, sduda1]@depaul.edu

ABSTRACT

Despite the growing promotion of the “open data” movement, the collection, cleaning, management, interpretation, and dissemination of open data is laborious and cost intensive, particularly for non-profits with limited resources. In this paper, we describe how non-profit organizations (NPOs) use open data, building on prior literature that focuses on understanding challenges that NPOs face. Based on 15 interviews of staff from 10 NPOs, our results suggest that NPOs use data to develop narratives to build a case for support from grantors and other stakeholders. We then present empirical results based on the usage of a data portal we created, which suggests that technologies should be designed to not only make data accessible, but also to facilitate communication and support relationships between expert data analysts and NPOs.

ACM Classification Keywords

H.5.m. Information Interfaces and Presentation (e.g. HCI): Miscellaneous

Author Keywords

Non-Profit Organizations; NPOs; open data; storytelling

INTRODUCTION

Though the term “open data” is not a new concept, over the past decade national and local government agencies in the US have become more aggressive in their efforts to make data publicly available for citizens to use [19, 21]. We define open data as facts, observations, and statistics collected together for reference and analysis that is publicly available for use. Studies have shown that use of open data has been pivotal in an institution’s execution of their goals [21]. However, while some may praise the US government’s effort in making more data available, there is still concern about the consumption of such data [26, 9, 16] – that is, the collection, cleaning, management, interpretation, and dissemination of data. Data

consumption is laborious and cost intensive, both factors that can make open data inaccessible for non-profit organizations (NPOs) with limited resources [31] and can affect fulfilling their goals (e.g., inciting policy changes, community development, obtaining grants). We argue that by understanding how NPOs use open data, we can design technology solutions that not only provide access to but also easier consumption of open data. Thus, we pose the following research questions: *How are non-profit organizations using open data to achieve their goals? How can we better design technologies to make open data more consumable for organizations with limited resources?*

The objective of this paper is to: (1) describe how non-profit organizations with limited resources use open data in the US and (2) evaluate the use of our data portal by NPOs. Our data portal, which can be publicly accessed through our research center’s website, was also designed to support collaboration between NPOs and our Field Expert Data Analysts (FEDAs). The data portal is an online tool that provides access to aggregated data (i.e., data combined by multiple factors such as geography) and reports created by our FEDAs. FEDAs are trained expert analysts with strong statistics, data mining, and information visualization skills that specialize in a specific field or subject (e.g., education, health, crime). FEDAs provide technical assistance, custom analyses, and reports comprised of data visualizations and in-field analyses to NPOs at no cost. We specifically focus on the relationship between NPOs and FEDAs as much prior HCI research has focused on inter-organizational collaboration amongst NPOs [15, 29], ignoring the potential benefits of collaborative analysis between NPOs and external field experts. Our approach builds on extensive CSCW work on organization knowledge management [3, 5, 28] by extending the context to NPOs.

To answer our research questions, we interviewed staff of various levels at NPOs that advocate for affordable housing locally, regionally, or nationally. NPOs reported using housing data to administer programs, develop policy, and to tell organization and community-specific narratives. NPOs further enrich these narratives with other contextual data that helps connect their story to their target audiences.

We then discuss NPOs’ responses to our data portal. We found that features in our portal that support access to aggregated

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from Permissions@acm.org.

CSCW '16, February 27-March 02, 2016, San Francisco, CA, USA
Copyright © 2016 ACM ISBN 978-1-4503-3592-8/16/02...\$15.00
DOI: <http://dx.doi.org/10.1145/2818048.2820068>

data play a limited role; instead, NPOs rely heavily on their relationships with specialized field experts to satisfy their data needs. Based on our findings, we suggest best practices around designing effective technologies that support NPOs in interpreting, consuming, and communicating open data to address their goals and objectives.

Our research contributions are two-fold. First, this paper provides insight into how NPOs use open data, building on prior literature that focuses on understanding challenges that NPOs encounter [12, 15, 24, 29, 31]. Second, we describe the collaboration practices between NPOs and individuals who are external experts (non-volunteers), extending previous HCI work on inter-organizational collaboration [29, 32] as well as organization knowledge and memory management [5, 28]. Our design implications, which encourage interpretation and consumption of data through collaborative analysis, add to the large body of research on designing technologies for NPOs [8, 12, 11, 15, 17, 20, 22, 23, 24, 29, 31, 32].

In the following sections, we begin by presenting prior literature on technology use by NPOs and general practices of open data. We provide an overview of our study design, including a detailed description of our data portal and the NPOs that we interviewed. We then present our research findings and conclude by describing design implications and opportunities for future work.

RELATED WORK

With the recent push towards open data, NPOs are rapidly gaining access to data that can help them lobby grantors, volunteers, and other government agencies for support. However, NPOs have limited resources to effectively analyze and visualize open data. Research on collaborative analysis suggests benefits to partnering with others to understand open datasets [18, 27, 33]; yet, few have explored how NPOs could take advantage of such techniques. In this section, we describe prior literature around NPOs' Information and Communication Technologies (ICTs) use, open data analysis practices, and collaborative analysis technologies.

Non-Profit Organizations and ICTs

NPOs play a pivotal role in society, economically and socially. For example, a study by the Urban Institute found that 2.3 million NPOs operate in the U.S., contributing over \$800 billion to the economy in 2010 [6]. Over 25% of U.S. citizens volunteered over 15 billion hours in 2011, an estimated worth of \$296 billion [6]. In addition, NPOs advocate for community wellness, encourage social change, and serve the underprovided with critical functions that empower individuals [15]. NPOs support the needs of many communities by compensating for the disparities of goods and services ranging from emergency resources to advocacy for policy change [31].

Despite their large economic and societal role, NPOs have limited resources that originate from painstaking efforts in fundraising and grant writing [24]. Such efforts motivate NPOs to primarily focus on their organizational goals, neglecting important in-house necessities such as technology and staff development [31]. Yet, NPOs still wish to function as efficiently as their for-profit counterparts [29], and ICTs play

an integral role in achieving such goals by supporting staff and intra-organizational structures [15].

Due to budget restrictions, many NPOs face a multitude of challenges regarding the use and adoption of ICTs [11, 15, 20, 24, 29, 31, 32]. Some NPOs rely on inexpensive, yet inadequate, "out-of-the-box" solutions that require much internal maintenance, leaving NPOs with the task of continually patching together and reconfiguring systems [31]. Such maintenance can result in lost, abandoned, or duplicated data [31]. Centralized data provides a means for NPOs to identify trends, which enables them to more effectively focus their campaigns to achieve their goals [8]. Furthermore, deficiencies in internal technologies and differences in technology adoption demotivate NPOs from engaging in inter-organizational collaboration [29]. Thus, lack of resources impact NPOs' current ICT practices and inhibit their ability to collaborate with others to accomplish their goals. This paper presents a novel perspective by describing how NPOs use our data portal to collaborate with external experts to supplement their limited resources.

Open Data Practices and Collaborative Analysis

There has been a surge in the availability of open data in recent years, trending from government agencies mandating the release of more datasets [4]. During its launch in May 2009, data.gov contained 76 datasets from 11 US government agencies. Eleven months later, data.gov contained 1,284 datasets from 170 government and public agencies [19]. The growth in open data has implications across private, public, and not-for-profit sectors, acting as a catalyst for innovation by stimulating potential for funding, providing information about markets, and inspiring development beyond applications [21].

With data becoming more available, scholars have begun to investigate the consumption practices of open data [4, 9, 14, 16, 19, 21, 26]. There have been numerous studies that suggest that availability of data does not ensure accessibility and equity in data consumption, which could empower those who have been traditionally marginalized [16, 26]. In addition, some have found that raw data (i.e., data presented in absolute number for the purpose of re-use [9]) is preferred over manufactured reports [19]. Davies & Frank, however, suggest that access to raw data does not necessitate proper reuse of data [9]. Instead, they found that repurposing data successfully requires creating context for users, which could stem from an understanding of how the data was created and will be used [9]. This suggests a complex relationship amongst those who create open data, those who use open data, and the technologies that support access to open data. Though relationships between data experts and consumers may shape the use of open data, few have examined the extent to which these relationships may impact data consumption.

Given the increased access to data, there has been more focus on effective data analysis techniques. One that has gained much attention is collaborative analysis, where data consumers (experts and non-experts) work together to analyze, interpret, and visualize data [9, 15, 18, 31]. The production of these analyses requires communication between users to assist in decision-making by "distributing information across group



Figure 1. Sample of Data Portal filters with populated viewing area

members" [18]. Organizing information can help users collaboratively analyze data, which can be done through tags and links [33]. Collaborative analysis technologies, which support such communication, have also become effective tools to assist users in understanding and discussing their findings [18, 27, 33]. Furthermore, collaborative help-seeking amongst users and experts have been found to increase organizational memory [1, 2], the sharing of organizational knowledge amongst its members [5]. Given the benefits of collaborative analysis technologies and the need for NPOs to find inexpensive, yet accurate, data analysis solutions, there is an opportunity to explore how NPOs can partner with others to assist in data analysis.

NPOs often lack the financial means, time, or technical skills to easily analyze and consume open data; however, NPOs are savvy in finding ways to support their need for data despite how time-consuming and inconvenient these methods may be. Use of collaborative analysis technologies on open data may be an opportunity for NPOs to use fewer resources when attempting to achieve their goals around open data consumption. Our study addresses this gap by describing how NPOs use data and collaborative ICTs that effectively support data interpretation and consumption by NPOs.

DESIGN STUDY

We examined the following research questions: (1) How are non-profit organizations using open data? (2) How can we better design technologies to make open data more consumable for organizations with limited resources? To address these questions, we established a cross-disciplinary team of researchers, graduate students, and staff of our research center, which is housed at a large private university in the Midwest. We conducted semi-structured interviews with staff at 10 NPOs. During interviews, we asked participants to reflect on their current data use and their experience using our existing data portal. We then analyzed the interview transcripts using inductive qualitative analysis techniques.

Research Center Background

The Institute for Housing Studies (IHS) is a research center conceived through a grant from the John D. and Catherine T. MacArthur Foundation. The mission of the research center is to provide reliable, recent, unbiased data to NPOs so that

they can make informed decisions regarding housing policy. IHS focuses explicitly on providing data that informs issues including, but not limited to availability and preservation of affordable housing, housing development finance, community development, and emerging trends in neighborhood housing markets. In addition to property-level data, the research center also collects related data such as U.S. census data, Home Mortgage Disclosure Act data, crime data, and USPS vacancy data.

Data Portal Background and Reports

The initial features of the data portal were based on the experiences of the IHS staff and feedback from NPOs. In May 2013, nine months prior to our study, our team deployed an online data portal that can be accessed freely from the research center’s website. Both the IHS website and the data portal were built by an outside vendor using a Django framework. The research center’s staff updates the data manually via a content management system (CMS). Of the 16 aggregated datasets, 15 are updated yearly and one is updated quarterly. Prior to our data portal, local housing NPOs would often rely on national research centers.

Staff of the research center, which included Field Expert Data Analysts (FEDAs), developers, and community development liaisons, led the development of the portal. The data portal provides aggregated data through the use of three filters: data type, geographic level, and year range. The data type filter includes housing stock composition, property sales activity, mortgage activity, foreclosure filing activity, foreclosure auctions, and long-term residential vacancy rates. The geographic levels comprise of county region, municipality, community area, and ward. The corresponding viewing area presents the data as a table. Figure 1 shows an image of the filters and populated viewing area.

In addition, our research center provides access to reports created by FEDAs, who continually research and analyze the local housing market. Reports, which are created quarterly at a minimum, are located outside the data portal but still within the research center’s website. Each report is unique, containing in-depth analyses, synopsis of the subject, data sources, data visualizations, and/or key findings. The format of the reports varies from academic articles and white papers to more applied reports such as policy briefs and fact sheets.

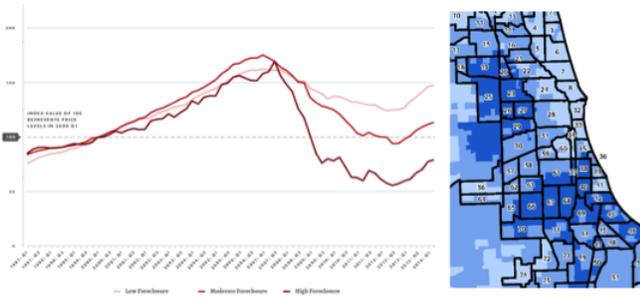


Figure 2. Sample Data Visualization from Reports

Figure 2 shows an example data visualization from a report, which can be downloaded as a PDF from the website.

For further assistance or inquiries, the website prompts users to reach out to the research center by providing contact information for a FEDA.

Field Expert Data Analysts (FEDAs)

Field Expert Data Analysts (FEDAs) are staff members at the research center that perform in-depth analyses of housing market trends in the city and surrounding municipalities. During this study, there were six full-time FEDAs, all compensated through the same grant as the research center. Four of the FEDAs have Masters degrees in Urban Planning and Policy, one in Geography, and one a Ph.D. in Economics. All the FEDAs’ training emphasized housing studies, which included statistics, data mining, and data visualizations (e.g., mapping). With professional experience as local housing experts that range from five to seventeen years, our FEDAs are well respected amongst housing organizations across the city. Furthermore, their reports have received national attention from media outlets such as the *New York Times*, *Chicago Tribune*, *Wall Street Journal*, and *Washington Post*. In addition to generating the datasets, conducting in-depth analyses, and creating reports of local housing trends, FEDAs complement the data portal by collaborating with NPOs to provide custom datasets for their target area and valuable market insights. Before the availability of FEDAs at our research center, NPOs often requested assistance from national organizations that may have lacked local experts.

Study Participants

We interviewed 15 staff members from 10 NPOs that used our data portal, which is free. Though the shared mission between all the NPOs was to promote affordable housing, the size and structure of the organizations as well as the method of achieving the mission differed. We first identified 10 NPOs varying in size, structure, and area coverage. The NPOs yearly budgets ranged from \$148,000 to over \$293,000,000, where the median was \$2,800,159. Despite the budget, all the NPOs reported a lack of resource allocation towards data analysis. Five NPOs were region-based, and five were community-based. Region-based organizations covered an area that reached beyond the metropolitan area, while community-based organizations focused on neighborhoods in the city. Table 1 summarizes the participants and their NPOs.

NPO	Area Coverage	Focus of Work	Participation Code	Individual Role
1	Region	Community Development	R1	Senior Project Manager
			R2	Data Analyst
2	Region	Community Development	R3	Director
3	Region	Community Development	R4	Project Manager
4	Region	Housing	R5	Data Analyst
			R6	Director
5	Region	Housing	R7	Program Officer
			R8	Director
6	Community	Housing	C1	Director
7	Community	Community Development	C2	Senior Organizer
8	Community	Housing	C3	Executive Director
			C4	Field Manager
9	Community	Housing	C5	Field Manager
			C6	Director
			C7	Vice President
10	Community	Community Development		

Table 1. Overview of Participating Individuals and NPOs

Region-based Organizations

Region-based organizations collaborate with other NPOs, local governments, businesses, and community leaders to explore community development, policy, and financing issues. Two region-based organizations (NPO 4 and 5) interact directly with residents and focus only on housing-related issues. The other three focus on public policy issues and overall community development. NPO 1, 4, and 5 support development by providing financial lending services to potential residents and other NPOs.

Community-based Organizations

Community-based organizations interact and develop relationships with individual residents in a targeted neighborhood. Two community-based organizations (NPO 6 and 7) accomplish their goals by organizing community leaders and residents; the other three establish programs, secure grants, and execute strategic plans. NPO 8, 9, and 10 focus solely on affordable, subsidized, or integrated housing, while the other two community-based organizations focus on overall community development.

To recruit participants, we contacted the NPOs and asked to interview people who worked directly with data or made decisions regarding how data is used in the organization. Thus, our participants had varying hierarchical positions in the organization. We interviewed participants with varying positions to not only understand how data is manipulated by those who work with data on a regular basis but also organizational strategy and rationale of data use decisions.

Seven participants were higher-level, and eight had associate to mid-level positions. We identified higher-level positions as those whose role in the organization was strategic decision-making (e.g., grant seeking, deciding on future directions for organization). Higher-level employees did not work directly with data but instead were in a position to ask someone else to gather and/or analyze certain data. High-level positions included director, executive director, and Vice President. Associate to mid-level positions were staff that were more likely to have direct contact with data on a fairly consistent basis. Such

positions included project managers, data analysts, program officers, senior organizers, and field managers. Participants' data analysis skills varied, with 12 identifying themselves as amateur and five as experts (R2, R5, C1, C3, and C7).

Participants' ages ranged from 25 to 45 years old with a median age of 34. Four participants were male (36%), and eleven were female (64%). Eleven participants were Caucasian (73%), three Latino (20%), and one African-American (7%). The highest level of education for six participants (40%) was a Bachelors degree, and nine participants (60%) had a Masters degree.

Data Collection

We conducted 15 semi-structured interviews, where we asked about: interviewee background information and role in the organization, data usage and practices, data portal usage, and technologies used at work. We created two interview scripts to fit the interviewee's profile in regards to their position level and degree of contact with data. The major differences between the two interview scripts were the level of depth around particular topics. Specifically, we asked high-level staff questions regarding how data fits into the long-term organizational objectives while reserving questions about detailed data manipulation practices to those who consistently have direct contact with data. At the end of each interview, we administered a survey to understand work-related technology use and to gather demographic information.

Data Analysis

We qualitatively analyzed the interviews by developing a codebook with 41 codes initially. After achieving an inter-rater reliability score of 95%, two of the authors applied the codes line-by-line on the interview transcripts. We then iterated on the codebook by consolidating overlapping codes, reducing the number of codes to 31. Examples of codes are organizational mission, data availability, data format, and analysis techniques. We then grouped the codes into 4 categories: data question types, portal use, challenges and NPO profiles. During our final iteration of analysis, we identified unique differences in how organizations use data to develop narratives, differences in the types of data questions, and salient challenges. We also identified design opportunities based on data portal feedback and data needs.

FINDINGS

Staff from NPOs reported that they use open data to narrate stories to donors and stakeholders. Our findings suggest that NPOs would benefit from technologies that facilitate the opportunity for collaborative analysis of open data.

Storytelling with Data

NPOs primarily use data to narrate local conditions to seek support for their goals. Data provides "*ways for information to come alive*" (C1) and establishes evidence for future strategic planning. NPOs use storytelling to make data consumable for their target audience, which can include internal and external stakeholders, government officials, funders, financial investors, residents, and community leaders. Using Boje's

study of organizations using storytelling, we define to storytelling as "how people [or organizations] perform stories to make sense of events, introduce change, and gain political advantage" [7]. We recognize that there are numerous definitions of stories and narratives but in this paper, we use the terms interchangeably to refer to the creation of "part of the story line, a bit of interpretation, and some facts that confirm a line of reasoning" [7]. There has been much work on organizations using storytelling as a mechanism for organizational change by gaining support from stakeholders, sharing facts and knowledge, and co-creating narratives that inform future direction of the organizations [10]. Our results suggest that NPOs use data to narrate their reports, presentations, and data visualizations. R2, a data analyst in a region-based organization, describes how her NPO communicates their stories:

"In presentations [...] I'm not up there just pointing to numbers but translating that into more of a story or an explanation of methodology."

Storytelling may differ depending on the objective (i.e., grant applications, internal benchmark evaluations, or for local government officials), which affects the process in which NPOs utilize data. This aligns with Eisenhardt who suggest that while "[g]ood storytelling may make studies entertaining [...], their theoretical impact comes from rigorous method and multiple-case comparative logic" [13]. The remainder of this section describes two types of stories that NPOs tell: *heterogeneous* and *homogeneous* data stories. We conclude by describing how NPOs incorporate context to strengthen their stories.

Heterogeneous Data Stories

Heterogeneous data stories amalgamate multiple data to tell one composite story that depicts current conditions of a specific area. R6, a director at a region-based organization, describes an example:

"When we talk about telling our story, I think we also use broader economic data, both population changes, not only in the city and the suburbs, but within specific community areas, racial, ethnic change within a particular area, income information, unemployment information."

Heterogeneous data stories often try to answer questions such as, "How has our organization affected foreclosure rates in our targeted neighborhood?" or "What policy changes will affect development?" NPOs tell heterogeneous data stories by compiling complex data. These are custom built data, derived from multiple sources and datasets (such as housing, crime, demographics, etc.) and can span across large geographies and time frames. We found that in-house data analysts rarely compile these complex data; instead, NPOs rely on relationships with external data analysts (such as our FEDAs) to obtain and interpret these datasets.

To tell heterogeneous data stories, NPOs often need to undergo a complicated process to identify the data needs. NPOs without a data analyst rely on relationships with FEDAs for this process. In-house data analysts, such as R2, use their own unique, complicated analysis process to identify key data components:

“This department was just trying to get an aerial snapshot of what’s happening there. “What” and “what” kinds of issues are overlapping with each other? The way I went about that wasn’t exactly to figure out [the] research question, but to compartmentalize issues that I thought went together.”

Homogeneous Data Stories

Homogeneous data stories consist of one dataset that can tell multiple stories. NPOs may manipulate one specific dataset to tell various stories. Our analysis suggests two types of homogenous data stories: parcel and trend based.

NPOs tell parcel-based homogeneous data stories using parcel level data – that is, information about individual units or buildings. Often, these stories try to answer questions such as “Is this unit vacant?” The purpose of these stories is to help identify the status of specific units or buildings to, ultimately, aid in administering and identifying new program opportunities (e.g. foreclosure prevention). Parcel-level questions could also be applied to other topics; in health, for example, a parcel-level question may translate into a question about whether a household has been exposed to a particular disease. In other words, parcel-level questions are very low-level questions that may be important for “on-the-ground” program implementation.

Parcel-based homogeneous data stories pertain to small geographic areas (sometimes as small as one square mile), over short time periods – such as weeks. C7, from a community-based organization, describes an example of using parcel level data to tell a homogeneous data story:

“We track losses of Single Room Occupancies because when we go to alderman, we want to say, ‘Look, this is happening really fast, and we need an ordinance to prevent this bleeding from continuing in our neighborhoods.’ And that’s a really stark number that people respond to.”

Packaged datasets (e.g., existing data cleaned and ready for use) may not exist to tell this type of parcel-based story, leaving NPOs to manually obtain this data through fieldwork (e.g., physically visiting buildings and documenting the conditions), calling landlords or tenants, or performing Internet searches. These homogeneous data stories require very detailed data, which most open datasets do not provide.

Trend-based homogeneous data stories use aggregated datasets to help NPOs understand patterns and trends that communicate changes in neighborhood market conditions over a designated period of time – retrospectively and prospectively. These types of homogeneous data stories may answer questions such as, “In the past year, have foreclosures increased or decreased in this neighborhood?” This data relates to larger geographical areas (e.g., neighborhoods, municipalities) and over annual time periods.

These stories help measure and narrate organizational progress, create benchmarks, and/or programmatic goals and evaluations. For example, C3’s organization tells a story about their progress in reducing racially segregated housing by *“look[ing] at the change from this census to that census [to see if] the integration of [this community] has improved.”*

Data Context to Strengthen Stories

NPOs use data to describe a specific context to forge a stronger connection with their targeted audience. Adding context to narratives can be achieved via geographical or statistical data. NPOs typically decide between using aggregated (e.g., percentages) or raw data (e.g. data before it is processed into percentages). C7 articulates the importance of these contexts:

“When you go meet with the legislator for policy, if you show up with info about the city or state as a whole, and they’re concerned about their district, it’s time wasted. [But] if you can walk in there and say that ‘Out of the 10,000 housing units in your community, 20% are distressed – that’s why we’re trying to get these resources from this particular source to do this,’ you make a real connection for them in what’s happening right now in their area.”

To personalize their narratives, NPOs may pair qualitative data with statistical facts. Personal stories create richer narratives to increase the impact of the organization’s work. When discussing the use of residents’ stories in her presentations to donors and policy makers, C1 states:

“I think we do our most powerful work where the human voice is there and then the data backs up the experiential story.”

To summarize, NPOs use open data to communicate two types of stories to achieve their goal. NPOs reported that they used both quantitative and qualitative data to describe the context to strengthen their connections with their targeted audience.

Data Portal Experience

To answer our second research question, we explored the use of our data portal. Our data portal was designed to provide free access to housing data in varying formats (e.g., aggregated data, reports). As a complement to the portal, users also had the ability to consult with FEDAs. In the following sections, we discuss NPOs’ experiences using our data portal, focusing on the three themes that emerged. Specifically we describe how NPOs (1) access data and face challenges, (2) utilize reports that we generated, and (3) build relationships with FEDAs.

Data Access and Challenges

The aggregated datasets available on our data portal help NPOs quickly gather facts, make decisions, and substantiate their suppositions about targeted areas. Participants transform the aggregated data into more consumable formats (e.g., fact sheets or high-level overviews) for higher-level colleagues, such as organizational directors. Upon receiving the transformed data, higher-level colleagues use the data for storytelling, targeting specific audiences (e.g., policy maker, grant organizations).

While all the NPOs appreciated the type of data available through our portal, not all were pleased with the format of the data. We provide aggregated data in percentages, a design decision we based on previous studies indicating that raw data may not be as helpful to these organizations [9]. However, we found that when NPOs need a deeper comparison, they prefer to do quick calculations in spreadsheets (e.g., counts).

Though our data portal allows users to download .csv files, five (R2, R5, C1, C3, and C7) participants expressed frustrations because of the difficulties of doing calculations on percentages rather than absolute numbers.

In fact, we found that NPOs generally have challenges accessing the data in absolute numbers. Absolute numbers facilitate the ability to freely manipulate the dataset, making the data reusable for other purposes. The lack of absolute numbers limits the ability to repurpose data for homogeneous data stories. At times, NPOs may only acquire datasets in percentages, which removes the context of the data, especially when comparing across geographies. Staff such as C3 prefers both formats, explaining:

"I always want both – but if I had to choose, I'd rather have raw. [The data] doesn't tell you that [one community] has 90,000 people and other communities have 30,000."

Our data portal provides aggregated data that is associated with specific geographical units (e.g., neighborhoods, wards, counties, municipalities). NPOs require varying levels of geocoded data to accurately compare matching data to produce neighborhood trajectories, comparisons, and heterogeneous and homogeneous data stories. Unavailable geocoded data affects the story being told. NPOs compromise by using related, available data and then modify their message accordingly. R5 recounts:

"If it's not at the level we want it, we have to help interpret it or come up with a way of talking about it that is useful in our geography. [...] It's almost like crafting your message."

Accessing raw data can be challenging for NPOs. Packaged geocoded raw data may be especially difficult to obtain, which may pertain to privacy issues – which is the case with health and crime data. Also, there may be restrictions regarding which formats the data can be accessed due to data subscription agreements (a challenge we also face). Subscription-based data sources refer to data that may have been initially open, but after it is processed and cleaned, it is now licensed (free or for a cost). The license many times limits the format of delivered data. The subscription model of data access can be an issue with data in numerous topics.

NPOs struggle with identifying reliable and useful open data sources and often expend excess time and labor in doing so. The lack of a centralized retrieval hub led our interviewees to visit multiple sources, requiring arduous tracking of their many options. Meanwhile, newer sources were obscure. R5 states:

"Another challenge would be keeping up-to-date with what [sources are] out there and available. There are so many different sources. I know there are different things I could be using, but I don't know about them. There are always new resources that are coming out. You don't know where to go and what's available."

Subscription-based obstacles plague both region and community-based NPOs, though more commonly the latter

due to little financial resources. Participants reported that they sometimes "borrow" subscriptions from colleagues at other organizations. NPOs may also bypass subscription-based sources through intricate and time-consuming processes. R2 explains neutralizing this issue by securing the data through collaborations with our FEDAs:

"The process is to look at your data portal, looking at that overview that was published a while ago, trying to get an understanding of what they have and what their sources are, and then deciding if it's hard to get that data. They calculated this data with the County Assessor's, which we don't have that data and we're not in the position to purchase it. So then I reach out [to your field expert data analyst]."

Reports Generated by Field Expert Data Analyst

Our FEDAs generated reports containing data overviews. We found that NPOs more frequently chose to use our reports over the available aggregated data due to time, labor efficiency, ease of use, and overall convenience.

Our reports satisfy multiple data needs. We discovered that our reports could resolve issues regarding costly subscriptions by providing data otherwise unavailable to them. Additionally, NPOs rely on the advanced analyses and findings, which allow them to select applicable pre-processed data points to insert into their local narrative. NPOs stated that our reports inform their current and future work by apprising them of current housing issues.

In addition to resolving some data challenges, our reports help with data consumption through examples, such as sample data visualizations. For instance, R2, a data analyst at a region-based organization, stated:

"We looked through a lot of data points that were presented in the report that you created. It was for the city, but we wanted to recreate that. [...] We liked a lot of those charts because you can take a lot away really quickly. So we replicated those by getting the data from [your FEDAs] and then also putting in some of our own."

Contrarily, we found that the reports may be overwhelming; in fact, four participants (R3, R5, R6, and C2) expressed difficulty understanding the content. Participants that lacked technical data analyst skills considered the visualizations unsuited for their needs and were unable to repurpose the visualizations accordingly. In fact, participants at NPO 5 relied entirely on our FEDAs to create custom reports and visualizations (e.g., charts and graphs). When asked how we can improve the reports, R8 – a director at NPO 5 – did not have the technical understanding to provide suggestions. This leads to interesting questions about the role of in-house analysts, whose skill may not result in-depth statistical calculations but whose familiarity with data may allow them to easily translate analyses and apply them to their NPO's narrative.

Building Relationships with Field Expert Data Analysts

Though access to our aggregated data and generated reports help NPOs meet their data needs, it also initiates communication with our FEDAs. Through this communication, NPOs

begin to build meaningful relationships with FEDAs. FEDAs collaborate with NPOs throughout the use of open data through personalized services (which include personal consultations), technical support, cultivating in-field awareness, and custom datasets to craft heterogeneous data stories.

Relationships with FEDAs bridge the gap between NPOs' limited resources [15, 29, 31, 30, 32] and their technical data skills [9] by providing a means of collaborative analysis. We define collaborative analysis as the symbiotic relationship between NPOs and FEDAs, where FEDAs typically provide technical support (e.g., detailed analysis, context for data) while NPOs provide insight given their local knowledge that allows for a more sophisticated, informed interpretation of results that extends beyond simple quantitative analysis. During one-on-one consultations, NPOs brief FEDAs on their goals. FEDAs then use their expertise to devise a plan and/or solution. During the analysis, FEDAs and NPOs deliberate on the interpretation of the data by answering questions such as "What do the results mean for this community area?" Given their perceptive understanding of the local area that they service, NPOs' interpretation of the quantitative analysis is vital.

NPOs with and without in-house data analysts benefit from these collaborations. FEDAs collaborate with NPOs without in-house data analysts by providing data converted into digestible visualizations, walkthrough analyses, and unknown data sources. Our FEDAs helped R3 achieve her goal by connecting her to different sources and providing helpful situational information around her issue:

"We wanted to encourage the communities to do more than just look at a chart for all 42 towns and only look at their town. We want to have them think about what's a collective issue. [...] It was honestly [your FEDAs] giving us a great chart with the data sources – 'here's the issue and limitations you may experience, and here's a link to the direct data to the portal.' She made it really easy for us."

Additionally, results suggest that relationships with FEDAs support the work of in-house data analysts by disseminating vital in-field expertise. R5, a data analyst, may not need technical assistance, but still collaborates with our FEDAs for market specific issues:

"We've started to build relationships [...] and have point people that we can call for advice and help for things that are beyond understanding or scope. This happens a lot. [...] [The FEDAs] helped us with neighborhood data and questions with that."

Participants in higher positions (e.g., directors, vice presidents) also work with our FEDAs to build awareness for specific in-field concerns. High-level staff trusts FEDAs to define and clarify issues that affect the NPO's overall goals. Such collaborations help define data issues to develop large-scale goals and underlying issues. For example, R8 explains:

"[The FEDAs] helped us sort [this issue] in however we want to cut that. What should our goal be? [...] We don't really know what types of buildings need a retrofit. He

helped us figure that out. Then he asked us questions – 'What do you think are the issues that would matter here?' That has been really helpful."

Above all, NPO and FEDAs collaborations maintain a symbiotic nature. Inquiries from NPOs in return help inform the FEDAs' research, which ultimately result in new data, analysis, or reports. R8 says:

"We're always the ones asking the questions. [...] Once they have the issue, they take it in some direction."

We observed a collaborative relationship between NPOs and FEDAs as opposed to a one-sided consultancy given that from the NPOs, FEDAs can acquire access to a new, previously unavailable dataset and establish a new data source for other NPOs. For example, FEDAs may rely on NPOs for parcel level datasets – that is, data concerning the status of individual units or buildings. FEDAs do not manually collect data at this level (e.g., visual fieldwork) and depend on other NPOs to create and then deliver parcel level datasets. In one interview, C2, a community organizer from a community-based NPO, described a unique, detailed parcel level dataset created by her organization. Upon learning about this dataset during a debriefing of the interview, our FEDAs stated:

"This dataset you've just described – it's a completely new dataset. I don't think this exists in any other organization in the city [...] but I get calls for this all the time."

Consequently, our study suggests that the benefits of our data portal extend beyond the use of aggregated data and generated reports. We found that NPOs prefer data delivered through our reports; more importantly, the relationships built with our FEDAs increase data consumption and, thus, supports the missions of NPOs.

DISCUSSION

Our findings suggest that the relationship between NPOs and FEDAs are critical to the use, consumption, and interpretation of open data despite other features of our portal. The notion of user and expert knowledge sharing has been thoroughly studied [3, 5, 28]. For example, Ackerman and colleagues' work on Answer Garden, a system that allowed users and experts within an organization to share knowledge, illustrates the benefits of computer-mediated communication amongst users with varying expertise. Our study extends this literature by illustrating similar results in a different context – organizations that are not-for-profit. NPOs may face unique challenges (as compared to other organizations) such as limited budgets, temporary unpaid staff (volunteers), and prosocial objectives, which could influence how the organization uses data. Future work may explore organizational factors (e.g., goals, budget, strategy, culture) that may impact how various organizations use open data and engage in collaborative analysis.

Other studies regarding organizational memory demonstrate the importance of designing effective technologies that support information sharing practices in organizations [5, 28]. Our findings suggest that not only are experts beneficial to users [3, 2], but that the relationship between the two can be mutu-

ally beneficial. Thus, technologies that support collaborative analysis should support relationship building and knowledge sharing symbiosis between users and experts. Future work may explore characteristics of relationships (e.g., trust) between FEDAs and NPOs and how such factors may influence decision-making around data usage. Furthermore, our study was unique in that our experts were subject-matter experts as well as data analysts. Prior literature suggests that these are two distinct competencies that have specific outcomes [25]. Future work could examine the impact of various skill levels amongst experts on how NPOs' choose to adapt data analysis and build relationships with experts.

Given the significant relationship between FEDAs and NPOs in our study, we propose four design implications that target collaboration: (1) browsing datasets; (2) recommending content; (3) nurturing relationships with FEDAs and other NPOs; and (4) developing custom data visualizations.

Browsing Datasets

We found that NPOs face difficulties gaining access to open data and must scour through different sources to identify new datasets. Our data portal also limits users to only our aggregated datasets. To address this challenge, we suggest creating a centralized database hub for other external datasets. To connect sources to the users [9], NPOs could enrich the database content by uploading and sharing useful datasets – a feature gaining popularity among web-based platforms such IBM's cityforward.org. Users could discover datasets by browsing the contents through different filter, search, and tagging mechanisms to promote collaborative analysis [33].

Recommending Content

Results from our study suggest that when seeking assistance, NPOs rely on and trust recommendations from FEDAs to use open data, ranging from specific datasets to visualizations. The current design of our data portal does not include recommendations. Instead, FEDAs provide recommendations upon request, which may not be a sustainable model. To help FEDA services become more scalable, future work could examine the impact of developing a system that recommends content (i.e., data, analysis techniques, appropriate visualizations) to users based on their stated needs. These recommendations could be based on previous suggestions made by FEDAs to NPOs or users with similar profiles. We found that FEDAs provide unique solutions (such as creating customized complex datasets) that reflect the NPO's data questions and goals. A profile comprised of such elements would provide parameters for developing an automated system. In addition to recommended content, FEDAs would be able to make personalized suggestions directly to specific NPOs as needed.

Nurturing Relationships with FEDAs and Other NPOs

Our study suggests that FEDAs perform a crucial role in consuming data. NPOs often emphasized that their organizations are not research facilities. In fact, R1 proclaimed, *"I'm not a researcher; this might not be appropriate but sometimes I don't know where the data is or where to start."* Eight out of ten NPOs maintained a direct relationship with our FEDA to consult on housing issues and data questions. Our findings suggest

that communicating with FEDAs play a pivotal role in data consumption. To build on this relationship, we suggest future iterations of our portal to include an online discussion forum. An online discussion forum would give NPOs the ability to better collaborate with FEDAs, devise solutions, and more easily be directed towards different sources [1]. Moreover, it would allow NPOs and research centers to communicate and share information, which would not only strengthen relationships between NPOs and FEDAs, but also inter-organizational communication between NPOs [2, 32]. Given that NPOs sometimes learn from example analyses (R2), online discussions would allow NPOs to share past experiences, best practices, data processes, and stories amongst each other.

Developing Custom Data Visualizations

Seven out of ten NPOs used data visualizations from our reports to communicate their narratives and to advance their goals. Our portal does not currently support the ability to create custom visualizations. Future work could examine the benefits of collaborative visual data analysis tools, which could provide a means for groups to explore data by manipulating data visualizations. Such tools help users more easily consume data by highlighting insights and important findings [27], which could be particularly useful when creating homogeneous data stories. Furthermore, visual analysis may allow NPOs and FEDAs to collectively analyze data and create solutions [18].

CONCLUSION

This paper moves HCI designers and researchers beyond creating open data tools that are simply clearinghouses, databases, or storage facilities where NPOs with limited resources are forced to scour through massive amounts of data. Instead, our findings suggest that technologies should be designed to not only provide data in consumable ways, but also facilitate communication and relationship building between FEDAs and similar NPOs. Though our study focused solely on open data practices of NPOs with housing missions, many of our insights may be applicable to NPOs that focus on other topics and areas. Future work may explore open data use among NPOs with varying missions – such as sustainability, health, and crime data. Such work could begin to unpack the similarities and differences of NPOs' open data consumption practices across topics.

ACKNOWLEDGMENTS

We would like to thank our study participants and the John D. and Catherine T. MacArthur Foundation for their support of the Institute for Housing Studies. We would also like to thank the reviewers for their thoughtful comments.

REFERENCES

1. M. S. Ackerman and T. W. Malone. 1990. Answer Garden: A Tool for Growing Organizational Memory. In *Proc. of COCS (COCS '90)*. ACM, New York, NY, USA, 31–39. DOI:<http://dx.doi.org/10.1145/91474.91485>
2. Mark S. Ackerman and David W. McDonald. 1996. Answer Garden 2: Merging Organizational Memory with Collaborative Help. In *Proc. of CSCW (CSCW '96)*.

- ACM, New York, NY, USA, 97–105. DOI: <http://dx.doi.org/10.1145/240080.240203>
3. Mark S. Ackerman, Volker Wulf, and Volkmar Pipek. 2002. *Sharing Expertise: Beyond Knowledge Management*. MIT Press, Cambridge, MA, USA.
 4. Charalampos Alexopoulos, Lefkothea Spiliotopoulou, and Yannis Charalabidis. 2013. Open Data Movement in Greece: A Case Study on Open Government Data Sources. In *Proc. of PCI (PCI '13)*. ACM, New York, NY, USA, 279–286. DOI: <http://dx.doi.org/10.1145/2491845.2491876>
 5. Liam J. Bannon and Kari Kuutti. 1996. Shifting Perspectives on Organizational Memory: From Storage to Active Remembering. In *Proc. of HICSS (HICSS '96)*. IEEE Computer Society, Washington, DC, USA, 156. <http://dl.acm.org/citation.cfm?id=795700.798548>
 6. Amy Blackwood, Katie Roeger, and Sarah Pettijohn. 2012. *The nonprofit sector in brief: Public charities, giving and volunteering, 2012*. Technical Report. The Urban Institute, Washington DC.
 7. David M. Boje. 1991. The Storytelling Organization: A Study of Story Performance in an Office- Supply Firm. *Administrative Science Quarterly* 36, 1 (1991), pp. 106–126. <http://www.jstor.org/stable/2393432>
 8. Eleanor Burt and John A. Taylor. 2000. Information and Communication Technologies: Reshaping Voluntary Organizations. *Nonprofit Management and Leadership* 11, 2 (2000), 131–143. DOI: <http://dx.doi.org/10.1002/nml.11201>
 9. Tim Davies and Mark Frank. 2013. 'There's No Such Thing As Raw Data': Exploring the Socio-technical Life of a Government Dataset. In *Proc. of WebSci (WebSci '13)*. ACM, New York, NY, USA, 75–78. DOI: <http://dx.doi.org/10.1145/2464464.2464472>
 10. Robert F. Dennehy. 2001. The Springboard: How Storytelling Ignites Action in Knowledge-era Organizations. *Journal of Organizational Change Management* 14, 6 (2001), 609–614. DOI: <http://dx.doi.org/10.1108/jocm.2001.14.6.609.2>
 11. Lynn Dombrowski, Jed R. Brubaker, Sen H. Hirano, Melissa Mazmanian, and Gillian R. Hayes. 2013. It Takes a Network to Get Dinner: Designing Location-based Systems to Address Local Food Needs. In *Proc. of UbiComp (UbiComp '13)*. ACM, New York, NY, USA, 519–528. DOI: <http://dx.doi.org/10.1145/2493432.2493493>
 12. Lynn Dombrowski, Amy Volda, Gillian R. Hayes, and Melissa Mazmanian. 2012. The Labor Practices of Service Mediation: A Study of the Work Practices of Food Assistance Outreach. In *Proc. of CHI (CHI '12)*. ACM, New York, NY, USA, 1977–1986. DOI: <http://dx.doi.org/10.1145/2207676.2208342>
 13. Kathleen M. Eisenhardt. 1991. Better Stories and Better Constructs: The Case for Rigor and Comparative Logic. *Academy of Management Review* 16, 3 (1991), 620–627. DOI: <http://dx.doi.org/10.5465/AMR.1991.4279496>
 14. Lisa Gitelman. 2013. *Raw data is an oxymoron*. MIT Press.
 15. Jeremy Goecks, Amy Volda, Stephen Volda, and Elizabeth D. Mynatt. 2008. Charitable Technologies: Opportunities for Collaborative Computing in Nonprofit Fundraising. In *Proc. of CSCW (CSCW '08)*. ACM, New York, NY, USA, 689–698. DOI: <http://dx.doi.org/10.1145/1460563.1460669>
 16. Michael B Gurstein. 2011. Open data: Empowering the empowered or effective data use for everyone? *First Monday* 16, 2 (2011).
 17. Darrene Hackler and Gregory D. Saxton. 2007. The Strategic Use of Information Technology by Nonprofit Organizations: Increasing Capacity and Untapped Potential. *Public Administration Review* 67, 3 (2007), 474–487. DOI: <http://dx.doi.org/10.1111/j.1540-6210.2007.00730.x>
 18. Jeffrey Heer and Maneesh Agrawala. 2007. Design Considerations for Collaborative Visual Analytics. In *Proc. of VAST (VAST '07)*. IEEE Computer Society, Washington, DC, USA, 171–178. DOI: <http://dx.doi.org/10.1109/VAST.2007.4389011>
 19. Becky Hogge. 2010. *Open data study*. Technical Report. Transparency and Accountability Initiative. http://www.soros.org/initiatives/information/focus/communication/articles_publications/publications/open-data-study-20100519
 20. Youyang Hou and Cliff Lampe. 2015. Social Media Effectiveness for Public Engagement: Example of Small Nonprofits. In *Proc of CHI (CHI '15)*. ACM, New York, NY, USA, 3107–3116. DOI: <http://dx.doi.org/10.1145/2702123.2702557>
 21. E. Lakomaa and J. Kallberg. 2013. Open Data as a Foundation for Innovation: The Enabling Effect of Free Public Sector Information for Entrepreneurs. *Access, IEEE* 1 (August 2013), 558–563. DOI: <http://dx.doi.org/10.1109/ACCESS.2013.2279164>
 22. Christopher A. Le Dantec and W. Keith Edwards. 2008. Designs on Dignity: Perceptions of Technology Among the Homeless. In *Proc. of CHI (CHI '08)*. 627–636.
 23. Christopher A. Le Dantec and W. Keith Edwards. 2010. Across Boundaries of Influence and Accountability: The Multiple Scales of Public Sector Information Systems. In *Proc. of CHI (CHI '10)*. ACM, New York, NY, USA, 113–122. DOI: <http://dx.doi.org/10.1145/1753326.1753345>
 24. Cecelia Merkel, Umer Farooq, Lu Xiao, Craig Ganoe, Mary Beth Rosson, and John M. Carroll. 2007. Managing Technology Use and Learning in Nonprofit Community Organizations: Methodological Challenges and Opportunities. In *Proc. of CHIMIT (CHIMIT '07)*. ACM, New York, NY, USA, Article 8. DOI: <http://dx.doi.org/10.1145/1234772.1234783>

25. Florence Millerand and Geoffrey C Bowker. 2009. Metadata Standards. In *Standards and their stories: How quantifying, classifying, and formalizing practices shape everyday life*. 149–165.
26. William P OHare. 1988. The rise of poverty in rural America. (1988).
27. Benoît Otjacques, Mickaël Stéfas, Maël Cornil, and Fernand Feltz. 2012. Open Data Visualization Keeping Traces of the Exploration Process. In *Proc. of WOD (WOD '12)*. ACM, New York, NY, USA, 53–60. DOI: <http://dx.doi.org/10.1145/2422604.2422612>
28. Dave Randall, Jon O'Brien, Mark Rouncefield, and John A. Hughes. 1996. Organizational Memory and CSCW: Supporting the Mavis Phenomenon. In *Proc. of OZCHI (OZCHI '96)*. IEEE Computer Society, Washington, DC, USA, 26–.
29. Jennifer Stoll, W. Keith Edwards, and Elizabeth D. Mynatt. 2010. Interorganizational Coordination and Awareness in a Nonprofit Ecosystem. In *Proc. of CSCW (CSCW '10)*. ACM, New York, NY, USA, 51–60. DOI: <http://dx.doi.org/10.1145/1718918.1718930>
30. Amy Volda. 2011. Shapeshifters in the Voluntary Sector: Exploring the Human-centered-computing Challenges of Nonprofit Organizations. *interactions* 18, 6 (Nov. 2011), 27–31. DOI: <http://dx.doi.org/10.1145/2029976.2029985>
31. Amy Volda, Ellie Harmon, and Ban Al-Ani. 2011. Homebrew Databases: Complexities of Everyday Information Management in Nonprofit Organizations. In *Proc. of CHI (CHI '11)*. ACM, New York, NY, USA, 915–924. DOI: <http://dx.doi.org/10.1145/1978942.1979078>
32. Amy Volda, Ellie Harmon, and Ban Al-Ani. 2012. Bridging Between Organizations and the Public: Volunteer Coordinators' Uneasy Relationship with Social Computing. In *Proc. of CHI*. 1967–1976.
33. Wesley Willett, Jeffrey Heer, Joseph Hellerstein, and Maneesh Agrawala. 2011. CommentSpace: Structured Support for Collaborative Visual Analysis. In *Proc. of CHI (CHI '11)*. ACM, New York, NY, USA, 3131–3140. DOI: <http://dx.doi.org/10.1145/1978942.1979407>