Empowering Communities with Data: Role of Data Intermediaries for Communities' Data Utilization

Ayoung Yoon
Indiana University Purdue University Indianapolis (IUPUI), USA. ayyoon@iupui.edu

Paula Jo Anders McNally
Indiana University Purdue University Indianapolis (IUPUI), USA. pjmcnall@iu.edu

Andrea Copeland
Indiana University Purdue University Indianapolis (IUPUI), USA. ajapzon@iupui.edu

ABSTRACT

Data have significant potential to address current societal problems not only at the federal and state levels, but also in smaller communities, in neighborhoods, and in the lives of individuals. While the proposition for this potential is that data are and will be shared with and reused by and for communities at different levels, not all data are not systematically or routinely shared for reuse with communities due to social, structural and technical infrastructure barriers. Data intermediary organizations can play a significant role in removing existing barriers while unlocking the potential of data for all, particularly for communities with limited human or financial resources, limited access to existing data infrastructures, and underserved populations. Considering the significance of the data intermediary organizations on local communities, this study aims to explore the role of intermediaries that usually facilitate community members/organizations' data utilization. The findings of this study reveal that data intermediary organizations play four major roles that are crucial in communities' data utilization: (1) democratizing data, (2) adding value to existing data, (3) enhancing communities' data literacy, and (4) building communities' data capacity. This study has several important implications to offer a solution to overcome the challenges of data reuse at the local level.

KEYWORDS

Data reuse, open data, data infrastructure, data intermediaries, data curation

INTRODUCTION

Data have significant potential to address current societal problems (e.g., education, health, economic development, and the environment) not only at the federal and state levels, but also in smaller communities, in neighborhoods, and in the lives of individuals. Many have argued that data have value as a source for community economic development, disaster planning, and decision-making in citizens' daily lives (Kassen, 2013; Levin & Schneir, 2015). In the report *A World that Counts, Mobilizing the Data Revolution*, the United Nations Secretary-General's Independent Expert Advisory Group (IEAG, 2014) underscored the significance of data particularly for underdeveloped countries or neighborhoods, arguing that "data are the lifeblood of decision-making" (p. 4) and that the recent data revolution has created opportunities to design, monitor and evaluate effective policies for communities.

While the proposition for this potential is that data are and will be shared with and reused by and for communities at different levels, much data—including, particularly, data collected by university and research institutes—are not systematically or routinely shared for reuse with communities (Research Data Alliance [RDA] Data for Development Interest Group, n.d.; Urquilla & Shelton, 2015). Government data are traditionally shared with citizens, and while recent open data initiatives accelerate this tradition, Douglass et al. (2014) pointed out that there are still structural and technical infrastructure barriers to be overcome for data to be fully accessible and easily usable. Research data are usually shared within scientists' own communities of practice, although researchers often share data interpretations to support community development. The IEAG (2014) pointed out that unequal access and use of data limit the role that data can play in the realization of sustainable development and argued that it is critical for all stakeholders (e.g., government, regional institutions, public/private data producers, academics, and citizens) to work to create a more equitable data ecosystem.

Data intermediary organizations can play a significant role in removing existing barriers while unlocking the potential of data for all, particularly for communities with limited human or financial resources, limited access to existing data infrastructures, and underserved populations. Van Schalkwyk, Canares, Chattapdhyay, and Andrason (2016) emphasized the importance of these data intermediaries in open data ecosystems, as they can connect different stakeholders and facilitate the flow of data. However, despite the significance of the role and impact of data intermediary organizations on local communities, these groups have been paid little attention in scholarly works, and little is known about how data are utilized among local communities and at the neighbourhood level (Van Schalkwyk et al., 2016).

This study aims to explore the role of data intermediary organizations that usually connect community members/organizations

with data sources and provide partnerships and consultations to support communities' data utilization (Yoon, 2017). We adopt our definition of data intermediary organizations from Van Schalkwyk et al. (2016), broadly defining such groups as agents positioned between two other agents in a data supply–demand chain and that facilitate the use/reuse of data. By uncovering the role of data intermediary organizations and analyzing the services they provide to the populations they serve, this study contributes to our understanding of how local communities utilized data to have meaningful impacts on their communities. The findings of this study have several important implications to offer a solution to overcome the challenges of data reuse at the local level.

LITERATURE REVIEW

A growing body of literature discusses the power and value of data to enhance data-enriched citizen and public engagement in policy- and decision-making relevant to citizens' daily lives (Gurstein, 2011; Kassen, 2013; Levin & Schneir, 2015). Keserű and Chan (2015) argued that open data not only impact economic development, but also increase state or institutional responsiveness, reduce levels of corruption, build new democratic spaces for citizens, empower local and disadvantaged voices, and enhance service delivery and effective service utilization. Levin and Schneir (2015) also underscored the importance of making reliable public data accessible to anyone, anytime, for any purpose, as such data have the power to help citizens make better decisions for their communities.

In the U.S. context, several cases demonstrate the potential of data reuse for community members. Chicago's open data portal driven by open government data initiatives resulted in several citizen-initiated projects that developed tools to serve Chicago's diverse community by improving the transparency of local governments' political processes (e.g., chicagolobbyists.org) and maintaining local neighborhoods as safe, open, and healthy green spaces for the local community (e.g., Mi Parquet Project; Kassen, 2013). Often, local organizations lead these projects. For example, the Green and Healthy Homes Initiative (GHHI) in Baltimore utilized health and housing data to provide healthy homes for communities and advocate for increased efforts to prevent lead poisoning (Urquilla & Shelton, 2015). These cases show how different community members (e.g., city planners, local government officers, economic developers, non-profit educational consultants, foundations and corporate organizations, and the public, including parents with concerns about childcare and school systems) can utilize or reuse data on regional areas and population groups collected by the private and public sectors to better understand the factors influencing their communities.

This emerging interest in using data for the community has sparked several discussions regarding how to make data open and sharable, many of which call for action on the part of the public sector. However, the movement to make data fully accessible and easily usable still faces several challenges, such as structural and technical infrastructure barriers (Douglass et al., 2014). In this context, several studies have explored the concept of intermediaries as agents capable of removing some of these barriers (e.g. Van Schalkwyk et al., 2016). The concept of Intermediaries was first discussed in the context of information and communication technology (ICT) as bridging or support organizations that link different agents by transmitting information from one to another (Brown & Kalegaonkar, 2002; Sapsed et al., 2007). They were later explored in different organizational contexts, such as business and e-government, as agents for diffusing services (Janssen & Zuiderwijk, 2014; Sein & Furoholt, 2010). In the context of open data, Van Schalkwyk, Willmers, and Czerniewicz (2014) reported that data intermediaries play several possible roles, such as increasing data accessibility and utility, serving as 'keystone species' in data ecosystems, and democratizing the impacts and use of open data. Several other studies have echoed the potential and significance of intermediaries in the contexts of government data (Chattapadhyay, 2014), developing countries (Davies, 2014; Van Schalkwyk et al., 2014), and broadly in ICT (Diaz Andrade, & Urquhart, 2010) and demonstrated intermediaries' crucial role in linking different agents in open data supply chains, including grassroots communities (Van Schalkwyk et al., 2016).

While the interest in and need for data intermediaries is growing, little empirical research has been conducted to understand their roles and the processes of data diffusion and utilization in local communities. By empirically investigating the roles and functions of data intermediary organizations, this study contributes to understand one facet of data utilization at local communities.

METHODS

This study takes qualitative approach to explore the role of data intermediaries. We broadly define a data intermediary as an organization that facilitates the (re)use of data to impact communities. The project team interviewed one to three key staff at nine data intermediary organizations across the U.S. We used several strategies to identify data intermediary organizations in different U.S. cities. As previous research has suggested that community research centers often help community members/organizations with their data needs—and, thus, serve as a form of data intermediaries (Yoon, 2017)—we first conducted a Google search using several key words, such as *community research center* and *community research institute*. Using this strategy, we identified approximately 50 community research centers that existed as non-profit organizations or were affiliated with universities. We also reviewed the 42 partner organizations across 30 cities included in the National Neighborhood Indicators Partnership (NNIP, http://www.neighborhoodindicators.org/), a collaborative effort by the Urban Institute (http://www.urban.org/)

and local partners to further the development and use of neighborhood information systems in local policy-making and community building. We carefully reviewed each organization's website for information regarding their missions, services offered, and past projects. After assessing each organization's fit for this study, we then chose data intermediary organizations in small, medium, and large cities (e.g., Grand Rapids, MI; San Antonio, TX; and Boston, MA) that agreed to participate in the study. Since one organization dropped out during the interview process, we interviewed a total of 15 staff members (one to three per organization) from a total of nine organizations.

We conducted a semi-structured phone interview with our participants following pre-developed interview protocols. The interview questions covered the organizations' missions, the role of the staff, types of services to support communities' data utilization, the interviewees' experiences working with community members/organizations, and challenges associated with support communities' data work. All interviews were audio-recorded and fully transcribed by professionals. The sample sets of data were inductively analyzed to develop a set of codes representing emerging themes. Then, the rest of the data were deductively coded following the developed coding scheme. A qualitative data analysis tool, Nvivo, was used to code the transcripts and analyze the data. The inter-coder reliability among the project team members was 93%, and we had more than 90% agreement for all codes except one, which still achieved an acceptable 86% agreement rate. As this study aims to "understand and describe as accurately as possible the phenomena" (Groenewald, 2004, p. 5), we followed a phenomenological and interpretive analysis approach (Smith, Flowers, & Larkin, 2009).

RESULTS

The nine participating organizations were drawn from small, medium, and large cities across the U.S. (Dallas, TX; Columbus, OH; Boston, MA; San Antonio, TX; Grand Rapids, MI; New Haven, CT; Indianapolis, IN; Milwaukee, WI; and Providence, RI). Four were independent non-profit organizations, and five were university-affiliated organizations. Six organizations were part of the NNIP network. The sizes of the organizations varied from 7 to 30 staff members, with average staff of 15. Of the 15 staff who were interviewed for this study, 4 were male and 10 were female. Their ages ranged from 30s to 60s. Table 1 provides basic demographic information about the participants, including their education levels and years of experience at the organizations. In the following section, we present the participating organizations' primary missions and services, as well as some of the challenges they have faced in fulfilling their missions.

Organization	Staff	Age	Gender	Position	Years at the Organization	Level of Education
O1	S11	30s	F	Program management	2.5	MS
	S12	30s	F	Co-director	5	PhD
	S12	40s	F	Co-director	4	PhD
O2	S21	40s	F	Director	4	PhD
	S22	30s	F	Director of research	2.5	PhD
O3	S31	50s	M	Director	10	PhD
O4	S41	50s	M	Senior director	6	PhD
	S42	50s	F	(Former) research associate	5	PhD
O5	S51	30s	F	Director	5	PhD
O6	S61	30s	M	Executive director	9	BA
O7	S71	50s	F	Director of collaborative research	22	MLA
	S72	30s	M	Research analyst	1	MA
O8	S81	60s	F	Director	1.5	PhD
	S82	30s	F	Data analyst	1.5	BA
O9	S91	30s	F	Research associate data analyst	2	PhD

Table 1. Demographic information

Organizational Missions and Serving Communities

An organization's mission is crucial in justifying its existence, differentiating it from other organizations, facilitating organizational innovation, and helping it achieve intended performance (McDonald, 2007). The ways in which the staff explained their organizational missions reflected the ways in which they perceived these missions in relation to their job responsibilities. From their descriptions of the missions, we found that data were either explicitly or implicitly related to the intermediary organizations' overall missions. Staff from most intermediaries related that their mission was to "strengthen communities with data" by "connecting the communities with data" (e.g., easy access, data aggregation, and analysis) and/or by "helping the communities to make a better decision with data" (O2, O5, O6, O7, O8, O9). Staff from two organization specifically stated that they contributed to "the democratization of data" through different types of data services for public use (O8, O9). Other staff emphasized how their organizations bridged communities and different partners (e.g., academics, business, etc.) by promoting collaborations and community research (O1, O3, O4). One staff member from O4 underscored the organization's mission to support community development by serving as "an information resource for community members." Although these three organizations (O1, O3, O4) did not explicitly mention data in their missions, data were inherently associated with their projects and community services. For instance, S11 discussed the significance of data in all projects, noting that data helps

community organizations gain a better understanding of their communities at large.

Most organizations tended to focus their services on their regional geographic areas, such as their neighborhoods, counties, and cities; however, some stated that they also served other cities in and out of the state. Their clients or partners were predominantly non-profits (e.g., community foundations, advocacy organizations, community service organizations like the YMCA, etc.) and local government or quasi-government entities, but they also occasionally worked for business associations and individual companies, "although that's less common" (S41).

Emerging Data Needs for Communities

The staff interviewed for this study experienced emerging data needs with their targeted communities. To meet these needs, they interacted with and used a wide range of data. Many reported working with publicly available data published by various government organizations, such as the Department of Education, the Department of Health, and the Bureau of Labor Statistics. Census data and the American Community Survey were among the sources commonly mentioned, but the interviewed staff also utilized data from local agencies to learn more about health, economic indicators (income, unemployment, etc.), housing, education, and business. Though the majority relied on open data, several requested private, proprietary data for specific community projects.

From the staff members' perspectives, one reason for communities to utilize data is to understand their neighborhoods in order to "make community decisions based on data" (S22), such as developing community-specific programs or services. Data are "important in looking at trends and opportunities" (S91) because data "help [community organizations] understand what the needs of their constituencies are and how best to position themselves for the future" (S41). Several echoed the need for data "to set new goals for [community] programs" (S91) or "to start a new initiative" (S51). These initiatives pointed to community efforts to make decisions based on data. Data were also crucial in "adjust[ing] some of [the community organizations'] services if [the communities] have determined that the data show that they're really not having the same impact that they anticipated" (S91). Often, this need was tied to the organizations' program evaluation or outcome measures, since understanding the impacts of their programs guided the organizations to change what they had been working on and how.

Communities' data needs were also highly associated with communities' efforts to apply for grants and secure funding. Staff from almost all intermediaries mentioned the importance of grants for community organizations, most of which are small, non-profit, and in great need of financial support. Data provided evidence of the impact of the community organizations' proposed projects; therefore, community organizations increasingly tried to "visualize these data or to create data tables that they can use for grant applications" (S91). S91 further explained that grant agencies require data for project/program measures, meaning that community members "are asked to be very data fluent in not only their grant applications, but also their grant reporting. […] They're being asked to demonstrate their impacts in terms of quantitative outcomes." Sometimes, funders required community organizations "to do an external evaluation […] to prove that the program is working" (S61), creating another strong data demand on the community side.

Community Data Services

To meet their communities' various data needs, the surveyed intermediary organizations provided different types of data services. The depth and breadth of the services they offered varied depending the intermediary organizations' missions and scales.

Support Data Reuse

Supporting communities' use of secondary data was one of major areas in which most data intermediary organizations offered services, which could be fee-based or pro bono, depending on the scale of the work and the intermediary's service model. One way of supporting data reuse was to simply help "citizens, non-profits, [and] government agencies have easy access to data and knowledge that comes from that data" (S22). Staff from the majority of intermediaries noted that community members occasionally knew what data they needed and wanted, but more often were "not so familiar with pulling secondary data" (S22) and did not know where to search.

The data intermediaries also offered more sophisticated levels of service to support data reuse through data aggregation or integration. Many intermediaries either compiled or aggregated publicly available data by importing or integrating the data into their own system, which S71 described as an activity "adding value" to enhance data utilization. S21 said "a lot of people [from the community] were intimidated" when using secondary data because, even though such data are publicly available, "they are not necessarily publicly available in a ready to use format" (S71). Intermediary organizations took data from various sources, re-organized datasets to meet communities' needs (e.g., by geography or neighborhood), broke geographic data down into smaller geographies, made the data clear and easy to explore, and dug down into the neighborhood level for local stakeholders. S61 described this process as transforming data "from nothing to useful for something," noting that publicly available raw data are often not user-friendly to community members.

Data tools were essential in supporting communities' data reuse. Several intermediaries developed data tools from their data aggregation and integration efforts to facilitate communities' data utilization. S71 said that "provid[ing] tools to community members to help them really compile the information that's of interest to them […] is a very important" piece of a data intermediary's work. While the functions of such tools differed across the intermediaries, common features included interactivity and the capability to present the data at the local level. Using the tools, community members "could select interested indicators as well as geographies and instantly map relevant information" (S22), functionalities that significantly supported the communities' data utilization. The tools also allowed communities to break data down to the neighborhood level, allowing community members to "build an index and look at comparisons between neighborhoods and trends over time" (S51). Being able to dig into data at the local level was crucial for communities' data reuse, as data at the state and federal level were often too general to accurately represent communities' current situations. Intermediary organizations also offered free training sessions to "teach people how to utilize those tools" (S51, S71) to maximize the benefits to the communities.

Collect Local Data

Several intermediaries were actively involved in collecting local data that accurately represented their communities, largely due to the lack of existing secondary data at the local level. S41 pointed out the "deficiencies in the [existing] data," stating this was why intermediary organizations had to "go out and [collect] primary data." S61 echoed that "often you can't get that granularity of data [at local level] from a government agency." The fact that publicly available data rarely provided enough information on a small scale was one of challenges hindering communities' data utilization. S81 said that the lack of data representing local communities was the major issue in supporting communities' data work, leading intermediaries to collect data to complement existing data.

Help with Analysis and Interpretation

While some intermediaries stated that they "don't do analysis" (S51), several others engaged with the analysis and interpretation process when asked to do so by their clients (i.e. community members). S91 said, "most of the smaller [community] organizations, […] they just don't have the capacity to analyze [data] themselves." When the intermediaries analyzed data for the communities, they typically did more than simply share the numbers in written reports. Instead, they "visualize [the data] or map [the results] for [the organizational needs] so that [the community organizations] can then take that output and use [it right away]" (S81). S71 described the analysis process as "a partnership, where we don't just interpret the data and send it back to [the clients], but there's multiple sessions where they're talking about what the questions are, and then a preliminary data analysis is presented, and so it is an interactive process, and there's a lot of collaboration." Further, O7 described "working with [the clients] to understand their needs and their culture, and everything that would go into forming whatever final data project they need" (S71). These reflections show that the intermediaries actively communicated with their clients throughout the analysis process to make sure the clients accurately interpreted and understood the data for their purposes.

Promote Data Sharing and Share Data with Communities

A few intermediary organizations also contributed to the data-sharing landscape by supporting community organizations' data-sharing practices. S91 said that O9 had been working with different organizations that wanted to share data-driven information with members of the community and had "been […] creating data-sharing relationship agencies [that currently do not share their data] to obtain new data sets." Using the data from these agencies, O9 was "linking the data and sharing back out the aggregate to them" as part of its data reuse support. The intermediary organizations also shared the data they collected with the communities. S61 said, "[regarding] our own data, our goal is to make everything public. […] There's some data we released more in an unidentified format or in its original and raw format. […] Pretty much everything we [collect], we make public."

While intermediary organizations occasionally shared raw data, including both secondary data they acquired and primary data they collected, several intermediaries shared data in more "processed" forms, considering the community members' different levels of data literacy and different abilities to search, understand, and interpret data. The most common form of "processed" data was "data reports" with tables and infographics on important indicators that of communities' interests. Often these reports were included in the clients' requests. For instance, they might "compare how [the city] is doing compared to other peer cities on a variety of indicators, such as poverty levels, educational levels, housing, things like that," and were "shared around the community a lot" (S22) through the clients. In other cases, intermediaries produced data reports for general community members or citizens so that any stakeholder could learn about the community and different topics/indicators and use the data for their own purposes. Not all of the intermediaries tracked who accessed and used these reports, but some had a clear understanding of their users, including legislators, community advocates, government officials, and the general public.

Improve Communities' Data Practices

Although it was not common, two intermediaries helped community organizations improve their internal data practices. Community organizations often needed to collect internal data for program evaluation and outcome measures for grant reporting, but not all implemented best practices. S51 ran into "a lot of issues where organizations don't support their own data or don't

keep [the data] in the same format all the time," generating "more process concerns." O9 helped these community organizations implement the best data organization method to "do or plan out their evaluation plan [early] so that they're not caught at the end of their grant cycle to figure out how demonstrate what their impact was in anything more than an anecdotal fashion" (S91). O1 was more actively involved in community organizations' internal data practices, helping them "curate the data they collect, and the method by which they collect, and the methods by which they store, and their quality assurance and improvement of that, […] and [track] longitudinal patterns" (S13). This was one of the most active forms of data support that intermediary organizations provided communities.

Foster Community Research

A small number of intermediaries actively participated in community research either by independently conducting research "to understand specifically what's going on in [the county]" (S11) or through partnerships with research institutions and community organizations. S61 called this effort "community-engaged research," which could take the form of a survey study "in conduction with the foundations, or hospitals, or health departments, or city offices" to answer community-specific research questions. These community research projects usually produced valuable data specific to the local context and helped to overcome some of the limitations of using secondary data.

Challenges to Support Communities' Data Work

While data intermediary organizations supported communities' data utilization through the various services described above, they also experienced different challenges doing their work. These challenges were associated with several different factors, such as a lack of organizational resources, a lack of local data, a lack of data literacy among community members, and difficulties building a partnership between communities and academic institutions.

Lack of Organizational Resources

Despite growing community demands for data utilization, not all data intermediary organizations were equipped with sufficient financial and human resources to either maintain current services or extend/develop new services for communities. Financial resources were crucial to the intermediary organizations' sustainability, and several mentioned that their business model was "100% self-funded" with no state support (and, thus, fully funded by projects). This limited opportunities for community organizations that could not afford a service fee, although some intermediaries noted that they did smaller-scale "pro bono work" (S21) for smaller communities. Limited financial resources could also limit the development of a new service or the extension of an existing service. As S71 mentioned, these "resource constraints prevent us from working on [new area because] we don't have the funding to [do all the work]." For smaller intermediaries, a lack of human resources was the factor limiting their community data support, as no single staff member can have all different types of expertise and data skills. S61 said, "the staff time and staff level of expertise" also mattered for collecting and working with complex data.

Lack of Local (Neighborhood) Data

While data intermediary organizations shared some common challenges relating to data reuse, such as difficulties accessing secondary data, difficulties aggregating data due to interoperability issues, and difficulties working with sensitive data (Lucero et al., 2015; Yoon, 2016), one notable challenge that staff from all intermediaries identified was a lack of data representing local areas or neighborhoods. Although all staff utilized publicly available data for community data work, they agreed that "the public data are fairly general, and usually the scope is fairly large (S41)." Many noted that there were no data available on particular topics (e.g., health data, unemployment data) at the neighborhood or small-town level. Alternatively, they used or let communities use "data at a zip code level to portray what's going on in their neighborhood," even though such reports are "just basically generally wrong" (S81) because "zip codes are not necessarily related to neighborhood[s]" (S21). S21 said this made community members really "angry […] and the people in a particular neighborhood within that zip code are like, 'that's so not true' because their shared daily experience is different than what the data suggested, and so they're freaking out." This was the major challenge the intermediaries faced when supporting communities' data work, and it was also why data intermediaries needed "to figure out a way to develop our own [data]" (S41). In addition, publicly available data were "notoriously old, [and] it was really common to find data that would be six years old [or] a couple-year lag" (S42). Usually, communities needed current, up-to-date data representing current community problems; in such cases, old data were not particularly useful.

Community Members' Data Literacy Skills

One critical issue relating to interactions with communities was differences in community members' data literacy skills. Many intermediary staff mentioned that different levels of data literacy influenced how they communicated with both their clients and the general public. S72 said, "general citizens are probably not being very accurate in terms of interpreting [data]," which can be a challenge when working with them. Intermediary organizations shared the results of data analyses and/or interpretations in formats "designed to be digestible" (S61). Data reports also aligned with these efforts. Still, a lack of data literacy could hinder intermediaries' communication processes. As S11 shared, "it's hard for [community members] to understand what goes into analyzing and cleaning data. Our community partners get frustrated when they have a request and they don't understand

why we can't generate it that second, why it might take a few days."

Building a Partnership between Academic Institutions and Communities

For those intermediaries that had tried to build or already built a partnership between communities and academic institutions to support communities' data work or community research, a distance between the two entities often posed a challenge. A partnership with university researchers can have great benefits for communities' data work because academic researchers are experts in topical areas or domains. S42 described one case in which he was "a little suspicious about the motivation for doing the community need assessment [requested by one community organization]." When he consulted with experts from a local university, one of them "flat out said, 'It's a fad, and this is what's going on politically' because they had done years of research on the area." He said the researchers immensely helped to clarify where to look for data sources while crystalizing the real problems to address using the clients' data sources. However, there was still a degree of distance between communities and academic institutions because the two parties tended to have different motivations and needs. Academic researchers tended to focus on education and research; thus, "they push a publication, research design […] that would be required and understandable in [an academic] environment" (S81). By contrast, communities "may not see much value in research itself […] They may be struggling with day-to-day challenges […] and don't see any short-term or immediate benefits from [academic research]" (S31). S81 also found that, "sometimes, data people [from academia] are not very respectful of the [community members]." When academic researchers lacked a deeper understanding of communities' needs, community members found their collaboration intrusive.

DISCUSSIONS

This study explored data utilization at the local community level through the work of data intermediary organizations. The findings of this study reveal that data intermediary organizations play four major roles that are crucial in communities' data utilization: (1) democratizing data, (2) adding value to existing data, (3) enhancing communities' data literacy, and (4) building communities' data capacity.

Democratizing Data

Data democratization involves helping all citizens access and understand data. Although only two intermediary organizations in this study identified data democratization as one of their missions, all intermediaries contributed to democratizing data in practice. Easy access to data when needed at any given moment is an important first step towards data democratization, as previous research has suggested that access to data is the single most important component of successful data reuse (Yoon, 2016). Most intermediaries tried to improve communities' access to data. However, democratizing data means more than just ensuring easy access. Intermediary organizations were actively involved in helping communities make sense of data, ensuring that data could be easily understood and interpreted for the purpose of the communities' data utilization. Aggregating different data sources, breaking data down into desired units (e.g., neighborhood, county), developing self-service data tools for analysis and reuse, and visualizing data were all initiatives in which intermediary organizations engaged to support communities' understanding of data. These efforts align well with the goal of data democratization, which is to expedite decision-making processes, eliminate barriers to accessing or understanding data, and uncover opportunities for organizations (Marr, 2017).

Adding Value to Existing Data

A second critical role that intermediary organizations play is adding value to existing data. Not all of the intermediaries seemed to be aware of this role, though S71 acknowledged that their work added value to data and data utilization. However, whether intermediaries are aware of this role or not, adding value to existing data significantly contributes to open data curation. Value-adding activities are the hallmark of data curation and perhaps the most meaningful and uniquely necessary step in supporting data reuse and utilization. The Digital Curation Center (DCC, n.d.) defined digital (data) curation as the effort to "maintain, preserving and adding value to digital data throughout its lifecycle," as well as to "enhance the long-term value of existing data by making it available" for current and future use. Data intermediary organizations may not contribute to curation of the complete data lifecycle, for example, as the interviewees rarely mentioned long-term data preservation or data sustainability. However, they notably contribute to several other curation components, particularly relating to maximizing access and data preparation. In the context of intermediaries' support of communities' data access, access does not simply mean connecting communities with data by identifying appropriate data sources. The studied intermediaries were also involved in the discovery process, during which they requested data on behalf of communities, suggested potentially good sources, and consulted with local domain experts to find good data sources. Further, much of the intermediaries' work was related to enhancing accessibility and understandability through thorough data preparation. Data preparation is known to be one of most difficult, time-consuming, and knowledge-intensive challenges in data curation. When obtaining data from different sources, the intermediaries had to address issues of interoperability, as well as to clean, merge, aggregate, and integrate data with their data system to make the data more accessible, easily retrievable, and understandable. Such steps enriched the data for communities' further use and analysis. When community members accessed data through their intermediaries' systems, the data were well-prepared for consumption and contained complete and self-explanatory information. Since the ultimate goal of data curation is to support

current and future use/reuse of data, the role that intermediaries play in curation cannot be underestimated.

Enhancing Communities' Data Literacy

Data literacy is known to be fundamentally important for extensively using data for decision-making (Cruosoe, 2016). In the context of communities' and individuals' data utilization, data literacy is generally understood as "the knowledge of what data are, how they are collected, analyzed, visualized, and shared, and the understanding of how data are applied for benefits or detriments" (Crusoe, 2016, p. 27). Staff from the intermediaries noted that it was important to have a certain level of data literacy to ensure smooth communication during the process of working with communities. Often, a lack of such skills was a barrier to data utilization. Data literacy skills were also critical for communities because they needed to interpret data "based on their understanding of their community" (S71). Due to the importance of data literacy for both the intermediary organizations and the communities themselves, many intermediaries either directly or indirectly educated communities on data literacy. For example, they asked about communities' data needs and perspectives on data, which helped them understand the process of data utilization, and debriefed them on the data analysis and the results, thereby helping communities understand the data and draw accurate conclusions for decision making. Many staff discussed how this process was iterative and collaborative. Several intermediaries also offered the training sessions relevant to data, including training on the data tools they had developed and, less commonly, trainings on secondary data. Since general citizens and community members had few opportunities to develop their data knowledge and skills, intermediaries' contributions to data literacy education are critical and a good fit for the niche.

Building Community's Data Capacity

Overall, data intermediary organizations significantly contribute to building communities' data capacity. In a way, the simple existence of data intermediaries helped communities be less intimidated by data utilization. This represents one important step in introducing communities to the power of data. By using the data services offered by data intermediaries, communities can become more comfortable with data and perhaps become aware of the need to develop further data literacy skills to fully benefit from data. Further, some data intermediaries were aware of their communities' lack of capacity to use and interpret data, and they sought to address these issues. This approach was reflected in both the collaborative nature of intermediaries' work with communities and their efforts relating to data education. The intermediaries' intentions to build community capacity is illustrated in the following statement:

> S71: "we don't wanna just do things for [the community] and walk away. We want to be showing them how to do it. Obviously, some people, they're not interested in knowing the whole thing, but you have to be interested […] our goal is to increase their ability to do similar things in the future, or at least a few steps forward and maybe adapt to just being more experienced in [data work."

Greater community capacity for data utilization would be a great asset for society, as it would enhance communities' ability to utilize data to empower communities. Ultimately, such capacity will support the goal of open data and data reuse in the community context, which will improve the lives of communities and individuals.

Despite the multifaceted roles that data intermediaries play in helping with communities' data work, the findings also illustrated that there are still many barriers to be improved for intermediaries to support communities and make data really useful for communities. It is worth noting that data intermediaries are not and should not be the only player in communities' data work, and collaboration and/or partnership with other data infrastructures is necessary to overcome some of the barriers and generate synergy to support communities' data utilization. Intermediaries were limited to existing financial and human resources, which made them unable to serve as many community as they wanted. This could include smaller communities that may not be able to afford the service fees, although intermediaries offer a pro-bono service for these organizations. In addition, the need to educate community members about data literacy skills emerged in this study; not all intermediaries had enough capacity to provide a formalized training session to the communities. Educational institutions (e.g., local universities) and local information organizations (e.g., public libraries) can share some of the responsibilities of educating communities about data literacy, and can help with communities' interaction with data, considering their expertise in education, community engagement and services. These collaborations and partnerships will complement the services provided by data intermediary organizations while meeting the needs of communities that intermediaries cannot fully afford. The best method and model for these partnerships should be investigated further by considering each entity's capacity.

CONCLUSION

As the findings demonstrate, data intermediary organizations pay several important roles that affect communities' data utilization. By democratizing data, adding value to existing data, enhancing communities' data literacy, and building communities' data capacity, data intermediaries contribute to empowering communities with data, particularly regarding the communities' decision making relating to important community problems and improving community members' everyday lives.

While this study reports new findings about the role of data intermediaries and data utilization at the local community level, it also has several limitations. First, because the study focused exclusively on intermediaries, the findings reflect only intermediaries' perspectives regarding communities' data utilization. Most of the staff who participated in this study were highly skilled in the realm of data. Speaking directly to community members, who would have much more diverse backgrounds and levels of data skills, is necessary to fully understand how data are used and what kinds of challenges communities face in using data to make a real impact.

ACKNOWLEDGMENTS

REFERENCES

Brown, L. D., & Kalegaonkar, A. (2002). Support organizations and the evolution of the NGO sector. *Nonprofit and Voluntary Sector Quarterly*, *31*(2), 231–258. https://doi.org/10.1177/0899764002312004

Chattapadhyay, S. (2014). Access and use of government data by research and advocacy organisations in India: A survey of (potential) open data ecosystem. *Proceedings of the 8th International Conference on Theory and Practice of Electronic Governance (ICEGOV)*, 361–364. DOI: 10.1145/2691195.2691262

Crusoe, D. (2016). Data literacy defined pro populo: To read this article, please provide a little information. *The Journal of Community Informatics*, *12*(3). Retrieved from http://ci-journal.net/index.php/ciej/article/view/1290

Davies, T. (2014, July). *Open Data in Developing Countries - Emerging Insights from Phase I*. Retrieved from http://opendataresearch.org/content/2014/704/open-data-developing-countries-emerging-insights-phase-i

Díaz Andrade, A., & Urquhart, C. (2010). The affordances of actor network theory in ICT for development research. *Information Technology & People*, *23*(4), 352–374. https://doi.org/10.1108/09593841011087806

Digital Curation Center. (n.d.). *What is digital curation?* Retrieved from http://www.dcc.ac.uk/digital-curation/what-digital-curation

Douglass, K., Allard, S., Tenopir, C., Wu, L., & Frame, M. (2014). Managing scientific data as public assets: Data sharing practices and policies among full-time government employees. *Journal of the Association for Information Science and Technology*, *65*(2), 251–262. https://doi.org/10.1002/asi.22988

Groenewald, T. (2004). A phenomenological research design illustrated. *International Journal of Qualitative Methods*, *3*(1), 1–26.

Gurstein, M. B. (2011). Open data: Empowering the empowered or effective data use for everyone? *First Monday*, *16*(2). Retrieved from http://firstmonday.org/ojs/index.php/fm/article/view/3316

Janssen, M., & Zuiderwijk, A. (2014). Infomediary business models for connecting open data providers and users. *Social Science Computer Review*, *32*(5), 694–711. https://doi.org/10.1177/0894439314525902

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

Keser, J., & Chan, J. K. (n.d.). *The social impact of open data*. Sunlight Foundation. Retrieved from http://www.opendataresearch.org/http://www.opendataresearch.org/dl/symposium2015/odrs2015-paper20.pdfdl/symposium2015/odrs2015-paper20.pdf

Levin, P. L., & Schneir, H. (2015, December 2). Imagine a world of good data. *Stanford Social Innovation Review: Informing and Inspiring Leaders of Social Change*. Retrieved from http://ssir.org/data_for_community_driven_solutions/entry/imagine_a_world_of_good_data

Lucero, R. J., Kearney, J., Cortes, Y., Arcia, A., Appelbaum, P., Fernández, R. L., & Luchsinger, J. (2015). Benefits and risks in secondary use of digitized clinical data: Views of community members living in a predominantly ethnic minority urban neighborhood. *AJOB Empirical Bioethics*, *6*(2), 12–22. https://doi.org/10.1080/23294515.2014.949906

McDonald, R. E. (2007). An investigation of innovation in nonprofit organizations: The role of organizational mission. *Nonprofit and Voluntary Sector Quarterly*, *36*(2), 256–281. https://doi.org/10.1177/0899764006295996

Marr, B. (2017, July 24). What is data democratization? A super simple explanation and the key pros and cons. *Forbes*. Retrieved from https://www.forbes.com/sites/bernardmarr/2017/07/24/what-is-data-democratization-a-super-simple-explanation-and-the-key-pros-and-cons/

Research Data Alliance (RDA). (n.d.). *Data for development interest group*. Retrieved from https://www.rd-alliance.org/groups/data-development.html

Sapsed, J., Grantham, A., & DeFillippi, R. (2007). A bridge over troubled waters: Bridging organisations and entrepreneurial opportunities in emerging sectors. *Research Policy*, *36*(9), 1314–1334. https://doi.org/10.1016/j.respol.2007.05.003

Sein, M. K., & Furuholt, B. (2012). Intermediaries: bridges across the digital divide. *Information Technology for Development*, *18*(4), 332–344. https://doi.org/10.1080/02681102.2012.667754

Smith, J. A., Flowers, P., & Larkin, M. (2009). *Interpretative Phenomenological Analysis: Theory, Method and Research*. SAGE Publications Ltd.

The United Nations Secretary-General's Independent Expert Advisory Group on a Data Revolution for Sustainable Development (IEAG). (2014). *A world that counts, mobilizing the data revolution*. Retrieved from http://www.undatarevolution.org/wp-content/uploads/2014/11/A-World-That-Counts.pdf

Urquilla, M., & Shelton, J. (2015, December 18). Bridging communities and government through data. *Stanford Social Innovation Review: Informing and Inspiring Leaders of Social Change*. Retrieved from http://ssir.org/data_for_community_driven_solutions/entry/bridging_communities_and_government_through_data

van Schalkwyk, F., Wilmers, M., Czerniewicz, L., & Czerniewicz, L. (2014). *Case study: Open data in the governance of South African higher education*. Retrieved from http://www.opendataresearch.org/sites/default/files/publications/Open%20Data%20in%20the%20Governance%20of%20South%20African%20Higher%20Education%20WEB.PDF

van Schalkwyk, F., Canares, M., Chattapdhyay, S., & Andrason, A. (2016). Open data intermediaries in developing countries. *The Journal of Community Informatics*, *12*(2). Retrieved from http://ci-journal.net/index.php/ciej/article/view/1146

Yoon, A. (2017, September). Data for community development and decision making. *Research Data Alliance (RDA) 10th Plenary Meeting*. Montreal, Canada.

Yoon, A. (2016, October). Red flags in data: Learning from failed data reuse experiences. *Proceedings of the 79th Association for Information Science and Technology (ASIS&T) Annual Meeting, 53*(1), 1–6. Copenhagen, Denmark. DOI: 10.1002/pra2.2016.14505301126