Understanding social impact of data on local communities

Ayoung Yoon and Andrea Copeland

Department of Library and Information Science,
Indiana University–Purdue University Indianapolis, Indianapolis, Indiana, USA

Abstract

Purpose – The purpose of this paper is to understand the social impact of data on communities from cases of community data utilization.

Design/methodology/approach – This study took an interpretive qualitative approach and conducted a semi-structured phone interview with 45 participants from data intermediaries and local community organizations.

Findings – The results demonstrate both direct and indirect impacts of data on local levels, including resolving local problems from data-driven decisions, realizing unknown problems or correcting misrepresented problems, changing community data practices, strengthening community identity and enhancing the community's data skills.

Practical implications – The research shows that communities' data utilization supported community-led actions and initiatives from the bottom-up perspective, which demonstrates the need for supporting communities' data work.

Social implications – Minimizing inequality in data utilization should be resolved so that all communities can benefit from the power of data.

Originality/value – By demonstrating evidence of data being critical to encouraging communities' data utilization, this study fills the gap in existing research, which lacks a clear explanation for how the potential of data can be realized at the local level.

Keywords Data reuse, Open data, Social impact, Data curation, Community informatics, Data impact

Paper type Research paper

Introduction

Many acknowledge the value of data as a source, not just for scientific knowledge, but also for a community's economic development, disaster planning and decision making (Heidorn, 2008; Kassen, 2013; Levin and Schneir, 2015). An exponential growth in data volume supports data's potential as a community source – with a recent movement toward open data making local, regional and national data publicly available (The United Nations Secretary-General's Independent Expert Advisory Group on a Data Revolution for Sustainable Development, 2014; Gurstein, 2011). Community organizations, such as NGOs, nonprofits and local governments with other public sector organizations, become more data-driven when making decisions to fulfill their missions of improving their communities and furthering society. Private sector organizations also utilize data to serve clients and provide meaningful work for their employees (Howson *et al.*, 2018).

While previous studies have reported growing and urgent data needs in communities (Yoon *et al.*, 2018), they have also pointed out the lack of clear explanations for how data's potential can be realized at the local level (Bertot *et al.*, 2014; Kassen, 2013; Yoon *et al.*, 2018; Keserű and Chan, 2015). Existing literature has also suggested that evidence of the social and political impact of open data is incredibly scarce, and analysis of data utilization cases and their impacts is rarely done (Kassen, 2013; Keserű and Chan, 2015). Despite some studies discussed impact of open data through use cases, mostly in the context of open government data (e.g. Janssen *et al.*, 2012; Ruijer *et al.*, 2017), other argued that many were still conceptual (Bertot *et al.*, 2010; McDermott, 2010), and examined a large city at urban

level (Janssen *et al.*, 2012; Martin *et al.*, 2013; Thakuriah *et al.*, 2017), or developing countries (Aronson, 2004; Bentley and Chib, 2016; Mutuku and Mahihu, 2014), which lacks in demonstrating the potential of data at local level, a smaller community. This is partially because the concept of open data, or data revolution, is still new, and it is also inherently difficult to measure social changes generated by open data and data utilization at the local level (Kassen, 2013; Keserű and Chan, 2015).

Providing compelling evidence of the role and impact of data on local, community levels is crucial for promoting communities' data utilization. However, despite the many opportunities data can bring for bettering communities, data utilization can be a tough investment for small, under-resourced community organizations due to the lack of skills and resources for leveraging data (Howson *et al.*, 2018; Open Data Watch, n.d.). Demonstrating strong evidence for the value of data can be one way of encouraging data utilization among community members and helping them find ways to overcome data utilization challenges.

The purpose of this study is to analyze cases of community data utilization to understand the social impact data has on communities. In this study context, "communities" refer to groups of people with a common interest in a particular area. Usually related to a physical, local area, this common interest can be a certain topic about the neighborhood (such as education, aging and economic development) or the physical area itself. This research operationalized the term "social impact" in this study as a significant, positive change that addresses human behavior and cognition, organizational or community culture and community problems. This definition was developed and contextualized based on the authors' understanding of the term proposed by the Interorganizational Committee on Guidelines and Principles (1994). They broadly viewed social impact as a term that also includes cultural impact, defining it in two ways: "the consequences to human populations of any public or private actions that alter the ways in which people live, work, play, relate to one another, organize to meet their needs and generally cope as members of society" and "the changes to the norm, values, and beliefs that guide and rationalize their cognition of themselves and their society" (Interorganizational Committee on Guidelines and Principles, 1994). Via qualitative data analysis, this study will explore how data and data utilization impact communities' lives in a way that addresses community challenges and changes community culture.

Literature review

The idea of data revolution, which is "an explosion in the volume of data, the speed with which data are produced, the number of producers of data, the dissemination of data, and the range of things on which there is data, coming from new technologies," has gained notable attention in recent years from different sectors, as data have become "the lifeblood of decision-making" in many contexts (The United Nations Secretary-General's Independent Expert Advisory Group on a Data Revolution for Sustainable Development, 2014). The IEAG argues that the data revolution, when properly and effectively utilized, will empower people with better policies, better decisions and better outcomes for society (The United Nations Secretary-General's Independent Expert Advisory Group on a Data Revolution for Sustainable Development, 2014).

Recently much literature has discussed data's power and value for driving positive changes on such topics as climate, criminal justice, trade-and-commerce, human trafficking, poverty, food and many more (Levin and Schneir, 2015; Gurstein, 2011). Levin and Schneir particularly focus on the concept of civic data (health, energy, education, economic, transportation, urban development, environmental, national security and criminal justice), all of which are essential for citizens to build stronger communities and better lives (Levin and Schneir, 2015). While many studies discussed the benefits of data driven by open government data initiatives, open data also include data outside of governments and public

sectors, such as research and science data as well as private sector data (Mohammed, 2015). Thakuriah *et al.* (2017) particularly discussed the significance of data from business transactions and opinion-monitoring systems such as real estate, food, house energy, water consumption or customer reviews to understand the patters of communities.

As a response, local communities, including cities and counties, have been trying to utilize data to solve local problems and improve citizens' lives. For instance, the Array of Things Project in Chicago collected community environmental data and shared them with city planners, community organizers, businesses and the general public, which allowed policy makers and community members to work together to identify and address community needs (Urquilla and Shelton, 2015). In Boston, a nonprofit organization utilized data to identify the most successful interventions for at-risk youth to help youth avoid incarceration and escape poverty (Urquilla and Shelton, 2015). While many efforts were driven by open data initiatives led by local governments, community organizations were also voluntarily involved in data utilization and led projects to make community decisions based on data, such as developing strategic plans or community services, conducting program evaluations and outcome measures and applying for grants to secure funding for community projects (Yoon *et al.*, 2018).

Still, there are a number of limitations for data utilization at the local level. Lack of data literacy skills and expertise for leveraging data, lack of resources within community organizations, lack of public resources and limited data availability are major barriers for communities doing any data work (Howson *et al.*, 2018; Yoon *et al.*, 2018; Urquilla and Shelton, 2015). Perhaps due to these barriers, connecting data – particularly civic data – to citizens' lives through policy, advocacy or other actions does not happen very often, although it seems rational considering the potential of data (Levin and Schneir, 2015). Data intermediaries play a significant role in helping communities' data capacity and enhancing their data literacy skills (Yoon *et al.*, 2018), but data intermediaries' services are not always available to all communities, as many of them offer only fee-based consultation.

Existing literature also suggests that, while the notion of open data and the potential of data for communities has been gaining attention, the potential realized at the local level is not clearly explained (Kassen, 2013). Several recent empirical studies on data utilization have been conducted at local level, involving examination of community organizations and marginalized communities (Bopp *et al.*, 2017; Johnson, 2015; Kassen, 2013; Thinyane *et al.*, 2018; Yoon *et al.*, 2018), and the relative novelty of the field itself also makes providing compelling evidence difficult (Keserű and Chan, 2015). Many local data projects were initiated with the launch of the official US Government data portal in 2009, and many were relevant to developing tools for utilizing data, which made it hard to track use cases beyond the projects (Kassen, 2013). Keserű and Chan also argue that it is difficult to detect the impact of data on social changes when the changes require multiple stakeholders – and even when the changes are associated with behaviors – which is a known struggle for providing evidence of social impact in general (Keserű and Chan, 2015).

These previous studies present existing research gaps in regard to demonstrating compelling evidence of the role and impact of data on the local level. However, strongly demonstrating this evidence is critical for encouraging communities' data utilization, helping them obtain the benefits of open data and aiding communities' development. One way to understand the impact of data is to build stronger narratives about the changes (Keserű and Chan, 2015), which is an acceptable methodology for understanding impact (Arvidson and Lyon, 2014; Garbarino and Holland, 2009). This study fills these gaps in existing research by using qualitative methods to understand the impact of data on communities.

Method
This study takes and interpretive qualitative approach to explore the impact of data on local communities based on communities' data utilization experiences. To find community

members with experiences using data, this study first identified data intermediary organizations in different US cities. Data intermediaries are known to support communities' data utilizations and work closely with community organizations (Yoon *et al.*, 2018). To locate these intermediaries, the authors used a Google search and a review of 42 partner organizations across 30 cities included in the National Neighborhood Indicators Partnership (NNIP, www.neighborhoodindicators.org/). (The NNIP is a collaborative effort between the Urban Institute (www.urban.org/) and local partners to further the development and use of neighborhood information systems in local policy-making and community building.) From these results, the authors carefully chose and worked with nine data intermediary organizations in small, medium and large cities (e.g. Grand Rapids, MI; San Antonio, TX; and Boston, MA). The researchers first interviewed a total of 15 staff members (one to three per organization) from these nine intermediary organizations, then interviewed 30 community members using the network of data intermediary organizations.

The researchers conducted a semi-structured phone interview with all 45 participants, following two different, pre-developed interview protocols for data intermediaries and community members that were designed to address the present study's research questions. For data intermediaries, the questions addressed the organizational missions with a role of data to support their missions, and their experiences working with community members/ organizations to help communities with data work. They were also asked about any changes their data projects made (including changes at perception level) and evidence of those changes, if applicable. The interviews with community members explored their experiences working with data using specific example projects, such as goals of data project, processes, expected outcomes, and final outcomes or products. Using a critical incident technique, participants were asked to discuss the projects that were already completed to further elaborate the impact of their work on their communities. While they were encouraged to discuss the most recent data projects, community members also talked about any other projects if they deemed it necessary.

All interviews were audio-recorded and fully transcribed by professionals. Serial numbers were assigned to each transcription, starting with S for the staff at data intermediaries and C for the community members from community organizations. The sample sets of data were randomly chosen and inductively analyzed to develop a set of codes, using interpretive qualitative content analysis methods (Hefferon and Gil-Rodriguez, 2011; Mayring, 2014). The unit of analysis was each data project, as more than one participant could discuss the same project – especially when participants were from the same organization or had worked, as community organizations or data intermediaries, on the same project. The content analytical technique of inductive code formation was appropriate for answering this study's research questions concerning how communities utilized data and what impacts the data (projects) had on their communities. To capture the multifaceted nature of data's impact on communities, the impact was double-coded, when necessary, as internal vs external, direct vs indirect and evidence vs perception. The rest of the data were deductively coded following the developed coding scheme, using the qualitative data analysis tool, Nvivo. The inter-coder reliability among the three coders was 93 percent. This analysis was exploratory in nature, aiming to provide new insights on a social phenomenon that has not yet been studied clearly. As an exploratory study, this analysis did not seek generalizability but, rather, uniquely treated each unit of analysis.

Results
In total, 15 participants were the staff of data intermediary organizations who closely work with community members to help their data utilization. In total, 30 participants were community members from various community organizations – 19 from nonprofit community organizations (e.g. community foundations, advocacy groups, etc.), 4 from

educational organizations, 1 from healthcare, 2 from business and commercial enterprises, 2 from local governments and 1 from a religious organization. In total, 16 participants were female, and 14 were male. Their ages ranged from 20s to 60s; 1 was in his/her 20s; 12 were in their 30s; 18 were in their 40s; 7 were in their 50s; 5 were in their 60s; and 2 were in their 70s.

The level of experience working with data varied among participants, but all of them were involved in at least more than two data projects. Their data sources were also varied – from federal and local governments to research data to health and business data. Some examples include: the census, the American Community Survey, voter records, K-12 school district data, Kids Count, data from the Department of Labor and Training and the Bureau of Labor Statistics, GIS data from city systems and other small, local data sets. While many participants relied on publicly available data, some also utilized proprietary data when necessary, although the process of finding and accessing proprietary data was different than working with open data. Many participants were also involved in data collection, due to the limitations of existing data. Existing data could not always fully answer the community questions being explored, and the participants often wished to combine their own collected data with existing data.

The following section presents the various impacts data brought to the communities based on their experiences working with existing data, as well as the data collection process. Some participants acknowledged the difficulty in illustrating the impact of data utilization, making statements such as: "[I]t's hard to know what happens several steps down the road" (S22) and "[I]t's gonna take a long time […] that's like ten years out and we're not gonna see a turn right away" (C82). However, despite the difficulties, the data present different types of direct and indirect impacts at different levels. Some impacts were more visible than others; some took place in the short-term, while others required long-term tracking to be demonstrable.

Solving local problems based on data-driven decisions

Arguably the most visible and direct data impact is solving local problems based on data-driven decisions. Many participants discussed how local data were the useful sources for making changes in community programs and directions to properly meet community needs. Data became the compelling evidence for supporting project directions. For instance, a city planner (C76) discussed an expressway modeling project intended to resolve severe traffic issues, combining traffic data with data about drivers' behaviors. He said that different stakeholders had "different lens[es]" for understanding the problems, which caused much "block and tackling." In light of this, the data set had "real value" because it "allowed us to move [forward] with the project, [and] get the permits in place." The project was eventually successful based on "sound data [with] right modeling," and he was pleased with "the number of people who come up to us now and say, 'I can't believe how you improved traffic downtown.'" Similar cases were reported when community members utilized data for conducting community risk assessments, which led to proper resource, such as new station buildings, to make sure "we meet the communities' needs" (C81).

Realization of unknown or misrepresented community problems

A number of participants also reported that data utilization enhanced community understanding by correcting misrepresented community problems and unveiling unknown community issues. S91, who worked to resolve community safety issues, argued that sometimes the real problem is people's assumptions of community problems, which may cause wrong investments:

> We've seen where people were going to do major [education] projects for 14-24-year-old boys in a neighborhood, and, when you look at the data, there [are] no 14-24-year-old boys in the neighborhood.

She said that, when a real problem in her area has to do with disparities and discrepancies, people tend to focus on projects related to certain age groups. She usually pushed these neighborhood agencies to properly utilize data by "introducing consideration of disaggregated data, and it made a real difference in what they were doing." S92 reported a similar case with a community General Education Development (GED) program in which "the neighborhood agency moved to different programming." The community data "confirm [ed] why they hadn't been able to recruit [the right] people to their GED program […] [W] ithout the [data] on age breakdown, they didn't realize [the issue with their work]." Correcting misrepresented community issues may not create a direct, short-term impact, but it has a strong, indirect impact by helping organizations correctly address, and solve, community problems.

Unveiling unknown community issues is another change that data utilization brings. One staff member from a data intermediary reported her experience working with a community organization on a project concerning chronic absenteeism:

> I don't believe it was well-known in the state that there were strong longitudinal impacts on student performance based on their attendance rate. […] It became sort of a touchstone in the state, for an easy way to make an impact is to support students in understanding the importance of attendance. […] I think we tend [to] raise awareness about issues that may not have really been well understood before. (S101)

Although her work might not cause any short-term direct change, by raising awareness, she expected to create desirable change in the long-term. C76 echoed the critical, but indirect, impact of data utilization, saying that, in their situation, "the numbers didn't lie, and people were really shocked [when they saw the data]." C31 also saw their project – raising awareness on the city's environmental issue – as a driver to correct problems, since "significant money is being invested" by the organizations who were obligated to correct the problems.

Changing community data practice
According to S91, communities' data work also helps change the culture of communities and community organizations from conversation-based practice to evidence-based practice. When community members utilize their experience-based knowledge to discuss and make decisions – knowledge which may be partially true but may not fully represent the entire communities' situation or problems – data can be used as evidence to better support community members' thoughts:

> I think, as a community, we're very used to planning with brainstorming […] [T]he loudest voice in the room gets the idea of the day, and that becomes an initiative. I think, in a couple of instances, we've been able to insert data into the conversation and help people look for some evidence-based practices, or some data that would suggest that the approach that they're talking about makes sense. And I think that has made a big difference. (S91)

However, the changes did not just involve making decisions. C62 discussed how data utilization experiences also helped improve data practice at their organization in general, considering the impact data can potentially bring. For instance, her organization had attempted, through their data work, to set up a definition of variables to be the focus of their work. She stated: "[W]e've been able to […] get the county to convene around a common definition of attendance and start to measure things in a consistent way." Her organization collected community data on education – realizing the importance of collecting data consistently and how that consistency influenced data reliability, which, in turn, influenced data analysis. Creating definitions of variables of interests built solid data practice within her organization and regarding relevant stakeholders.

Strengthening community identity

Several community organizations reported that they engaged community members (e.g. volunteers, staff and the general public) in the process of data collection to better understand their community and to compromise the limitation of existing data. When community members engaged in data collection, the process itself created a sense of belonging and social support. For instance, several food banks were involved in a project that aimed to develop a better system of food distribution, and the staff were invited to the longitudinal survey study. A project principle investigator (PI) reported that a number of staff members from food banks said that answering survey questions "helps them build empathy […] because they are asking questions they otherwise might not ask, and that could then lead to […] increased social support" (S13). Having a feeling of empathy and social support helped them to build a strong community identity with affective attitudes on the community they served, which further helped the staff better interact with their clients and citizens. S13 saw this as a positive impact because "sometimes [the staff] can be a little transactional" when doing community support work.

Enhancing community data literacy

Engaging community members in the data collection process exposed them to the concepts of data and data utilization, which may be new to most of the general public, and helped them to develop their data skills. C103 said that one community member, who helped with data collection using Tableau, was able to find a job because of her experience and skills with Tableau. When engaging community members or the general public in the data collection process, the project team usually provided training, especially when the process was long-term or when the collection involved technology (such as Tableau). The training, together with the collection experience, helped community members develop interests and skills relevant to data.

The data collection process also helped community members understand community problems on a practical, rather than an abstract, level by observing and making sense of data. C33 shared his experience of having teenagers from different ethnic groups be interviewers on a project about occupational injury, after providing them with proper training. He explained "what was fascinating was that the children who had been conducting [adult] interviews were now looking at the data from their own interviews." The data collection experiences gave the teenagers clear ideas about the issues around them, as they were able to understand what the data said about those issues:

> So, engaging the community itself in the process in a way that was intelligible meant that, all of a sudden, they saw themselves reflected in the data, which they hadn't up until that point. […] For them the data about their community was very abstract, but, when it was the data about themselves, it suddenly became personal, and they were quite capable of understanding how to ask questions with a couple of peers. (C33)

This enhancement of data literacy skills is another important, indirect impact of data utilization at the local level. Communities' data literacy development is very critical to supporting their data work, as data are only meaningful when people can understand and make sense of them.

Discussion

This study's results demonstrate both direct and indirect impacts of data on the local level. By providing compelling and visible evidence of local problems, data helped resolve, and correct the misrepresentation of, these problems while unveiling unknown issues. Often, data were the drivers to community changes by initiating policies and actions to correct community problems, but, other times, they were useful tools for promoting

community awareness of issues; that community awareness eventually led to action and problem solving.

Community engagement in the data collection process also noticeably impacted community members – making them learn about their communities, helping them recognize data as practically connected to their own lives and providing opportunities for them to develop their data literacy skills. Lack of data literacy skills was a major barrier to communities' data work, and, therefore, a program must be developed to help with communities' data skills (Yoon *et al.*, 2018). The results of this research suggest that a potential way to provide this literacy training is through community engaged data work, exposing community members to data.

Communities' data utilization also changed the working cultures and procedures for local initiatives from experience-based to evidence-based decision making. This is an interesting change driven by data. Community organizations and members have so much knowledge of and affection for their region, which, combined with great understanding from their long-committed working experiences, built a compendium of experience-based knowledge about the community. Although there is value to experiential knowledge, it is true that this knowledge may inaccurately represent the communities because perceptions often play a strong role in forming the knowledge. Data can supplement this imperfection by combining valuable knowledge with evidence and, thus, providing strong support for community work.

Overall, this research shows that community data utilization supports community-led actions and initiatives from a bottom-up perspective. Community members have cumulative knowledge of their communities, understand their communities more than any others, and have a right to suggest what they need for making their lives better. Because top-down approaches – with limited facts and cursory understanding – sometimes do not reflect a community's real problems, which may cause poor community services (Levin and Schneir, 2015). Data, however, enables communities to lead changes from a bottom-up approach, and this is the real power of data utilization in local communities.

Conclusion
This study explored different impacts data can bring to local communities. While this study demonstrated the power of data to enable community-led initiatives for making the community better, there are various challenges that hinder communities' data utilization, such as making data easily accessible and keeping data current, accurate and usable (Levin and Schneir, 2015). In addition, there are notable gaps in data access and use, depending on communities' socioeconomic statuses – such as between developed and developing countries, or between information-rich and information-poor people (The United Nations Secretary-General's Independent Expert Advisory Group on a Data Revolution for Sustainable Development, 2014). The inequality of data utilization should be minimized so that all communities can benefit from the power of data.

While this study focused mostly on the positive impacts of data on communities, it is important to note that there can also be negative impacts. Data breaches, misuse of personal information, political manipulation and social harm and data errors are some well-known concerns (Redden, 2017). Little empirical research has been conducted to understand negative data impacts, and further exploration is necessary to obtain a full picture of how data impacts society. The need for the development of data literacy skills at equitable rates across all communities becomes more critical in view of the potential harms of data to individuals and communities alike. This research identified that there is a lack of necessary skills to data literacy skills for community members to engage holistically in the data ecosystem. The mutually beneficial relationship between data centers and community organizations identified in this research establishes the basis for extending that relationship structure to other community-based entities such as the public library, where data literacy skills can be taught.

References

Aronson, B. (2004), "Improving online access to medical information for low-income countries", *New England Journal of Medicine*, Vol. 350 No. 10, pp. 966-968.

Arvidson, M. and Lyon, F. (2014), "Social impact measurement and non-profit organisations: compliance, resistance, and promotion", *VOLUNTAS: International Journal of Voluntary and Nonprofit Organizations*, Vol. 25 No. 4, pp. 869-886, available at: https://doi.org/10.1007/s11266-013-9373-6 (accessed November 11, 2018).

Bentley, C.M. and Chib, A. (2016), "The impact of open development initiatives in lower- and middle income countries: a review of the literature", *The Electronic Journal of Information Systems in Developing Countries*, Vol. 74 No. 1, pp. 1-20.

Bertot, J.C., Butler, B.S. and Travis, D.M. (2014), "Local big data: the role of libraries in building community data infrastructures", *Proceedings of the 15th Annual International Conference on Digital Government Research'14 ACM Press*, Aguascalientes, pp. 17-23.

Bertot, J.C., Jaeger, P.T. and Grimes, J.M. (2010), "Using ICTs to create a culture of transparency: e-government and social media as openness and anti-corruption tools for societies", *Government Information Quarterly*, Vol. 27 No. 3, pp. 264-271.

Bopp, C., Harmon, E. and Voida, A. (2017), "Disempowered by data: nonprofits, social enterprises, and the consequences of data-driven work", *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems, ACM Press*, Denver, CO, pp. 3608-3619.

Garbarino, S. and Holland, J. (2009), *Quantitative and Qualitative Methods in Impact Evaluation and Measuring Results*, GSDRC, University of Birmingham, Birmingham.

Gurstein, M.B. (2011), "Open data: empowering the empowered or effective data use for everyone?", *First Monday*, Vol. 16 No. 2, available at: http://firstmonday.org/ojs/index.php/fm/article/view/3316 (accessed November 11, 2018).

Hefferon, K. and Gil-Rodriguez, E. (2011), "Methods: interpretative phenomenological analysis", *The Psychologist*, Vol. 24 No. 10, pp. 756-759.

Heidorn, B.P. (2008), "Shedding light on the dark data in the long tail of science", *Library Trends*, Vol. 57 No. 2, pp. 280-299.

Howson, C., Jones, L.C., Idoine, C. and Beyer, M. (2018), "How to use data for good to impact society", available at: www.gartner.com/doc/3880666/use-data-good-impact-society2018 (accessed November 11, 2018).

Interorganizational Committee on Guidelines and Principles (1994), *Guidelines and Principles for Social Impact Assessment*, US Department of Commerce, NOAA Tech. Memo. NMFS-F/SPO-16, 29pp.

Janssen, M., Charalabidis, Y. and Zuiderwijk, A. (2012), "Benefits, adoption barriers and myths of open data and open government", *Information Systems Management*, Vol. 2 No. 4, pp. 258-268.

Johnson, M. (2015), "Data, analytics and community-based organizations: transforming data to decisions for community development", *I/S: A Journal of Law and Policy for the Information Society*, Vol. 11 No. 1, pp. 49-96.

Kassen, M. (2013), "A promising phenomenon of open data: a case study of the Chicago open data project", *Government Information Quarterly*, Vol. 30 No. 4, pp. 508-513.

Keserű, J. and Chan, J.K. (2015), "The social impact of open data", Sunlight Foundation, available at: www.opendataresearch.org/http://www.opendataresearch.org/dl/symposium2015/odrs2015-paper20.pdfdl/symposium2015/odrs2015-paper20.pdf (accessed November 11, 2018).

Levin, P.L. and Schneir, H. (2015), "Imagine a world of good data", Stanford Social Innovation Review: Informing and Inspiring Leaders of Social Change, available at: http://ssir.org/data_for_community_driven_solutions/entry/imagine_a_world_of_good_data (accessed November 11, 2018).

McDermott, P. (2010), "Building open government", *Government Information Quarterly*, Vol. 27 No. 4, pp. 401-413.

Martin, S., Foulonneau, M., Turki, S. and Ihadjadene, M. (2013), "Risk analysis to overcome barriers to open data", *Electronic Journal of E-Government*, Vol. 11 No. 2, pp. 348-359.

Mayring, P. (2014), "Qualitative content analysis: theoretical foundation, basic procedures and software solutions", available at: www.ssoar.info/ssoar/handle/document/39517 (accessed November 30, 2018).

Mohammed, H.U. (2015), *Creating Social Value Through the Use of Open Data*, University of Manchester, Manchester, available at: http://studentnet.cs.manchester.ac.uk/resources/library/thesis_abstracts/MSc15/FullText/Mohammed-HafizUsman-diss.pdf (accessed January 11, 2019).

Mutuku, L. and Mahihu, C.M. (2014), *Open Data in Developing Countries: Understanding the Impacts of Kenya Open Data Applications and Services*, iHub Research, available at: www.opendataresearch.org/sites/default/files/publications/ODDC%20Report%20iHub.pdf (accessed January 30, 2019).

Open Data Watch (n.d.), "Understanding the impact and value of data", available at: https://opendatawatch.com/blog/understanding-the-value-impact-of-data/ (accessed November 11, 2018).

Redden, J. (2017), "Six ways (and counting) that big data systems are harming society", The Conversation, available at: http://theconversation.com/six-ways-and-counting-that-big-data-systems-are-harming-society-88660 (accessed November 11, 2018).

Ruijer, E., Grimmelikhuijsen, S. and Meijer, A. (2017), "Open data for democracy: developing a theoretical framework for open data use", *Government Information Quarterly*, Vol. 34 No. 1, pp. 45-52.

Thakuriah, P., Dirks, L. and Keita, Y.M. (2017), "Digital infomediaries and civic hacking in emerging urban data initiatives", in Thakuriah, P., Tilahun, N. and Zellner, M. (Eds), *Seeing Cities Through Big Data*, Springer International Publishing, Cham, pp. 189-207.

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", available at: www.undatarevolution.org/wp-content/uploads/2014/11/A-World-That-Counts.pdf (accessed November 11, 2018).

Thinyane, M., Goldkind, L. and Lam, H.I. (2018), "Data collaboration and participation for sustainable development goals–a case for engaging community-based organizations", *Journal of Human Rights and Social Work*, Vol. 3 No. 1, pp. 44-51, available at: https://doi.org/10.1007/s41134-018-0047-6

Urquilla, M. and Shelton, J. (2015), "Bridging communities and government through data", Stanford Social Innovation Review: Informing and Inspiring Leaders of Social Change, available at: http://ssir.org/data_for_community_driven_solutions/entry/bridging_communities_and_government_through_data2015 (accessed November 8, 2018).

Yoon, A., Copeland, A. and McNally, P. (2018), "Empowering communities with data: role of data intermediaries for communities' data utilization", *Proceedings of the Association for Information Science and Technology*, Vol. 55 No. 1, pp. 583-592.

About the authors
Ayoung Yoon is Assistant Professor at Department of Library and Information Science, Indiana University School of Informatics and Computing (IUPUI). She was RDA/US Data Share Fellow for 2016–2017. She received a PhD Degree from the University of North Carolina at Chapel Hill and MSI Degree from the University of Michigan. Her research areas include data curation, digital preservation and data sharing/reuse. She is particularly interested in building community capacity for constructing effective data infrastructure for sharing and reuse through proper curation practices. Currently, she is conducting several research projects focusing on communities' needs of data sharing, reuse and curation. Ayoung Yoon is the corresponding author and can be contacted at: ayyoon@iupui.edu

Andrea Copeland is Associate Professor at Indiana University School of Informatics and Computing (IUPUI), and Chair of the Department of Library and Information Science. Her research focus is public libraries and their relationship with communities. Her current emphasis centers on connecting the cultural outputs of individuals and community groups to a sustainable preservation infrastructure. She is Co-Editor of a recent volume, *Participatory Heritage*, which explores the many ways that people participate in cultural heritage activities outside of formal institutions.