Research and Applications

Opening government health data to the public: benefits, challenges, and lessons learned from early innovators

Erika G Martin[1] and Grace M Begany[2]

[1]Nelson A. Rockefeller Institute of Government, State University of New York, and Department of Public Administration and Policy, University at Albany, State University of New York and [2]Department of Informatics, University at Albany, State University of New York, and Department of Business and Information Technology, State University of New York at Cobleskill

Correspondence to Erika Martin, Department of Public Administration and Policy, University at Albany, State University of New York, 1400 Washington Avenue, Milne 300E, Albany, NY 12222, USA; emartin@albany.edu; Tel: (518) 442-5243; Fax: (518) 442-5298.

ABSTRACT

Objective: Government agencies are rapidly developing web portals to proactively publish "open" data that are searchable, available in nonproprietary formats, and with unlimited use and distribution rights. In this dynamic environment, we aimed to understand the experiences of 2 early leaders in open health data, the US Department of Health and Human Services and the New York State Department of Health.

Materials and Methods: Semistructured interviews with 40 practitioners and policymakers elicited value propositions, capabilities required for successful open data programs, and strategies for improving impact and sustainability. Transcripts were analyzed using a grounded theory approach to identify common perspectives and divergent viewpoints.

Results: Respondents were optimistic about the value of open data, reporting numerous opportunities to advance the triple aim of lower costs, improved health care quality, and better population health. Benefits to agencies include enhanced data quality and more efficient operations. External benefits include improved health literacy, data-driven changes in health care delivery, consumer engagement, and community empowerment. Key challenges are resources, cultural resistance, navigating legal and regulatory issues, and data quality.

Discussion: The open data movement will likely continue, but success requires sustained leadership, resources, organizational cultural change, promotion of data use, and governance. Jurisdictions that are initiating open data programs can incorporate these lessons from early innovators.

Conclusions: The open data movement has a bright future but unknown long-term impact. To maintain momentum, important directions for the field include reconsidering legal guidance on protecting health data in the open data era and quantifying the return on investment.

Key words: health policy, open data, data sources, data curation

INTRODUCTION

Open data platforms, containing "data that can be freely used, reused and redistributed by anyone—subject only . . . to the requirement to attribute and share alike,"[1] are new resources for health research and practice.[2,3] Although many public health datasets are already online, open data are commonly in nonproprietary formats, searchable in centralized portals, and without data use agreements.[2,3]

The US open data movement was catalyzed by the Obama administration's 2009 Open Data Directive to promote government transparency, encourage public participation in government policies, and increase collaboration across agencies and with private institutions.[4] Federal agencies were required to publish relevant downloadable data in open formats that are easily discoverable, culminating in HealthData.gov, with over 1900 datasets from US Department of Health and Human Services (DHHS) agencies. Other states and cities have subsequently developed their own open data sites.[5]

Although portals are quickly evolving, agencies have limited guidance on releasing data to different audiences.[6] In this dynamic environment, we aimed to understand the experiences of 2 early leaders, DHHS and the New York State Department of Health (NYSDOH), which oversees the first state open data portal devoted to health, Health Data NY. As an open data leader, the NYSDOH has a supportive policy environment, including an executive order for state agencies to post open data.[7] Focusing on early leaders is consistent with a "positive deviance" approach, which deliberately selects top performers to identify successful strategies that can be subsequently tested in a representative sample.[8,9] We explored benefits, challenges, and lessons learned through semistructured interviews with policymakers and practitioners at NSYDOH and DHHS, with additional interviews in California, Chicago, and Washington to assess the generalizability of findings.

METHODS

Participant recruitment

Participants from NYSDOH and DHHS with national expertise were recruited via theoretical sampling, aiming for diversity and breadth of responses.[10–12] Participants had varying expertise and responsibilities for open data, and were recruited until themes repeated in subsequent interviews. Our focus was NYSDOH, with federal interviews to understand the historical context, outside perceptions of New York, and consistency of experiences. We subsequently recruited participants from 3 other jurisdictions described by participants as innovative (California, Chicago, Washington) to evaluate external validity. Of the NYSDOH and federal participants initially targeted, 5 did not respond and 5 declined, citing lack of expertise but suggesting other individuals. As shown in Table 1, participants had various professional positions and came from different organization types. The University at Albany institutional review board deemed the study exempt.

Data collection procedures

A semistructured interview guide (see Supplementary Appendix) contained questions about: the value of open data; political, management, and technical challenges; required capabilities; historical context; and integrating consumers into open health data ecosystems. It was adapted for participants outside the NYSDOH or with specialized expertise.

Thirty-two semistructured interviews, lasting approximately 1 hour, were conducted in person (for 28 participants in New York) or via telephone (for 12 participants outside New York). Most interviews were digitally recorded and transcribed, with a few exceptions, such as executive leaders requesting handwritten notes. Both authors were present for most interviews.

Data analysis

Transcripts were analyzed using a grounded theory approach to systematically discover and report themes and concepts, with a positivist epistemological perspective.[10–12] We independently reviewed all transcripts to identify themes and generate associated codes, compare results to reach agreement, and draft a coding guide. We independently coded 4 transcripts with the guide and subsequently discussed new themes and refined existing codes, thus improving interrater reliability. After pilot coding, G.M.B. subsequently coded all interviews, meeting with E.G.M. regularly to review interviews that were challenging to code and ongoing and emerging themes, and to discuss possible revisions to the coding guide. After coding was completed, we selected codes to review in detail for this analysis and independently reviewed and synthesized the coded data to minimize researcher bias; this also allowed E.G.M. an opportunity to confirm how G.M.B. applied the codes and assess whether codes adequately captured important themes. We describe common perspectives and divergent viewpoints. The Appendix contains the themes and code descriptions. For respondent validation,[13] participants were invited to review the draft.

Table 1. Characteristics of study participants with expertise on open health data release, 2015[a]

Characteristic	Number
Organization type	
New York State Department of Health	23
US Department of Health and Human Services	5
Other New York State agency (Office of Health Information Technology Services, Office of Mental Health)	4
Nongovernmental organization (Health Data Consortium, for-profit health insurer, not-for-profit health services organization)	4
Other state health agency (California, Washington)	2
City health department	2
Primary professional role[b]	
Division director/senior manager	16
Executive leader	7
Manager	7
Independent contractor	4
Legal affairs	3
Data owner	2
Data expert	1
Geographic location	
New York	28
Illinois	2
Maryland	2
California	3
District of Columbia	3
Georgia	1
Washington	1
Total interviews[c]	
Total participants	40
Total interviews	32

[a]*Source:* Authors' interviews with policymakers and practitioners

[b]Clarification of some professional roles is as follows: "executive leadership" refers to someone at the commissioner or C-level, "data owner" refers to someone with expertise on and administrative authority over a specific health dataset and with authority to release it to an open data platform, and "data expert" refers to someone understood by his/her peers to be a key contact and authority on the organization's health data assets. Individuals may have had multiple roles, so this reflects the role most relevant to the interviews.

[c]There are more participants than interviews because some interviews had multiple participants.

Table 2. Anticipated benefits of posting government health data to open data platforms[a]

Benefit	Illustrative examples
More efficient public health operations	• Removal of internal data silos • Faster internal clearance to publish presentations and reports that use data publicly available on open data platform • Fewer Freedom of Information Act requests • Reduction in volume of repeated public queries about specific datasets • Use of food safety data to prioritize which restaurants to inspect first to find violations more efficiently
Improved data quality, timeliness, and usefulness	• End users may have questions about the data or find errors, thereby improving quality • Releasing data may improve metadata and other documentation as they are prepared for external audiences • There may be development of new processes to create data in ways that will improve future data publication • Agency staff may be pressured to release more timely data due to public demand • Data can be automatically refreshed on the platform, vs static webpages, thereby making it more timely • Government contracts may start requiring datasets to be included in final reports, enabling more timely delivery of data • Data can be downloaded in different formats, previously not possible with older content management systems • Open data platforms contain analytic tools that allow end users to interact directly with data online
Improved access to data for external researchers	• Scientific research beyond what agencies have the capacity to do in-house • Preliminary analyses for pilot studies and grant applications • Mechanism to develop new collaborations between public health practitioners and academic partners
Use of data to improve health care delivery and the built environment	• Promote data-driven improvements in health care delivery • Empower local communities to take action on public health issues such as improving the built environment
Improved health literacy	• Promote awareness of health issues such as prevalence of childhood obesity and vaccination rates among children • Opportunity to provide public health messaging on topics such as hospital-acquired infections to accompany the dataset being released • Help consumers locate medical providers • Help consumers identify which restaurants have critical violations • Promote awareness of the value of public health activities
Ability to reach new audiences	• Varied users such as the general public, local health departments, community planning groups, journalists, and developers bring new perspectives and innovative ideas • External users can disseminate information more quickly than agency staff whose products undergo internal review before publications
Creation of new applications for the public	• Code-a-thon challenges to develop new applications using agency data, such as iChoose600 application to help restaurant consumers select meals under 600 calories
Promotion of government transparency and fairness	• All users have access to the same datasets

[a]*Source:* Authors' interviews with policymakers and practitioners.

RESULTS

Current and future benefits of opening government health data

Participants were optimistic about the value of open data. Specific opportunities, summarized in Table 2, included internal benefits to agencies releasing data and improvements in health care delivery, the built environment, community planning and mobilization, and consumer decision-making.

Public health agencies have already experienced instances of more efficient operations. The first is reduced internal data silos, which limit NYSDOH staff access to other divisions' data. Releasing data has reduced cultural resistance to data sharing (described below), and staff can easily access datasets on Health Data NY. The NYSDOH estimates that 700 fewer Freedom of Information Act requests were filed within 3 months of launching Health Data NY's predecessor site (Maximizing Essential Tools for Research Innovation and eXcellence), and that the detailed metadata have reduced response time to dataset queries. Chicago's use of publicly available food safety data to prioritize restaurants for inspection improved efficiency.

The process of making data publicly available may improve data quality, timeliness, and usefulness. Data consumers may have questions, identify errors, or express a desire for additional data, thereby improving the quality, amount, and timeliness of data. One leader explained, "[A] major apprehension people have for making their data available is that somebody's going to find something wrong with it. . . . [But] I want people to use the data and tell me where the problems are." All those in leadership roles expressed this sentiment. Participants described instances where data owners corrected outdated web pages when reviewing the content to create metadata. Platform features enabling users to download data and analyze them online improve their usefulness. Preparing datasets for open data release is also prompting data systems redesigns: "Including open data design early in the process of a surveillance data collection activity improves the overall design . . . [and] . . . thinking about the data elements and actually defining them in a way that are standardized will not just improve open data . . . but will also improve the public health business."

Improved data access could enhance scientific developments, health care delivery, the built environment, and health literacy.

Although open data are not ideal for academic research, they are useful for pilot studies and promoting collaborations between agencies and researchers. Readily available data allow local communities to identify public health problems, make evidence-based decisions, and mobilize community support. Consumer benefits include health education and improved decision-making, such as selecting medical providers. The public may appreciate the value of public health services including surveillance activities; one participant explained, "Our little dream [is] about people really understanding what local public health and state public health agencies do . . . for decision-making [and] investments in public health programs."

Reaching new audiences is critical to achieving goals. Academic researchers are longstanding data customers, but open data can target new audiences with fresh ideas: "The smartest people around a particular problem probably don't work in your agency. . . . If you can get the problem out there for other people to think about . . . - that's tremendous." To facilitate innovative data uses, agencies work with external partners such as the Health Data Consortium to organize challenge competitions, and the NYSDOH public affairs group directs journalists to Health Data NY. Although competitions have produced few widely used health-related applications, there are striking examples of data journalism, such as the Chicago Tribune's crime statistics visualizations and New York news stories with maps of local childhood obesity rates.

When describing their hopes for open data, many respondents expressed uncertainty about specific uses but stressed that this uncertainty makes open data exciting. Examining data across sectors may generate discoveries as the volume of data increases. Changes in the health care environment may expand opportunities for data-based applications. One participant summarized, "The next five to ten years are going to show an incredible set of leaps forward in terms of cost containment, efficiency in healthcare, utilization of personal health data . . . to create a more comprehensive set of information about us as individuals in our healthcare system at large. So the opportunity is there. It's going to be fun to see what happens next."

Challenges of opening government health data

While open data provide numerous opportunities, there are various barriers to their publication, summarized in Table 3.

Almost all participants reported human resource constraints. At NYSDOH, the governor's early retirement incentive program reduced the workforce by approximately 10%, creating gaps in institutional knowledge and capacity. Although NYSDOH is a large organization, many staff are grant-funded and cannot work on unrelated projects. Some participants asserted that different technical skills are necessary for releasing open data, such as data management and writing for lay audiences, and that New York's consolidation of information technology staff into 1 agency has generated skill gaps. There was limited consensus on whether these issues affect data publishing.

Cultural resistance was universally mentioned as a critical barrier. Participants frequently discussed the pervasiveness of data silos, which reflect the following: historical legacies such as New York City being separate federal reporting jurisdictions; disease-specific funding streams, legal restrictions on data sharing, and data system structures; fears that shared data will be misinterpreted, compromise confidentiality, or increase workload; and feelings of ownership. Open data is a new public health business model, as data production previously entailed collecting and synthesizing data, publishing reports, and providing data upon request. One respondent explained, "There's culture change that has to coincide with making data available. People take a lot of time to accumulate the data that they have for their specific project or purpose and it's not often a first reaction for someone to say, 'Oh sure, I'll turn around and make that data open.'" Most participants had already perceived cultural changes, although they cautioned that overcoming cultural resistance is difficult, newer attitudes about data sharing have not fully permeated agencies, and backsliding could occur under less supportive administrations.

The legal review process was another major theme. Overlapping federal and state regulations govern stewardship and the collection and release of health data. Data stakeholders must understand how to protect health information, but many respondents found this knowledge elusive. For example, the Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule allows health information to become designated as "de-identified" through Expert Determination (whereby individuals with sufficient knowledge and experience determine re-identification risks to be negligible) or Safe Harbor (listing 18 identifiers including names, dates, and geographic information to remove or report at high aggregations, such as 3-digit zip codes). Expert Determination has unclear federal guidance: "As much as there's a national movement toward transparency in open data, there's also a lack of guidelines in what constitutes an expert for Expert Determination. I keep hoping . . . for Expert Determination certification or guidelines or training." As more data become available, statistical approaches to assess re-identification risk may evolve to consider the "mosaic effect," the increased re-identification risk when combining multiple datasets. While some participants desired standardized approaches, others cautioned that one-size-fits-all de-identification methods are impossible because of heterogeneous dataset characteristics and legal restrictions such as not using administrative data for nonprogrammatic purposes and statutory regulations on certain datasets. Complex layers of federal, state, and programmatic rules, unclear federal guidance on Expert Determination, and evolving techniques to assess re-identification risk have engendered tensions between data owners wanting to publish their data and legal staff applying existing regulations to protect agencies from breaches. Not all participants perceived a combative relationship; one explained, "The lawyers were the chief barriers [to releasing outpatient hospital discharge data]. And they were rightly so. They were very concerned about re-identification of data. They even did re-identify one data set and saved us from trouble, but they eventually became the champions of it too."

The last major barrier was ensuring high-quality and timely data. All participants expressed a desire to prioritize high-quality datasets over large quantities. However, assuring quality is difficult. To improve its metadata, the Health Data NY team developed detailed forms for data owners. However, NYSDOH has no centralized data management system, documentation varies internally, and metadata preparation takes time. De-identification lowers value; for example, datasets cannot be linked at the individual level, and there are high levels of aggregation. There is limited guidance on balancing value and re-identification risk. Another common concern is that datasets use different definitions and syntax for data elements, making standardization and interoperability difficult to achieve. Although staff members invest considerable resources into cleaning data, agencies that rely on local partners' data have limited quality control.

Several other challenges were frequently mentioned, although not perceived as insurmountable. Technical issues included

Table 3. Challenges encountered when opening government health data[a]

Challenge	Illustrative examples
Human resources	• Reduction in civil service workforce, creating knowledge gaps on specific data sources and limited staff time • Restrictions on ability to reassign grant-funded staff from specific projects to open data activities • Different technical skills required to release open data • Capacity to support information technology within agencies reduced during New York's transition to a single state agency for information technology
Cultural resistance	• Breaking down of internal data silos • New business model for creating and publishing data • Tension between posting public use files to existing program websites and new open data portals
Legal and regulatory issues	• Data release governed by complex set of overlapping federal and state rules and regulations • Only data owners have authority to release data, even if multiple groups use the data to generate reports
Data and metadata quality	• Data must be high quality and timely, and with clear metadata • Tension exists between maintaining value while minimizing disclosure risks • A lack of standard definitions for data elements has made interoperability between datasets extremely difficult to achieve • Agencies relying on local partners for data collection have less control over data quality • There is a need to develop methods to alert end users to data updates
Technical	• Extracting data from legacy software systems • Demand for more sophisticated capabilities from open data platform software, including more analytic tools and queryable interfaces • How to make open data interfaces more usable by different end users • Technical errors when uploading data to open data platforms, particularly for large datasets • Demand for tools to automate the upload process to refresh datasets as they are produced
Knowledge gaps among agency staff and leadership	• Understanding the goals and activities of the open data team • How to use open data platform technology • Methods to appropriately de-identify data and perform expert determinations • Identifying different end users and their data needs
Addressing the needs of end users	• End users must be trained to use the platform to discover data, conduct analyses, and interpret findings appropriately • Open data platforms must meet the needs of multiple audiences with different demands and skills
Political	• Having supportive executive leadership is important • Postings may cause tensions among some interest groups

[a]*Source:* Authors' interviews with policymakers and practitioners.

problems extracting data from legacy systems, limited analytic functions for users such as query tools, difficulty navigating interfaces, and inefficient processes to automatically upload new data. Several staff members expressed uncertainty over the Health Data NY team's activities and how to prepare data and interact with the open data platform. The largest area of uncertainty was regarding the identities of end users and their data needs. Participants acknowledged that their end users have different skills and motivations, but were uncertain how to adapt their data products for multiple audiences.

Successful strategies for opening government health data

When sharing successful strategies, 6 themes emerged; all were described as important. These themes covered leadership, funding, managing culture, improving the legal review, and making data releases higher-impact and more efficient.

First, active executive leadership and internal champions are crucial: "Leadership has to be there. It's not just leadership saying, 'Here, do this.' It's active leadership." Successful executive leaders develop and maintain strong visions for their open data initiative, and establish clear governance processes to execute data releases. Participants from all jurisdictions asserted that New York's governance, with an open data handbook articulating a vision, guidance for releasing data, and technical specifications, sets the state apart. Other jurisdictions, including California, have recognized the handbook's value and adapted it for their environment.

Second, sufficient resources are required to develop and sustain platforms: "Open data is not a free endeavor. It takes time and money to make data open." Successful jurisdictions have designated open data teams, and data owners need time to prepare datasets. The technical infrastructure needs continued funding, especially to support new desired capabilities such visualizations and query tools. As open health data initiatives are longer-term propositions, resources must be maintained: "If you're going to set something up, you don't want to do it for one or two years and then resources die and boom . . . support [it] on an ongoing basis."

Third, participants offered suggestions to overcome cultural resistance. Leadership must understand that open data is not a technical issue. Managing culture requires recognizing data-sharing attitudes and taking actions to establish buy-in (such as staff meetings to explain the initiative's purpose), identifying datasets offering immediate value to staff and external audiences, creating collaborative data release processes whereby open data staff and data owners jointly identify and prepare datasets, and providing status reports on the impact of data. In managing culture, leadership should understand when to protect data for confidentiality and "strike that balance between not under-protecting our personal data, but not overprotecting the data either."

Fourth, although the legal environment will remain complex, jurisdictions can facilitate legal reviews. Properly de-identifying data requires technical knowledge and tools to streamline this process; several NYSDOH respondents researched federal guidance and soft-

ware solutions, although they offered no recommendations for other jurisdictions. Transparent legal review processes are important for management. This includes clarifying legal considerations, such as "a better understanding of what [the legal department does]...what legal analysis is, and why legal makes these determinations and the fact that we don't make the law, we just apply it."

Early innovators had several strategies to increase impact. They tailor data products to their audiences: "Think up front about how the public will receive data, [have an] understanding [of] how that data is going to be used and...have the data stewards who are the experts in the data try to think about how the data will be used, and interact with the data and design their publishing system to accommodate that." They suggest starting small, with 5 to 10 "low-hanging fruit" datasets that are readily de-identifiable, not politically contentious, easy to use, and of wide interest. Aggressively promoting their use to demonstrate value can create a tipping point whereby demand for additional datasets increases and staff members experience tangible benefits. Successful open data teams have established feedback mechanisms to track progress and identify how data are used, ways to improve value, and how to prioritize future datasets for release.

Finally, respondents strongly cautioned against developing technology platforms and management processes de novo. Rather, jurisdictions should seek out, learn, and apply lessons from experienced jurisdictions: "If you can take an existing data system and modify it for your own needs...[rather than] build something from the ground up because you convince yourself that your needs are unique and you're special, I would caution folks to resist that urge. Sometimes you've got staff on hand who are programmers and they're here to stretch their wings, but that is often not the most cost effective way and you're not gaining the lessons learned from someone else who's built the system already." NYSDOH participants recalled a quick start after their commissioner invited DHHS leadership to visit and explain the open data concept and its value propositions. They subsequently adapted the DHHS model.

DISCUSSION

While open data platforms are emerging across the country, we interviewed NYSDOH and DHHS policymakers and practitioners about the value of releasing health data and required capabilities. They reported that open data may improve the triple aim of health care quality, lower cost, and improved population health.[14] Benefits to government agencies include enhanced data quality and more efficient operations. External benefits include improved health literacy, data-driven changes in health care delivery, consumer engagement, and community empowerment. Although open government initiatives are intended to promote government transparency, this benefit was discussed the least. The comments about improved data quality, timeliness, and usefulness signal that open data is changing data production and management practices more broadly beyond its policy goals. Releasing open data was not perceived as a technical issue; rather, critical challenges are resources, cultural resistance, navigating legal and regulatory issues, and data quality. The open data movement will likely continue, but success requires sustained leadership, resources, cultural change, promoting data use, and governance. Findings from other innovative jurisdictions were consistent, suggesting that the NYSDOH and federal experiences are generalizable.

These results provide several contributions to the literature. Open data platforms are in their infancy, and jurisdictions creating new portals can incorporate lessons from early innovators. Many recommendations, such as fostering strong executive leadership, establishing transparent governance processes, adapting existing tools, developing strategic plans, monitoring progress, and making human resource investments, are unsurprising and consistent with other domains.[15–18] There is growing literature on how open government improves transparency, but less attention to how it improves government. There are some published use cases, such as use of nursing home bed census data for New York City's Hurricane Sandy response, increased media coverage of childhood obesity in New York, and improved planning and citizen engagement around pedestrian injuries in San Francisco.[3,19] Our interviews supplement these use cases by describing additional benefits and anticipated returns on investment.

Our findings suggest several recommendations to increase open data's impact. The first is to start small, with "high-value," noncontroversial datasets that are easy to release. End-user feedback is critical to prioritize future datasets and tailor data products to consumers. Incorporating public health messaging into data releases can increase data awareness and encourage their use. Continued efforts to improve data quality and database design can enhance value and streamline the data publication process. Code-a-thon challenges and other community events can engage users and encourage innovation. To expand their scope, professional organizations such as the Association of State and Territorial Health Officials and the National Association of County and City Health Officials, which are not yet visible open data stakeholders, could actively promote use among public health practitioners.

Early innovators' experiences also suggest ways to improve the sustainability of open health data ecosystems. Executive leaders and champions should foster organizational cultures that are friendly to open data principles. Stakeholder analysis to identify relevant actors, knowledge and beliefs about open data, and potential misunderstandings or oppositions can improve the success of implementation.[20] The NYSDOH open data team worked closely with data owners to secure buy-in, improve the data release process, and identify ways to promote data resources. Open data handbooks are critical to communicate visions and establish governance, and New York's handbook can be a guide.[21] Standardized metadata templates and expert determination forms can improve workflow efficiency and establish data and metadata quality standards. Leaders must commit to sustained investments in human resources and technology. After developing an open data site, moving data publication activities into an established program area can routinize them and protect them from budget cuts. Finally, open data managers should collect ongoing feedback from end users to improve data products and communicate early successes to agency staff and the public.

An important direction for the field is updating legal guidance on protecting health data in the open data era. Participants frequently discussed the challenges of legal reviews and expert determinations, because they require deep expertise in overlapping laws and regulations governing health data and technical skills to evaluate disclosure risk. Federal guidance is nonprescriptive because datasets are unique, yet many participants desired clearer guidance for de-identifying data and determining appropriate levels of risk. These issues will become more complex as increased volumes of data create mosaic effects whereby traditional de-identification methods may become obsolete. Reconsidering existing legal guidance, including potentially modifying Safe Harbor data elements such as geographic aggregation, and improving de-identification techniques may balance value and risk tradeoffs.

This study has several limitations. These cases only include early innovators, and although a range of participants were recruited, individuals with optimistic views are overrepresented. Future work could seek counterarguments against open data and test whether the factors that early leaders attributed to their success (such as leadership and workforce capacity) are absent from jurisdictions slow to release data. As participants were recruited to ensure diversity, the output is a description of the potential benefits, challenges, and lessons learned and should not be interpreted as representative beliefs. Future nationally representative quantitative surveys could explore how knowledge, attitudes, and beliefs vary within and across agencies and jurisdictions with varying political support and institutional capacity. There may be researcher bias, which we minimized by having 2 researchers with different expertise attend interviews and synthesize data, and inviting participants to review findings.

The open data movement has a bright future but unknown long-term impact, which is partly what makes it so exciting. To maintain momentum and ensure success, it is critical to quantify the value of data. Developing and maintaining open data platforms is costly, many benefits will not occur for years, and taxpayers and policymakers may demand evidence that releasing open data is an efficient use of resources. Documenting process outcomes such as reduced Freedom of Information Act requests is straightforward, but our interviews indicate that this is the tip of the iceberg. To start to quantify return on investment, researchers and practitioners should publish use cases on how open data has advanced the triple aim objectives.

CONTRIBUTION

FUNDING

COMPETING INTERESTS

The authors have no competing interests to declare.

ACKNOWLEDGMENTS

REFERENCES

1. Open data handbook. *Open Knowledge Foundation Website* opendata-handbook.org/guide/en/what-is-open-data. Accessed January 15, 2016.
2. Martin EG, Helbig N, Birkhead GS. Opening health data: what do researchers want? Early experiences with New York's open data portal. *J Public Health Manag Pract* 2015 Sep–Oct; 21(15): E1–E7.
3. Martin EG, Helbig N, Shah NR. Liberating data to transform healthcare: New York's open data experience. *JAMA* 2014;311(24):2481–2482.
4. Orszag PR. Memorandum for the heads of executive departments and agencies: open government directives. www.whitehouse.gov/sites/default/files/omb/assets/memoranda_2010/m10-06.pdf. Updated December 8, 2009. Accessed December 10, 2015.
5. Health data all-stars. Health Data Consortium. http://allstars.healthdata consortium.org. Accessed December 10, 2015.
6. Harrison TM, Pardo TA, Cook M. Creating open government ecosystems: a research and development agenda. *Future Internet* 2012;4(4):900–928.
7. Cuomo AM. Executive order 95: using technology to promote transparency, improve government performance, and enhance citizen engagement. http://www.governor.ny.gov/news/no-95-using-technology-promote-transparency-improve-government-performance-and-enhance-citizen. Updated March 11, 2013. Accessed December 10, 2015.
8. Bradley EH, Curry LA, Ramanadhan S, Rowe L, Nembhard IM, Krumholz HM. Research in action: using positive deviance to improve quality of health care. *Implement Sci* 2009;4:25.
9. Klaiman T, O'Connell K, Stoto M. Local health department public vaccination clinic success during 2009 pH1N1. *J Public Health Manag Pract* 2013;19(4):E20–E26.
10. Corbin J, Strauss A. *Basics of Qualitative Research* 3rd ed. Thousand Oaks, CA: Sage; 2007.
11. Patton MQ. Enhancing the quality and credibility of qualitative analysis. *Health Services Res* 1999;34(5 Part II):1189–1208.
12. Charmaz K. *Constructing Grounded Theory* 2nd ed. Los Angeles, CA: Sage; 2014.
13. Maxwell JA. *Qualitative Research Design: An Interactive Approach* 2nd ed. Thousand Oaks, CA: Sage Publications; 2005.
14. Shaw E. The impact of open: keeping you healthy. Sunlight Foundation. https://sunlightfoundation.com/blog/2014/06/25/the-impact-of-open-keeping-you-healthy/. Updated June 25, 2014. Accessed December 10, 2015.
15. Rainey HG. *Understanding and Managing Public Oranizations* 5th ed. San Francisco, CA: Jossey-Bass; 2014.
16. Quinn RE, Faerman SR, Thompson MP, McGrath MR, St. Clair LS. *Becoming a Master Manager: A Competing Values Approach* 4th ed. Hoboken, NJ: Wiley; 2006.
17. Government Accountability Office. *Library of Congress: Strong Leadership Needed to Serious Information Technology Weaknesses* (GAO Publication No. 15-315) Washington, DC: US Government Printing Office; 2015.
18. Government Accountability Office. *Farm Modernization Program: Farm Service Agency Needs to Demonstrate the Capacity to Manage IT Initiatives* (GAO Publication No. 15-506) Washington, DC: US Government Printing Office; 2015.
19. Wold C. In plain sight: is open data improving our health? California Health Care Foundation. http://www.chcf.org/publications/2015/01/in-plain-sight-open-data. Updated January 2015. Accessed December 10, 2015.
20. Schmeer K. *Guidelines for Conducting a Stakeholder Analysis* Bethesda, MD: Partnerships for Health Reform, Abt Associates Inc.; 1999.
21. Open data handbook: New York state open data initiative. New York State Office of Information Technology. http://ny.github.io/open-data-handbook/. Accessed December 10, 2015.