A research roadmap to advance data collaboratives practice as a novel research direction

Susha, Iryna; Pardo, Theresa A.; Janssen, Marijn; Adler, Natalia; Verhulst, Stefaan G.; Harbour, Todd

A Research Roadmap to Advance Data Collaboratives Practice as a Novel Research Direction

Iryna Susha, Örebro University, Örebro, Sweden

Theresa A. Pardo, CTG, University at Albany, SUNY, USA

Marijn Janssen, Delft University of Technology, Delft, The Netherlands

https://orcid.org/0000-0001-6211-8790

Natalia Adler, UNICEF, USA

Stefaan G. Verhulst, The Governance Lab, New York University, USA

Todd Harbour, New York State, USA

ABSTRACT

An increasing number of initiatives have emerged around the world to help facilitate data sharing and collaborations to leverage different sources of data to address societal problems. They are called "data collaboratives". Data collaboratives are seen as a novel way to match real life problems with relevant expertise and data from across the sectors. Despite its significance and growing experimentation by practitioners, there has been limited research in this field. In this article, the authors report on the outcomes of a panel discussing critical issues facing data collaboratives and develop a research and development agenda. The panel included participants from the government, academics, and practitioners and was held in June 2017 during the 18th International Conference on Digital Government Research at City University of New York (Staten Island, New York, USA). The article begins by discussing the concept of data collaboratives. Then the authors formulate research questions and topics for the research roadmap based on the panel discussions. The research roadmap poses questions across nine different topics: conceptualizing data collaboratives, value of data, matching data to problems, impact analysis, incentives, capabilities, governance, data management, and interoperability. Finally, the authors discuss how digital government research can contribute to answering some of the identified research questions.

KEYWORDS

Data Collaborative, Data Philanthropy, Data Sharing, Digital Government, Evidence Based Policy, Public Private Partnership

1. INTRODUCTION

The world's most complex and 'wicked' socioeconomic problems - from climate change to the spread of epidemics - cannot be tackled by a single authority in the public sector alone. It takes a village to understand and unpack these complex problems in order to make them understandable and actionable. Data can help, especially considering its exponential growth in the past few years. Access to new datasets and expertise can help provide a more accurate and comprehensive picture of the situation and empower decision-makers to tailor their interventions accordingly. However, much of the most relevant data that can help untangle these complex problems (and the expertise required to make sense of that data) often reside with corporations in the form of web clicks, like buttons, cell phone data, satellite data, etc. As a result, an increasing number of initiatives have emerged around the world to help facilitate data sharing and collaborations to leverage different sources of data to address societal problems. They are called *data collaboratives* (Verhulst and Sangokoya, 2015) and can be defined as "cross-sector (and public/private) collaboration initiatives aimed at data collection, sharing, or processing for the purpose of addressing a societal challenge" (Susha, Janssen, & Verhulst, 2017, p. 2691).

Data collaboratives is a novel research topic which expands the horizons of new approaches of tackling public problems and improving public services through data science and collaborative solutions. There is growing experimentation with data collaboratives (as mapped by The GovLab at datacollaboratives.org), however there is a lack of common terminology and shared understanding of the novelty and complexity of this new phenomenon. Therefore, we have convened a panel that brought together researchers and practitioners working on data collaboratives to showcase their results and discuss critical issues with the digital government community. The panel was held during the 18th International Conference on Digital Government Research on 8 June 2017 at City University of New York (Staten Island, New York, USA). It featured five speakers: a government official, two practitioners, and two academics (all co-authors of this paper). The panel was moderated by the first author of the paper. In this paper, we report on the outcomes of this panel with the aim of identifying a research agenda to help build bridges between research and practice in this area.

2. DATA COLLABORATIVES AS A NOVEL RESEARCH DIRECTION

Digital government fosters the use of information and technology to support and improve public policies and government operations, engage citizens, and provide comprehensive and timely government services. The global digital government research community is interested in the development and impact of digital government. The community exists at the intersection of computer and information science, social and behavioral science, and focuses on the needs and problems of government. Building new knowledge about data collaboratives as a complex phenomenon is in line with the interests of this community.

There has been a lot of work done in the field of data sharing in the digital government community (e.g. Landsbergen & Wolken, 2002; Gil-Garcia & Pardo, 2005) and only recently the focus has shifted towards the inclusion of private parties (Bharosa, et al, 2013). The same is true for work related to information sharing and collaboration among government organizations. Tung-Mou and Maxwell (2010, p.73) reviewed literature for factors affecting information sharing, which include promotion of a culture of information stewardship as opposed to ownership; strong leadership support to information sharing efforts; legislative and regulatory mandates; reward systems that promote information sharing both within and across organizations; the establishment of shared goals; and the development of ongoing trusted relationships based on mutual understanding of needs and concerns and shared responsibility. Gil-Garcia, Chun, and Janssen (2009) discussed the challenges to government information sharing and integration and grouped them in technical, organizational, political, and legal categories. Data collaboratives introduce new complexities to these collaborative engagements, with

the private sector playing a bigger role and the need for new structures, procedures, processes, and practices to shape new ways of working.

Data collaboratives can create value in five different ways: improving situational awareness and response, improving public service design and delivery, enabling forecasting and prediction based on the data, supporting evaluation and impact assessment of policies, as well as contributing to knowledge creation and transfer between the sectors (Verhulst and Young, 2017). In a data collaborative, 'doing good' for the society by contributing to solving societal problems is the central goal, and providing datasets and expertise should help to realize this goal. A key element to accomplish this is the collaboration and interaction between people, organizations, and communities. Apart from the contribution by organizations, the quality of the results is a key aspect as the reputation of both private and public sector organizations should not be harmed. This brings us to the following main features and principles underlying data collaboratives:

1. Defining a societal problem and considering how data can contribute to it;
2. The need to collect, share or exchange data among diverse information systems;
3. Processing capabilities by having some kind of shared infrastructure;
4. The need for mobilizing expertise that are often not available by a single organization;
5. Domination by private organizations in which governments might play a marginal role;
6. Quality controls for ensuring correct and accurate results of data analysis;
7. Sharing the results in a broader community and stimulating next steps.

An important tension in data collaboratives is that there is a large dependency between the resources and capabilities of different parties, moreover parties might have different interests and incentives for collaboration. The solving of a societal problem is a binding factor, and a collaboration might work only once for a certain purpose and not for other purposes. Collaboration might be further complicated due to heterogeneous datasets and systems and various levels of readiness. Also, the cost might not be evenly shared and might be hard to access. The interdependence creates new complexities and uncertainties.

Data collaboratives encompass different kinds of technologies, such as big data, advanced analytics, visualizations and require a change of roles and new capabilities from government (Janssen & Helbig, 2016). Public sector organizations might only be one of the many players and have no hierarchical or legislative control over the other organizations. Yet, they might be the most interested in the outcomes of the analysis and most equipped to take actions on the basis of the analysis. At the same time, they should trust the data and the outcomes and also be involved to ensure the acceptance of the outcomes and the taking of further steps.

3. A REAL-LIFE CASE OF DATA COLLABORATIVE

The panel highlighted many different societal challenges which have the potential to be addressed through data collaboratives: disaster response, childhood obesity, urban mobility, gender issues, environmental monitoring etc. The presentations from UNICEF and The GovLab showcased concrete examples of applied collaborative solutions to real-life issues thanks to data collaboratives.

One of the examples highlighted was the growing number of suicides among young people in India. The problem was selected for two reasons. First, because suicide is a complex phenomenon that cannot be tackled by one organization alone, hence the need for collaboration. Second, current data collection methods are insufficient to understand the magnitude and drivers of the problem. Suicide attempts and suicidal ideation are grossly underreported by the Indian National Crime Records Bureau and drivers such as cyber-bullying, peer pressure, examination failure, forced marriage, conflict over dowries, and indebtedness are only identified and reported in individual instances after the fact. According to a report based on verbal autopsy, suicide is the second leading cause of death among

people aged 15-24 in India, accounting for about 60,000 deaths in 2013. Therefore, there was room for experimenting with novel data approaches.

UNICEF HQ and The GovLab worked with the UNICEF India Country Office toward launching a data collaborative to better understand this growing phenomenon and unlock access to more information about adolescents and youth with mental health issues and other indicators of suicidal tendencies. The goal of the data collaborative was to gain unique insight that can help UNICEF's efforts to spread awareness about mental health care and related advocacy aimed at suicide prevention. With buy-in from the UNICEF India Country Office, the teams in New York proceeded to scope existing literature and frameworks to develop a problem definition and subsequent data audit to shed light to emerging trends of suicide in India.

Data collaboratives should be problem-driven as opposed to data supply driven. It builds on well-established approaches that look at problems as springboard to address complex problems (Cameron, 1986; Oliver, 1992; Seo and Creed, 2002). Taking a page from a Problem Driven Iterative Adaptation (PDIA) framework (Andrews et al, 2015), the data collaboratives methodology developed by The GovLab focuses on locally nominated problems, expanding the authorizing environment to include not only the public sector, but several collaborators providing data and expertise. To attract these collaborations, the development of a problem-driven case study on suicide trends in India serves a twofold goal. First, it helps break down a complex social problem into more manageable parts that can be easily understood by a multitude of partners, including those in the private sector. Second, it makes a compelling case for action that helps mobilize data-driven collaborative solutions to tackle this 'wicked' problem.

Once the problem was defined, UNICEF HQ and The GovLab reached out to data scientists to brainstorm technically feasible data science solutions for the underreporting of suicide trends in India. Researchers from the Institute of Scientific Interchange Foundation (ISI Foundation), based in Turin, Italy, agreed to join the efforts through multiple brainstorming meetings. Several data approaches and data sources were identified, mostly related to online browsing behavior and other internet activity using social media, chat room messages, and internet search queries, all of which could help shed light to users' suicidal thoughts/ideation.

Following the data-audit, the next step was to reach out to corporations that had the available data in hand. Researchers from Microsoft Research Center, based in Israel, joined the data collaborative providing additional expertise and data based on anonymized and aggregated queries submitted to the Bing search engine and Yahoo Questions by users in India.

The ongoing research seeks to document the value of data collaborative by focusing on how to complement existing official baselines that tend to underreport suicide by looking at the correlation between the number of search queries of predefined keywords related to suicide and the number of official suicides records per state. Subject Matter Experts on suicide in India (mostly from NGOs so far) are periodically invited to provide feedback to these research findings and government officials will be involved as the work evolves. The research is expected to be finalized before December 2017, after which decision-makers in India will be contacted to work on implementation.

4. RESEARCH ROADMAP FOR DATA COLLABORATIVES

Below we present the critical issues and questions which were identified based on the panel discussions. We grouped them into *nine topics* which are explained in more detail below.

4.1. Conceptualizing Data Collaboratives

A digital government research agenda focused on data collaboratives should pursue questions about data collaboratives themselves: what is a data collaborative, what forms do they take, and how are they different from ongoing practice? New explanatory and descriptive research is needed to understand this phenomenon more fully. We need to understand what is new about data collaboratives, what

sets them apart from other public sector and private sector data sharing and integration activities and public private partnerships, how and in what way they vary by policy domain, by partners, by region, and what is the role of government. This also involves examining what new questions can be answered and what new methods can be used in data collaboratives. New understanding of the role of various actors in the formation and operation of a data collaborative is required; the idea is not where the data sits – but who is part of the conversation.

4.2. Value of Data

A digital government research agenda focused on data collaboratives must examine questions about the value of data itself as an individual, organizational, and societal asset. Further, questions about whether the growing recognition of the value of data to help solve societal grand challenges is enabling or constraining data collaborative formation, resiliency, and sustainability are important. For example, will growing recognition of the potential to monetize private data constrain data collaboratives or not?

4.3. Matching Problems to Data to Partners

A research agenda should pursue questions about how data collaboratives are formed around the identification of shared interest in a problem and the matching of problems (demand) to those with data and expertise (supply). In the case of the data collaborative in India, data scientists and data-driven companies were contacted via personal connections of UNICEF HQ and The GovLab. How does one move beyond these interpersonal connections to set up a data collaborative?

Even when the contact with potential data-holders and experts is established, how does the negotiation take place? Saying to potential partners "we need your data" is not enough to get it, especially considering how this asset is increasingly becoming more and more monetized and is being referred to as "the world's most valuable resource" (The Economist, 2017). As we have seen, grounding the need for data on a compelling problem-driven case studies helped UNICEF and The GovLab mobilize and convince partners to share their data and expertise pro bono. But how can this be replicated at scale? One of the biggest obstacles for engaging the private sector in data collaboratives is a widespread lack of capacity among the demand side to break down complex problems into smaller, better defined and more actionable challenges that can be easily understood by corporations willing to share data and expertise. New tools and techniques to understand all the various underlying aspects of this complexity and uncertainty of relevant problems may be needed, as well as new tools for conducting analyses of cost and risk for sharing and for not sharing.

Further, questions about the nature of such partnerships, for example, how to find the balance between sharing more data with more people with sharing some data with some people that is less open, more targeted, and created high value, must be pursued. Tools and techniques to model problems and systems are needed so that new understanding of how and where new kinds of data can help solve those problems is needed. In other words, we need to find ways to quickly identify what data sources are relevant to the real-life problems – how to find them, test their relevance and utility, and then make the data fully available. Finally, how can new understanding of problems, matching problems to data, and securing partners who are committed to the principles of a data collaborative happen quickly so that the resulting capability can contribute to communities in crisis?

4.4. Impact Analysis

A digital government research agenda focused on data collaboratives should pursue questions about the impact of data collaboratives. As an example, in the case of the data collaborative in India, the partners faced the challenge of how to translate the insights generated into actionable recommendations. How can we determine impact from data collaboratives? What kind of public value is being or can be created through data collaboratives? What role can corporate data play in helping governments meet their mission?

4.5. Incentives

From the practitioners' point of view, there are certain incentives at play for corporations to share data, such as corporate social responsibility and the chance to benefit from external data science expertise (Verhulst and Young, 2017). However, such incentives often conflict with the urgency and goals of the government agencies or research organizations requesting the data. Accessing corporate data on a short notice, for example when a disaster strikes, is difficult and may incur high costs. Moreover, it is not always obvious who has the relevant data for a particular problem and how to approach these data-holders outside interpersonal networks. Data is often guarded and over-protected by data-owners due to, among other reasons, keeping the organization's competitive advantage.

Various stakeholders need to work together and they need incentives for doing this. New knowledge is needed about why different actors would share data. What incentive structures drive data collaboratives? Are new incentives required? How do different structures incentivize different partners? How best to incentivize data collaborative formation, what conditions will contribute to value creation through a data collaborative model, and how data collaboratives can be sustained over time rather than formed as a consequence of a crisis? These are all related questions. For example, in the case of the data collaborative in India, one of the challenging questions was how to make the engagement sustainable, on-going and generalizable and how to scale up the effort and go beyond the progress made. Besides, how to facilitate a productive dialogue between different communities (such as corporate data scientists and social workers) with different cultural backgrounds and incentives?

4.6. Capabilities

A digital government research agenda focused on data collaboratives should build new understanding of the capabilities required to identify the need for a data collaborative, to launch that collaborative, and use it to create public value that can be sustainable over time. The agenda should include questions about capabilities required to ensure data is ready when needed, actors having the analytics skills necessary to leverage the availability of new data in support of a societal challenge, about capabilities needed to address information policy issues and technical infrastructure requirements. Additional questions include: What capabilities are required in what contexts? What role does trust play and how does the capability to build trust influence data collaborative value creation? How does the collaborative model create new capability models? Is it possible to create clarity around roles and responsibilities in building the data collaborative and creating value? Are there different capabilities needed to create an incident focused data collaborative versus a sustainable data collaborative or a problem- versus systemic capacity focused one?

4.7. Governance

For data collaboratives governance is about the distribution of decision-making authorities to ensure the proper working of a collaboration. Governance can be viewed as the defining and allocating of actions and decisions that ensure a form of collaboration that cannot be externally imposed (Stoker, 1998). A digital government research agenda focused on data collaboratives must examine the ways in which governance plays a role in building and sustaining data collaboratives. In particular, what data governance models work for data collaboratives in different contexts? What data governance models are best for the organizations who are contributors to data collaboratives – the supply side; how does the demand side influence data governance in supply organizations? Existing research highlights the importance of establishing rules, principles, and standards and putting in place the elements of responsible data governance. What are those rules, principles and standards? What are the elements of responsible data governance vis a vis a data collaborative? Issues of data ownership should also be researched, especially with this asset containing vital information that can help public interest (i.e. as in the case of epidemics). To what extent can data sharing change from voluntary to obligatory? Questions about data governance as an enabler to data collaboratives must be examined,

in particular the impact of data governance in getting data ready and fit for use within the often crisis conditions around which data collaboratives emerge.

4.8. Data Management

A digital government research agenda focused on data collaboratives must explore the role of data management and how and to what extent related capabilities and the use of best practices enable data collaboratives. Who owns which data? What is the quality of these data? For instance, in the case of the data collaborative in India, the partners faced a challenge of how to assure the data is representative. Besides, new understanding about how to create structured data out of unstructured data and make that newly structured data available as an asset within the data collaborative is needed.

Organizations have yet to fully realize the potential that their data holds. Despite ample evidence to the contrary, data often remains siloed, undiscoverable, and inaccessible by the enterprise. For organizations to begin using data as a critical decision-making asset, the organization must first understand that it needs to treat data as an asset, something that is inherently valuable and needs to be managed in a deliberate and explicit manner. It is important to note that organizations should not wait until things are settled, stable or finished before starting to manage their data. For those organizations that do not develop data strategies and learn to manage their data, they will painfully learn there will always be a good reason for not making changes to their organization and its behavior. This lack of organizational fortitude, commitment and action will have detrimental effects on the organization and its ability to continue meeting the demands of an ever-evolving, competitive environment.

Fundamentally, organizations need to develop an integrative discipline for structuring, describing and governing information data assets, regardless of organizational and technological boundaries, and to improve operational efficiency, promote transparency and enable business insight. No matter how an organization acquires data—through sharing relationships with government or industry (data collaboratives), commercial acquisition, or created using organizational resources—data governance is the foundational function to help ensure that the equities of all parties are protected.

4.9. Interoperability

Interoperability is the ability of diverse systems to collaborate with each other (Scholl & Klischewski, 2007). Interoperability is not only a technical issue, it also applies to organizational aspects (Goldkuhl, 2009). Data is stored in different formats and in heterogeneous systems which might hinder its easy use. Much efforts might be required to derive data from the systems and to link data and to analyze the data.

Recent advances in data technologies have resulted in significantly improved ways of interacting with organizational data assets—despite the vagueness of the big data promise. In other words, big data technologies and industry specific data scientists can provide organizations some of the resources required to do more with data, but organizations still need a well-defined target and an understanding of where they are relative to that gap. This can be improved by recognizing that we need to teach more data management principles as part of core data science curricula and to partner data scientists with capable and qualified data management professionals. Finally, while organizations long to exploit their data assets and perform advanced analytics, organizations need to realize that they must first crawl, then walk, then run about data and, if organizations truly want to trust the results of computational abilities, they must be able to account for data across the entire data lifecycle—from acquisition through final disposition.

Table 1 below summarizes the research questions and topics discussed above into a data collaboratives research roadmap.

To summarize, questions of mutual interest to those seeking to create new public value through data collaboratives and the digital government community range from the more technical to the more social, including what forms do data collaboratives take and why, what impact are they having and why, what incentives are driving participation in collaboratives and what factors influence the success and sustainability of data collaboratives.

Table 1. Research roadmap for data collaboratives: summary of research questions and topics

Topic	Research question
Conceptualizing data collaboratives	● What sets data collaboratives apart from more traditional public-private partnerships focused on data sharing or integration? ● What forms do data collaboratives take and how do these vary by policy domain, by partners, by region? ● What should be the roles of various actors, including government, in the formation and operation of data collaboratives?
Value of data	● What are the tensions between the value of data as an individual, organizational, and societal asset? ● How does the recognition of these different types of value of data enable or constrain data collaboratives practice? ● How is it possible to reconcile the value of monetizing private data with the societal value of using them for public good?
Matching data to problems to partners	● How should data collaboratives be formed? ● What tools or techniques can be used to 'break down' or model complex problems to understand where new data may be useful? ● What tools or techniques can be used to identify relevant data and test their utility for a certain problem? ● How can the costs and risks of sharing and not sharing data be assessed? ● What data sharing mechanism is optimal in a certain situation and can maximize the value of data? ● How is it possible to orchestrate data sharing under the conditions of urgency such as in times of crises?
Impact analysis	● How can the impact of data collaboratives be measured? ● What kind of value is or can be created through data collaboratives for various actors? ● How can data collaboratives help governments meet their mission and create public value?
Incentives	● What incentives motivate various actors to engage in a data collaborative? ● What new incentive structures may be needed to advance the uptake of data collaboratives? ● How is it best to incentivize the formation of data collaboratives as opposed to sustaining them in the long run? ● What incentives are likely to maximize the value of data collaboratives?
Capabilities	● What capabilities are required to provide timely access to corporate data? ● What data analytics capabilities are required to leverage available data? ● What capabilities are required at different stages of a data collaborative from formation to evaluation? ● What capabilities are required in various contexts (by policy domain, by partner, by region)? ● What is the role of trust building capability in facilitating data collaboratives and creating value from them? ● How does the data collaborative model create new capability models? ● What capabilities are required for one-off data collaboratives as opposed to long-term ones? ● How is it possible to create clarity around roles and responsibilities of actors involved in a data collaborative?
Governance	● What role does governance play in building and sustaining data collaboratives? ● What data governance models are optimal for data collaboratives in different contexts? ● What are the dynamics between supply and demand side parties when it comes to data governance? ● What elements of responsible data governance must be put in place vis a vis data collaboratives? ● What rules, norms, standards must be proposed? ● How can data governance impact timely data access in situations of urgency?
Data management	● What is the role of data management and related capabilities and best practices in enabling data collaboratives? ● What approaches can be used to transform unstructured data into structured and make them available within a data collaborative? ● How can data ownership be defined within a data collaborative?
Interoperability	● How can data be integrated and meet the requirements of each party? ● What approaches exist for supporting data interoperability? ● How can data be anonymized and the results be accomplished? ● What is the quality of data?

5. CONCLUSIONS AND FUTURE STEPS

In this paper, we proposed a research roadmap with nine topics and multiple research questions which can help advance our understanding of data collaboratives as a novel research direction. The roadmap was formulated based on the discussions during the panel at the International Conference on Digital Government Research (dg.o 2017). As an example, we described in more detail one of the real-life cases discussed during the panel - a data collaborative on suicides among young people in India initiated by the UNICEF and The Gov Lab. Many of the challenges experienced by the partners of this data collaborative are reflected in our research roadmap, such as matching data to problems, ensuring continuous engagement, translating data insights into impact, creating a dialogue among diverse stakeholders and others. This shows the value of deriving a research agenda from practices in the field and from discussions between practitioners and researchers.

The research roadmap shows that data collaboratives exhibit a case of extreme complexity, whereas the novelty of the phenomenon is poorly understood. We identified a number of technical, as well social challenges. Namely, there are different roles, levels of expertise, and capabilities which are required from the side of data providers and data recipients. It is important to establish rules, principles, and standards and put in place the elements of a responsible data governance. Moreover, data collaboratives should be looked upon in the context of larger data management regimes and practices of the participating organizations. The suggestion of the panel was to shift the focus from technology-centric to data-centric organizations, which would require thinking more about re-orchestrating the processes, policy instruments, and capabilities of organizations and building trust among partners around data sharing.

As a next step, we call on digital government community to further explore the questions raised in this paper. The digital government community's deep disciplinary expertise and extensive experience as interdisciplinary, applied scientists poises them to make substantial contributions to the emerging conceptualization and operation of data collaboratives as socio-technical phenomena. Digital government research can potentially offer useful theoretical frameworks and analytical tools to investigate some of the research questions surrounding data collaboratives. How each study is carried out and what each examination contributes to the understanding of data collaboratives will vary, but each has the potential to create new knowledge about this important and interesting evolution of the practice of data collaboratives, in particular when viewed through the eyes of the public sector partners.

ACKNOWLEDGMENT

REFERENCES

Andrews, M., Pritchett, L., & Woolcock, M. (2015). The Challenge of Building (Real) State Capability. Harvard Center for International Development Working Paper.

Bharosa, N., Janssen, M., van Wijk, R., de Winne, N., van der Voort, H., Hulstijn, J., & Tan, Y. H. (2013). Tapping into existing information flows: The transformation to compliance by design in business-to-government information exchange. *Government Information Quarterly*, *30*, 9–18. doi:10.1016/j.giq.2012.08.006

Cameron, K. S. (1986). Effectiveness as Paradox: Consensus and Conflict in Conceptions of Organizational Effectiveness. *Management Science*, *32*(5), 539–553. doi:10.1287/mnsc.32.5.539

Gil-Garcia, J. R., & Pardo, T. A. (2005). E-government success factors: Mapping practical tools to theoretical foundations. *Government Information Quarterly*, *22*(2), 187–216. doi:10.1016/j.giq.2005.02.001

Gil-Garcia, R. J., Chun, S. A., & Janssen, M. (2009). Government information sharing and integration: Combining the social and the technical. *Information Polity*, *14*(1, 2), 1–10. doi:10.3233/IP-2009-0176

Goldkuhl, G. (2009). The challenges of interoperability in e-government: Towards a conceptual refinement. *Statistics*, *10*(10).

Janssen, M., & Helbig, N. (2016). (in press). Innovating and changing the policy-cycle: Policy-makers be prepared! *Government Information Quarterly*. doi:10.1016/j.giq.2015.11.009

Kirkpatrick, R. (2013). Big Data for Development. Retrieved from http://online.liebertpub.com/doi/pdfplus/10.1089/big.2012.1502

Landsbergen, D. J. Jr, & Wolken, G. Jr. (2001). Realizing the Promise: Government Information Systems and the Fourth Generation of Information Technology. *Public Administration Review*, *61*(2), 206–220. doi:10.1111/0033-3352.00023

Oliver, C. (1992). The Antecedents of Deinstitutionalization. *Organization Studies*, *13*(4), 563–588. doi:10.1177/017084069201300403

Scholl, H. J., & Klischewski, R. (2007). E-Government Integration and Interoperability: Framing the Research Agenda. *International Journal of Public Administration*, *30*(8-9), 889–920. doi:10.1080/01900690701402668

Seo, M. G., & Creed, W. D. (2002). Institutional contradictions, praxis, and institutional change: A dialectical perspective. *Academy of Management Review*, *27*(2), 222–247. doi:10.5465/amr.2002.6588004

Stoker, G. (1998). Governance as theory: Five propositions. *International Social Science Journal*, *50*(155), 17–28. doi:10.1111/1468-2451.00106

Susha, I., Janssen, M., & Verhulst, S. (2017). Data collaboratives as a new frontier of cross-sector partnerships in the age of open data: Taxonomy development. In *Proceedings of Hawaii International Conference on System Sciences* (pp. 2691-2700). doi:10.24251/HICSS.2017.325

The Economist. (2017). The world's most valuable resource is no longer oil, but data. Retrieved from https://www.economist.com/news/leaders/21721656-data-economy-demands-new-approach-antitrust-rules-worlds-most-valuable-resource

Tung-Mou, Y., & Maxwell, T. A. (2011). Information-sharing in public organizations: A literature review of interpersonal, intra-organizational and inter-organizational success factors. *Government Information Quarterly*, *28*(2), 164–175. doi:10.1016/j.giq.2010.06.008

Verhulst, S., & Sangokoya, D. (2015). Data collaboratives: Exchanging data to improve people's lives. Retrieved from https://medium.com/@sverhulst/data-collaboratives-exchanging-data-to-improve-people-s-lives-d0fcfc1bdd9a

Verhulst, S., & Young, A. (2017). Five ways data collaboratives create public value. Retrieved from http://datacollaboratives.org/introduction.html#section5/5c

Iryna Susha is a postdoctoral researcher in the Department of Informatics at Örebro University, a guest researcher at Delft University of Technology, and a visiting scholar at The Governance Lab of New York University. Iryna's current research focus is on data collaboratives and how they can be efficiently used to address societal challenges. Her earlier works on open government, open data, and citizen participation have appeared in various international journals and conferences.

Theresa A. Pardo, Ph.D. is director of the Center for Technology in Government and a full research professor of public administration and policy at the University at Albany (UAlbany), SUNY. Under her leadership, the Center works in the U.S. and around the world with governments and private sector and non-profit organizations to carry out applied research and strategic consulting projects focused on sustainable innovation and value creation in the public sector. Theresa serves as OpenNY Adviser to NYS's Governor Andrew Cuomo, as Chair of the U.S. EPA's National Advisory Committee and as a member of the U.S. National Science Foundation's Business and Operations Advisory Committee. She is a member of the User Working Group of the NASA Socioeconomic Data and Applications Center and the Steering Committee of the NorthEast Big Data Hub, and a past president of the Digital Government Society. Theresa is the founder of the Global Smart Cities Smart Government Research Practice Consortium and serves as a member of the City of Schenectady's Smart Cities Advisory Commission. In 2016, she served as the first female chair of the Oman's Sultan Qaboos E-Government Award Jury. Theresa's work has been recognized with UAlbany's Distinguished Alumni and Excellence in Teaching Awards as well as the Rockefeller College Distinguished Service Award. Theresa is ranked among the top five digital government scholars in terms of citations to her published work and is a recipient of Government Technology Magazine's 2015 Top 25 Doers, Drivers, and Dreamers Award. Theresa holds a doctorate in Information Science from UAlbany, SUNY.

Marijn Janssen is a full Professor in ICT & Governance and chair of the Information and Communication Technology section of the Technology, Policy and Management Faculty of Delft University of Technology. His research interests are in the field of orchestration, shared services arrangements, and open and big data and infrastructures. He is Co-Editor-in-Chief of Government Information Quarterly (GIQ), associate editor of International Journal of Electronic Government (IJEGR), conference chair of IFIP EGOV series and is chairing mini-tracks at the DG.o, ICEGOV, HICCS and AMCIS conferences. He was ranked as one of the leading e-government researchers in surveys in 2009, 2014 and 2016, and has published over 400 refereed publications. More information: www.tbm.tudelft.nl/marijnj.

Natalia Adler is the Data, Research and Policy Planning Specialist at UNICEF HQ, where she's trying to leverage data science and expertise to tackle complex socioeconomic problems through data collaboratives with the private sector and academia. She has also conceptualized a "Cities for Children" initiative, looking at the intersection of urbanization, climate change and child rights in Latin America. Natalia has previously worked with UNICEF Nicaragua and Mozambique advancing human centered design, lean start-up techniques, entrepreneurial ecosystems, and public finance management for child rights.

Stefaan G. Verhulst is Co-Founder and Chief Research and Development Officer of the Governance Laboratory @ NYU (GovLab) where he is responsible for building a research foundation on how to transform governance using advances in science and technology. Before joining NYU full time, Verhulst spent more than a decade as Chief of Research for the Markle Foundation, where he continues to serve as Senior Advisor. At Markle, an operational foundation based in New York, he was responsible for overseeing strategic research on all the priority areas of the Foundation. He is also an Adjunct Professor in the Department of Culture and Communications at New York University, Senior Research Fellow for the Center for Media and Communications Studies at Central European University in Budapest; and an Affiliated Senior Research Fellow at the Center for Global Communications Studies at the University of Pennsylvania's Annenberg School for Communications. Previously at Oxford University he co-founded and was the Head of the Programme in Comparative Media Law and Policy at the Centre for Socio Legal Studies, and also served as Senior Research Fellow of Wolfson College. He is still an emeritus fellow at Oxford. He also taught several years at the London School of Economics. Verhulst was the UNESCO Chairholder in Communications Law and Policy for the UK, a former lecturer on Communications Law and Policy issues in Belgium, and Founder and Co-Director of the International Media and Info-Comms Policy and Law Studies at the University of Glasgow School of Law.

Todd Harbour is the Chief Data Officer (CDO) for New York State, where he orchestrates work to design and implement a data management regime for the Empire State. Prior to this role, Todd was a senior federal government official based in the Washington DC metropolitan area, where he led work to establish data strategies, business frameworks, and data management platforms, which helped provide a reliable basis for answering questions from Congress and organizational leaders. Previously, Todd served as senior vice president at FGM Inc., a software and systems engineering corporation in Northern Virginia.