A Collaborative Governance Approach to Partnerships Addressing Public Problems with Private Data

Iryna Susha
Department of Informatics, Örebro University &
Delft University of Technology
iryna.susha@oru.se

J. Ramon Gil-Garcia
University at Albany, State University of New York
& Universidad de las Americas Puebla
jgil-garcia@ctg.albany.edu

Abstract

The recent explosion of data, which is generated, collected, and exchanged, opens up new opportunities and poses new challenges. Actors in different sectors have recently began to explore how they can work together and leverage these data to help address 'wicked' problems. A novel form of cross sector partnership emerges, labelled "data collaborative", which is normally focused on accessing private sector data and using it to address complex public problems. While there is emerging knowledge about how data can be shared in such partnerships, less is known about the collaboration dynamics of these partnerships. In this paper, we examine this problem from the perspective of collaborative governance and propose a framework for understanding collaboration around data sharing for public good.

1. Introduction

In the last few decades, it has become clear that in order to solve complex social problems a multiplicity of actors need to collaborate and work together towards a common goal [1, 2, 3]. The need for collaboration is related to the fact that important problems involve diverse aspects of government and society and require creative solutions. In fact, horizontal relationships in which there is no clear hierarchy and chain of command are more and more common in several policy domains from environmental protection to public health and from homeland security to social welfare policies. However, collaborating across organizational boundaries is not always easy. For instance, lack of clarity in terms of roles and responsibilities and turf issues (non-cooperation between organizations with seemingly common interests) have been identified as important challenges [4, 5].

More recently, there is an acknowledgement about the importance of data in society. Open data, big data, and data analytics are just a few prominent examples of terms used to highlight the emergent role of data for government, citizens, businesses and other stakeholders [6, 7]. The overall expectation is that the availability of more data in terms of quantity and diversity will lead to more evidence-based and thus improved decision making. For instance, open data and big data are expected to produce important economic benefits as well as more effective policies and stronger accountability [8].

Studying collaboration across organizational boundaries and information sharing to face common problems is not new in the literature nor in practical government initiatives [9, 10]. There has been research about benefits and challenges of inter-organizational collaboration and information sharing in government settings [11, 12]. However, there has been limited research which examined the intersection between inter-organizational collaboration, the affordances created by the data revolution, and the increasingly complex and urgent societal issues in need of joined action [13].

Data collaboratives could be conceptualized as new organizational forms in which government agencies, non-profit organizations and private firms share specific datasets, including private ones, with the purpose to address an important societal problem and thereby create public value [14, 15]. An essential element of data collaboratives is the societal focus underpinned by the 'data for good' philosophy, as well as the coming together of participants from different sectors who collaborate and share resources to address a complex societal issue. Much of the data which is critical for addressing societal challenges of today rests in private hands [16]. Data collaboratives leverage this new resource for public problem solving by means of novel partnerships and data sharing arrangements.

One of the main differences between information sharing initiatives and data collaboratives is that, when dealing with government information sharing, most of the time all the data is public or at least it should be available to the public. This is not always the case and

there are cases in which government agencies treat data as their own. However, the nature of the data does not change and they are still public. In contrast, data collaboratives normally attempt to use private data to address important social problems. Government agencies, non-profit organizations, and private firms share public and private data for specific efforts. They do not integrate all their data in a permanent system, but take advantage of the availability of diverse and complementary public and private data to better understand a specific problem and propose a solution.

Collaboration mechanisms are important for any organizational effort, but they are particularly essential for partnerships in which there are multiple and diverse organizations such as data collaboratives. Such partnerships frequently do not depend on hierarchical relationships and, therefore, decisions are not made at the top, because there is no formal organizational apex [17]. Instead, they rely on mechanisms that take advantage of their collaborative nature and also attempt to deal with some of the challenges of quasi-horizontal relationships [18, 19]. However, there is no clarity about the nature of these mechanisms and their particular challenges in the data age. This paper attempts to answer the following question: Are there any distinct collaboration mechanisms and challenges that come into play for data collaboratives?

This paper is organized in six sections, including the foregoing introduction. Section two briefly explains current literature about inter-organizational collaboration and information sharing and how this previous research can help to understand data collaboratives. Section three presents the collaborative governance framework, which is an example of current conceptual efforts attempting to understand the role of governance in collaborations. Section four applies some of the main ideas of the collaborative governance framework to data collaboratives using the case of Data for Climate Action (D4CA) as an illustrative example and highlights some of the differences and similarities with respect to other collaborations. Section five discusses our main findings and outlines some implications for research and practice. Finally, section six presents some conclusions and suggests areas for future research about this topic.

2. Inter-organizational collaboration in the age of data

There are many ways to define information sharing and to understand the relationships between information sharing, data integration, and collaboration. Information integration actually is a composite of four concepts [20]: collaboration networks, information sharing, data integration, and

interoperability. So, information sharing, data integration, and collaboration are different, but closely related concepts and they are affected by some of the same variables.

It is now clear that the success of initiatives involving collaboration, information sharing, and data integration is a multidimensional phenomenon, which is affected by variables of very different nature [21, 12]. Some authors cluster these factors into a few categories including contextual conditions, institutional arrangements, organizational structures and processes, technology features, and data characteristics [9, 11].

Contextual conditions refer to economic, political and social factors that affect inter-organizational collaboration and information sharing initiatives. Institutional arrangements are laws, regulations and other formal and informal rules, which are particularly important in the public sector. Organizational structures and processes include variables that represent how organizations and organizational networks function and also how they are constituted in terms of configuration, size, and specialization, among other aspects. Technology features refer to complexity, newness, and compatibility, among others. Finally, data characteristics include variables related to the availability, diversity and quality of data. It could also include data standards, metadata, and shared assessment of the necessary data.

Governance has been identified as a very important factor affecting inter-organizational collaboration and information sharing [22, 23]. Governance in general has several definitions but in essence it can be understood as a set of coordinating and monitoring activities which ensure survival of a partnership [38]. In the context of inter-organizational collaboration, governance has been defined as "the decision-making structures that form within and across the formal and informal networks of organizations that are created to collaboratively formulate and implement cross-boundary information sharing initiatives" [22].

Governance structures are important because they help to clarify roles and responsibilities, generate trust and improve coordination. They are also important for establishing standards, selecting shared incentives, and developing sound strategies. Overall, governance structures and mechanisms have the potential to significantly affect the success of collaboration and information sharing initiatives [23].

Since data collaboratives share some of the characteristics of information sharing initiatives, but also have some particularities, it is not clear to what extent collaboration structures and mechanisms are similar and function in similar ways. For instance, the fact that in data collaboratives frequently the data can be used for a specific purpose only and all other uses

are restricted, might require collaboration mechanisms that are more flexible and dynamic and adapt to specific projects, even when the partners are the same. In addition, the complexities associated with the sharing of private (big) data within a data collaborative add a new dimension to the way inter-organizational collaboration is normally orchestrated. By building on the literature reviewed above, we explore whether and how data collaboratives demand new or different mechanisms for parties to collaborate given the data context. The following section explains the main elements of the collaborative governance framework we are proposing as a starting point for our analysis.

3. Collaborative governance framework

To answer our research question, we use the collaborative governance framework proposed by Emerson, Nabatchi, & Balogh [24]. This framework synthesizes and integrates previous empirical and theoretical developments about this topic and, therefore, it is a solid foundation for our study. In their interpretation, collaborative governance stands for "the processes and structures of public policy decision making and management that engage people constructively across the boundaries of public agencies, levels of government, and/or the public, private and civic spheres in order to carry out a public purpose that could not otherwise be accomplished" [24, p.2]. We selected this particular framework because of its comprehensiveness and potential for generalizability: (1) it integrates key elements of other relevant frameworks and (2) it has a broader focus and can be applied to different sectors, policy domains, process mechanisms, and geographical and temporal scales. The framework is depicted in Figure 1.

Figure 1
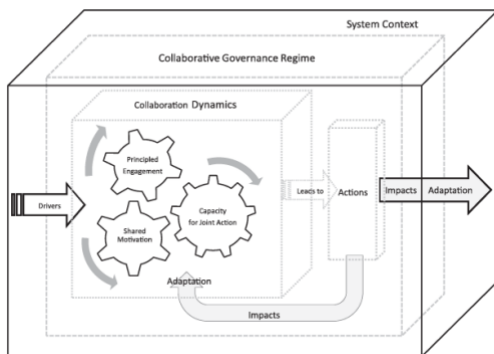The Integrative Framework for Collaborative Governance

Figure 1. Collaborative governance framework [24]

The framework incorporates three nested dimensions (boxes) and their respective components.

The outermost box represents a wider *system context* of political, legal, socio-economic, environmental, and other influences which affect and are affected by collaboration. From the system context emerge drivers of collaboration. The middle box depicts the *collaborative governance regime (CGR)*, where "regime" stands for a particular mode of public decision making in which cross boundary collaboration prevails [24]. The innermost box shows *collaboration dynamics* consisting of three components: principled engagement, shared motivation, and capacity for joined action. These components work together in an interactive and iterative way to produce *collaborative actions*. The latter can lead to impacts (results on the ground) and to the adaptation of the CGR and the system context. Each of these components are further characterized by multiple variables, but due to space limitation, we will discuss them in section five, in which we analyze and systematically apply the framework to an illustrative case. Here we offer this overview for the reader to understand the main components.

The purpose of this framework is to serve as a "conceptual map" [24] for practitioners and guide them through issues which are important for an effective collaboration. To put it bluntly, the framework addresses the questions of 'why', 'how', and 'so what' of collaboration. With our work, we respond to the call of the authors of the framework to test and critically apply the framework to different examples of collaborative governance. We argue that data collaboratives share some of the characteristics of other collaboration efforts, but also have particularities that makes them interesting for the application and potential adaptation of this framework.

4. Collaborative governance in data collaboratives

We structure our analysis highlighting how the collaborative governance framework [24] informs our understanding of data collaboratives following the logic of their model and discussing how data collaboratives emerge, what mechanisms put their process dynamics in motion, and how outcomes and impacts are generated. In our analysis, we refer to the Data for Climate Action case as an illustrative example to showcase how the framework sheds light on a particular instance of data collaboratives.

4.1. Data for Climate Action (D4CA)

Data for Climate Action was an initiative of the UN Global Pulse which ran from March to November

2017. This project brought together private sector companies with data scientists from academia and non-profits and was facilitated and intermediated by an intergovernmental organization, UN Global Pulse. This collaboration culminated in an innovation challenge which aimed to channel data science and big data from the private sector to fight climate change [25]. For this purpose, companies voluntarily shared relevant datasets with teams of researchers who were granted access to these data based on their research proposals. Nine companies provided data to this challenge which included: retail transaction data, social posts from different sources, lightning and weather data, sales of select products, call detail records, satellite imagery, air pollution data, weather observations API, and local road conditions. The initiative was made possible thanks to the support from Western Digital (an American corporation providing data storage solutions) and Skoll Global Threats Fund (an American non-profit supporting social entrepreneurship). We selected this case as an illustrative example because (1) it is set in an international context and (2) it is a finished project which showed results. This makes it possible to discuss all elements of the framework from the inception to the impact phase. We do so on the basis of desk research and review of literature as our work is conceptual at this point.

4.2 Inception of data collaboratives

The CGR framework puts forward a proposition that for collaboration to unfold one or more of the following drivers must be present: leadership, consequential incentives (e.g. a problem, crisis, threat, or opportunity), interdependence (i.e. inability of organizations to accomplish something on their own), and uncertainty (such as about how to manage a societal challenge). All four drivers are relevant for data collaboratives. Data collaboratives in their essence are driven by the recognition of a pressing societal issue (e.g. climate change, disaster aftermath, poverty, public health threats) which is often a 'wicked' problem requiring joined action from diverse stakeholders. On the other hand, a powerful driver is the opportunity offered by the data revolution to use data as a new resource for comprehending and addressing 'wicked' problems. Besides drivers, the framework posits collaboration is affected by a number of contextual variables, which can give momentum to or constrain collaboration. These include policy frameworks, political/power relations, socioeconomic/cultural issues, network connectedness, levels of trust/conflict, history of addressing the problem in question.

The D4CA campaign was publicly announced during the UN Climate Change conference COP21 in Paris in December 2015. It was conceived and implemented under the leadership of the UN Global Pulse who acted as an innovation intermediary and facilitated the contribution of resources by different players for the benefit of climate science. This collaboration aimed to trigger joined action to tackle a truly global challenge. Already in 2015 three companies gave their commitment to collaborate and provide datasets to advance the goals of the campaign. The second phase, the innovation challenge, was planned to take place and be completed by the end of 2016. However, it was not until March 2017 that the challenge could be launched.

This shows that in practice the four drivers from the CGR framework – consequential incentives, interdependence, uncertainty, and leadership – are necessary but not sufficient for data collaboratives to come to life. For data collaboratives to realize, another crucial driver is availability of *resources*, first and foremost financial. Although pro bono in many respects, data collaboratives bear significant costs and finding required data expertise is another concern, which may require additional resources. The CGR framework presents resources as an element of the system context, but in the case of data collaboratives the influence of funding on whether a data collaborative will proceed or not is very strong. This is similar to inter-organizational information sharing, in which financial resources are essential due their non-hierarchical and collaborative nature [11, 12, 39].

Another impactful driver of data collaboratives is the presence of *incentives* for data providers to donate commercial data for a societal purpose. In general, open data initiatives face important challenges and require a variety of economic, political, institutional and operational incentives [6, 7, 8]. In addition, in industry-academia partnerships firms are motivated to engage with universities for the following reasons [26]: (1) support for product development; (2) access to public research funding; (3) solutions to technological problems; and (4) finding research opportunities. In the context of data collaboratives particularly, incentives for companies to share data with data scientists for the benefit of society include 5Rs: reciprocity; research, recruitment, and insights (benefitting from free data science expertise); reputation and public relations; increasing revenue (when corporate data is offered at a cost) [27]. In the case of the D4CA, the factors of positive image, external data science expertise, publicity, opportunity to innovate, corporate responsibility, and access to research insights played a role in motivating companies to join the initiative.

4.3. Process dynamics of data collaboratives

The CGR framework also conceptualizes the inner workings of collaboration, or collaboration dynamics. Collaboration dynamics are put in motion by three interacting elements: principled engagement, motivation, and capacity for joined action. We will tackle them subsequently.

The first element is principled engagement which is enacted through participants engaging in (1) discovery of interests, concerns, and values; (2) defining shared purpose, terminology, and tasks; (3) deliberating; and (4) producing determinations, such as an agreement or course of action. Through these steps, a shared theory of action is supposed to be shaped. This element plays a particularly important role when it comes to data collaboratives. One key task in a data collaborative is *scoping and defining the problem* and breaking it down into questions which can be answered with available data [28, 31]. This is often an iterative process of going back and forth to find the perfect match between available data and the problem formulation at hand. Depending on the type of data collaborative, this process can be more problem- or data-driven. This also has an influence over who to collaborate with. The D4CA case can be seen as an example of the latter, as it is a challenge competition in which certain datasets are offered to the participants and the problem to be addressed is formulated rather loosely. Generally, articulating the actual problem and the outcomes to be achieved, instead of getting carried away with the data, is seen as a more successful strategy for data collaboratives [28].

As the participants define the terms of engagement, another crucial aspect in the context of data collaboratives is assessing *risks* and developing a risk mitigation strategy. The CGR framework does not give sufficient attention to this. As a minimum, this concerns data privacy in instances when personal data is shared. More broadly, this implies developing a data responsibility framework including principles, processes, and tools to leverage data for the benefit of affected populations [29].

The second element of the CGR framework describing the process dynamics of collaboration is shared motivation which can be understood as the social capital of collaboration. Shared motivation emerges from mutual trust, understanding, internal legitimacy, and commitment of participants. Many collaborations, and data collaboratives are not an exception, are built around interpersonal networks of *trusted partners* and from the history of previous collaborations. Going beyond that can be difficult, especially in case of data collaboratives, because convincing companies to trust their commercial data to an outsider can be challenging. Reputation of the data recipient is an important factor in this respect. Trust has a positive effect on the attainment of tangible benefits from collaborations [30]. Overall, trust is likely to be built in the course of repeated engagements.

The third element which is activated by the collaboration dynamics is capacity for joint action. Emerson et al. conceptualize this element as a combination of procedural/institutional arrangements, leadership, knowledge (transfer and management), and resources. Procedural and institutional arrangements mean process protocols and organizational structures which are necessary to manage the collaboration. They can be more formal (e.g. decision-making rules, regulations, operating procedures etc.) or informal (e.g. norms of reciprocity). In case of data collaboratives, a combination of both is typical as the participants often sign a memorandum of understanding and draw up terms and conditions governing data sharing and use. A range of issues are to be defined, such as data ownership, liability, public release and transparency, cost, intellectual property provisions [31]. Next to this, a particularly vital step in managing collaboratives is making available necessary resources and ensuring effective knowledge management. To this end, deploying *data stewards* is seen as one of the crucial steps towards making data collaboratives more scalable and de-risked [32].

4.4. Impact of data collaboratives

Any form of collaboration aims to achieve certain goals. The CGR framework conceptualizes the impact of collaboration as intentional and unintentional changes of state in the system context which can be physical, environmental, social, economic, and/or political. This also includes the added value of a social good or innovation developed as a result of collaboration actions. Impacts can be specific or broad, short term or long term. Ideally, their extent and nature are consistent with the desired results and the goal of collaboration. Furthermore, Emerson et al. put forward a proposition that the impacts of collaboration are more likely to match the desired ones and have fewer unintended negative consequences if they are specified and derived from a shared theory of action. The latter is understood as "the group's understanding of the size of the problem or challenge it is addressing, as well as the scope and scale of the group's chosen activities and interventions" [24, p.11]. Besides, according to the CGR framework, the impacts of collaboration trigger the adaptation of the system context and can potentially bring about transformative change to the problem in question. Next to the adaptation of the

situation being addressed, the impacts of collaboration also signal the adaptation of the collaborative governance regime, or of the institutional arrangements of the collaboration itself. It is a feedback mechanism to reflect on the effectiveness of the collaboration and on the 'return on investment' of the participants. For instance, if the impacts of collaboration are not in line with the targeted ones, the participants will be pressured to make adjustments to the way they collaborate or even to exit the partnership. The adaptation element of the framework is seen as critical, since Emerson et al. make a proposition that the sustainability of collaboration over time depends on how well the participants can adapt to the nature and level of impacts of their actions.

Data collaboratives as a form of collaboration are not limited to any particular domain or sector. Thus, the kind of impacts they (aim to) generate depend on the specific configuration of the collaborative (for a more detailed description thereof, see [15]). Essentially, all data collaboratives are driven by the need to fill in a data gap in relevance to a certain societal problem. Thus, one primary goal of a data collaborative is to close such a gap by making relevant data (or data services) available. The data is then analysed to produce problem-specific data insights, which is the immediate impact of a data collaborative. Thus, some data collaborative models result in generating new knowledge, while others – in developing a data product/service which is then offered as a tool to generate new insights.

Examples of the latter include Open Algorithms Project (OPAL)[1], Humanitarian Data Exchange (HDX)[2], and 23andMe Patient Centric Research Portal[3]. Other data collaboratives do not involve the development of any data infrastructure but rather produce one-off data insights. Any systematic assessment of impacts of data collaboratives as a novel form of partnership has to date been limited. That said, mostly, the impacts of data collaboratives tend to be quite specific and rather short than long term due to the experimental nature of this novel form of partnership. Some collaboratives can be considered more successful and impactful than others. One major difficulty is assessing to what extent the generated insights or data products managed to trigger transformative change to the societal problem. This often lies beyond the scope of the collaborative and involves actions from policy stakeholders not directly involved in the collaborative.

The D4CA project for example attracted more than 450 teams of researchers from 67 countries who applied to participate in the challenge; out of these, 97 were selected to access the data and develop innovation projects [33]. They were grouped in three thematic areas: climate mitigation, climate adaptation, and climate and Sustainable Development Goals. The output of these projects was either an application/dashboard/visualization or a research paper/poster. The winners of the challenge were announced on 12 November 2017 at the UN climate change conference COP23 in Bonn, Germany.

The grand prize of 16,000 USD was awarded to the team of researchers from Mexico's National Institute for Ecology and Climate Change and University of California, Berkeley. This team focused on the problem of air pollution in Mexico City and used the traffic data from Waze (in combination with other sources) to estimate emissions from the transportation sector and the potential impact of a number of electrification policies. Three more projects received thematic awards: (1) a framework to predict and alleviate road flooding in Senegal developed by the Georgia Institute of Technology; (2) a platform to predict and monitor the impact of forest fires in South Africa developed by the University of Cape Town and the University of Buffalo; and (3) an analysis of changes in spending patterns related to changes in air quality in Spain conducted by a team from Yale University. However, in these cases it is premature to speculate about policy impact beyond the timeframe of the D4CA project.

There is more clarity regarding the impact of the grand prize winner team's work. This team brought together both data scientists and public policy officials, which created a direct link to real life implementation of their data insights. After the D4CA, the team's work gained traction in the policy arena when several government organizations showed interest in their findings [34]. The team's visualization also fed into the discussions around the drafting of Mexico's national electro-mobility agenda. This example shows seeds of potential transformative change and adaptation of the system, as the CGR framework describes it. However, like most other innovation challenges, the D4CA presumed no explicit follow up of the winning projects. It is only at the discretion of the organizers that the contact between the research teams and field practitioners would be facilitated.

This problem of the one-off nature of challenges is highlighted in the literature [35] and is common across different types of innovation competitions. Typically, the cut-off point is the awards, whereas the actual presentation and advocacy of the data science or innovations to relevant policy makers is outside of the scope of such challenges. Moreover, the expectations of the organizers of innovation challenges are typically

---

[1] https://www.opalproject.org
[2] https://data.humdata.org
[3] https://www.23andme.com/en-eu/

expressed in a desired number of submissions. This is often used as a proxy to estimate how successful the competition was.

Besides the lack of explicit feedback loops to the system context (policy arena) as discussed above, it is often unclear how the adaptation of the collaborative governance regime itself can take place, including as in the case of D4CA. The D4CA challenge was preceded by the Big Data Climate Challenge organized by the UN Global Pulse in 2014, hence there was and is potential for learning and adaptation for the next initiative. One lesson to learn is the success of the D4CA grand prize winners of bringing together academics and government into the same team. For the following competition, this could become a requirement or preference as this creates better exposure of the data insights to the policy makers. Besides this, data collaboratives will benefit from the creation of a more formal community of practice. In this respect, the UN Global Pulse is uniquely positioned to play a leading role given their mandate to advance data-driven development and their continuous involvement in data innovation collaborations on the ground.

5. Discussion and Implications

The main purpose of our study was to understand whether there are any distinct collaboration mechanisms and challenges associated with data collaboratives. To answer this, we showcased how the CGR framework applies to data collaboratives as a novel form of partnerships. In our analysis, we discussed a number of issues specific to data collaboratives compared to other information sharing initiatives. These include the following and we briefly explain each of them in this section.

- Incentivizing private sector data providers is an important prerequisite/driver of data collaboratives.
- Defining the problem in relation to the available data is a key determinant of the theory of action in a data collaborative.
- Identifying and mitigating risks of sharing and using private data is a pivotal point in the design of data collaboratives.
- Data collaboratives require a distinct set of data stewardship capabilities to enact and manage data sharing and use.
- Regarding all issues above, trust between partners plays a decisive role and its role is different given the multi-sector nature of the partnership.

Collaboration efforts that involved mainly public sector organizations, such as information sharing initiatives among government organizations, do not always have to face challenges related to creating incentives for private companies to share their data. Within the public sector, incentives are related to the idea of the "common good", but also to legal requirements and political willingness [21, 11]. Data collaboratives face these same challenges, but also the fact that private sector data providers need to understand and assess the benefits from their participation, even if it is not an immediate financial gain. As can be seen from the D4CA case, without commitment and motivation of the data providers the data collaborative would not have proceeded.

Since data are the primary source of collaboration and potential results in data collaboratives, many times the problem needs to be defined in terms of the available data. This is not always easy, since the actors involved in the collaboration not always know what is available from the other actors and what data they are willing to share. This is also challenging because literature about information sharing argues that the problem should be defined first and then the partners should start making decisions about which information needs to be shared and which organization has that information [9, 21, 13]. Sometimes if the data do not exist they can be collected as part of the collaborative effort. In the case of data collaboratives, most of the time the data already exists and they are part of the normal operation of the organizations involved, including private partners. It is not clear if private sector organizations would be willing to collect new data and use their own resources to do that.

As mentioned early, data collaboratives could share specific data for very specific purposes and any other use is not allowed. This is particularly important for private sector organizations, since they are participating in the collaboration assuming that their data will be used only for the purpose they have already agreed upon [27]. Therefore, as part of the process, it is extremely important to identify and mitigate risks related to sharing and using private data. The governance structure and mechanisms should reflect how the data is going to be managed throughout the whole data life cycle for a specific initiative, including potential storage, reuse, and disposal. Besides risks, data collaboratives should also consider and take precautions against potential data harms [29]. Devising a comprehensive data responsibility framework can be helpful in this respect.

Within the public sector, data stewardship is very important since data is not owned by government agencies or public managers. They are just stewards of those data on behalf of the citizens and other

stakeholders [36]. To a certain extent, this is different for private companies since the data that they collect are generally accepted as their property. Recently, there are arguments against this, particularly when referring to data about individuals (personal data). However, in general, it is important to acknowledge the similarities and differences in stewardship capabilities and conceptualizations between private, public, and nonprofit organizations, since this could affect their willingness to participate in data collaboratives.

Again, the multisector nature of the partnership and its focus on specific datasets present some specific challenges in terms of trust among participants. Organizations from different sectors have different interests and concerns and they need to trust each other in order to share data and form a successful collaboration [37]. In addition, the fact that the collaboration needs to materialize in the context of sharing of specific datasets that organizations value, could hinder trust development processes. This is because it is not a high-level notion of trust, but partners need to trust each other in terms of specific uses of the data that have been allowed for the initiatives or not. Having a powerful shared goal could help to develop trust among participants, but it is not always enough. So, even for initiatives that involved a very good cause, it could still be challenging to develop the necessary trust. Innovation intermediaries with a good reputation, as in the case of D4CA, can greatly contribute to building trust between data providers and data users.

6. Conclusions

It seems clear that data collaboratives have some particularities that are essential to understand their collaboration dynamics and potential outcomes. More specifically, their multisector nature and their focus on specific datasets seem to affect how the collaboration is formed and how it evolves and produces results. We found that using the collaborative governance framework helps to understand the specific mechanisms and how the different elements work in multisector partnerships around the use of private data for public good. Our analysis also reveals that the framework does not sufficiently explain some specific complexities associated with data collaboratives and thereby should be extended in certain ways. As a way to conclude this paper, we elaborate on these two statements below. We also provide some suggestions for future research about this topic.

First, the value of the CGR framework is that it offers a conceptual map to navigate the different elements of collaboration. It represents collaboration as three nested dimensions – the way parties engage with one another to shape their collaborative actions (collaboration dynamics) is influenced by the system of rules, norms, principles (collaborative governance regime) which in turn are influenced by the contextual enablers and constraints from the political, economic, social, legal domains (system context) [24]. The framework helps to identify all these elements for the case of data collaboratives and better understand the collaboration mechanisms and dynamics.

For instance, the framework puts forward a number of propositions many of which underscore the importance of a *shared theory of action* (understanding of the size of the problem and the scope/scale of activities to be undertaken). Once such a theory is identified explicitly, the planned collaborative actions are likely to be implemented successfully (in the case of data collaboratives, data to be shared and used). Furthermore, such a theory should specify the impacts that the data sharing and use should achieve (i.e. addressing the societal problem in question); then the impacts are likely to match the desired ones. In practice, this is difficult to articulate and plan for beforehand, largely because relevant policy makers on the ground are typically not directly involved in the data collaborative as it happens.

Data collaboratives is an emergent experimental form of addressing public problems with private resources. This is a new opportunity for public managers and policy makers about which yet little is known in official policy circles. Involving relevant decision makers in early phases of data collaboratives may be challenging but extremely beneficial if one wishes to see data make its way into actual policy decisions. This can also help overcome the shortsightedness in terms of impact of data collaboratives and their one-off nature. Future research should study the effects of leadership and the involvement of specific public managers and policy makers on the overall results of data collaboratives.

Second, although we found the framework to be a useful analytical lens, we also argue that it requires some adaptation in the context of data collaboratives. First, it is an idealistic view to present drivers of collaboration as originating exclusively from the system context. Many data collaboratives, and other types of partnerships, get to a difficult start because of issues internal to the collaborative effort, such as capabilities, resources, data availability, (lack of) commitment, and turf issues, among others. We suggest adjusting this aspect of the framework to reflect both external and internal drivers. We argue that this is not unique to data collaboratives and such adaption could benefit the framework as a whole and its applicability to very diverse collaboration efforts.

We also observe that in the case of data collaboratives there is one more adaptation feedback loop – from impacts to drivers – as results of one data collaborative may trigger other data collaboratives, once more organizations see what is possible and how it can be done. Adding and explaining this additional feedback loop could also be a useful extension of the framework. Future research could explore more cases of data collaboratives and better understand how this new adaption feedback loop works in different contexts. Having more cases could also help to identify any other potential extensions of the collaborative governance framework. Our next step is validating our findings about the framework empirically.

8. Acknowledgements

9. References

[1] T. Christensen and P. Lægreid, (2007) The Whole-of-Government Approach to Public Sector Reform, Public Administration Review 67(6), 1059–1066.

[2] Gil-Garcia, J. Ramon. "Towards a smart State? Inter-agency collaboration, information integration, and beyond." *Information Polity* 17, no. 3, 4 (2012): 269-280.

[3] B. Klievink andM. Janssen, (2009) Realizing joined-up government – Dynamic capabilities and stage models for transformation, Government Information Quarterly 26, 275–284.

[4] Thompson, A.M., Perry, J.L. (2006) Collaboration processes: inside the black box. Pub. Adm. Rev. 66, 20–32.

[5] Sayogo, Djoko, J. Ramon Gil-Garcia, and Felippe Cronemberger. (2016). *Determinants of Clarity of Roles and Responsibilities in Interagency Information Integration and Sharing (IIS).* Paper presented at the 15th IFIP Electronic Government (EGOV) and 8th Electronic Participation (ePart) Conference 2016, Guimarães, Portugal, September 5-8.

[6] S. Dawes and N. Helbig, (2010) Information strategies for open government: Challenges and prospects for deriving public value from government transparency, in Proceedings 9th International Conference on E-government (EGOV), Lausanne, Switzerland, pp. 50-60.

[7] A. Zuiderwijk, M. Janssen, S. Choenni, R. Meijer, and R. Sheikh Alibaks, (2012) Socio-technical impediments of open data, Electronic Journal of eGovernment, vol. 10, no. 2, pp. 156 – 172.

[8] Janssen, M., Charalabidis, Y., & Zuiderwijk, A. (2012). Benefits, Adoption Barriers and Myths of Open Data and Open Government. Information Systems Management, 29(4), 258–268

[9] Dawes, S. S. (1996). Interagency information sharing: Expected benefits, manageable risks. Journal of Policy Analysis and Management, 15(3), 377–394.

[10] Fedorowicz, J., Gogan, J. L., & Williams, C. B. (2007). A collaborative network for first responders: Lessons from the CapWIN case. Government Information Quarterly, 24(4), 785–807.

[11] Gil-Garcia, J. Ramon and Djoko Sigit Sayogo. (2016). Government Inter-Organizational Information Sharing Initiatives: Understanding the Main Determinants of Success. Government Information Quarterly, 33 (3): 572–582.

[12] Yang, T. M.,& 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.

[13] Puron-Cid Gabriel, J. Ramon Gil-Garcia and Luis F. Luna-Reyes. (2016). Opportunities and Challenges of Policy Informatics: Tackling Complex Problems through the Combination of Open Data, Technology and Analytics. International Journal of Public Administration in the Digital Age 3 (2): 66-85.

[14] S. Verhulst and D. Sangokoya, "Data Collaboratives: Exchanging Data to Improve People's Lives," in Medium vol. 2015, ed, 2015.

[15] Susha, I., Janssen, M. and Verhulst, S. (2017), "Data collaboratives as a new frontier of cross-sector partnerships in the age of open data: taxonomy development", Proceedings of the 50th Hawaii International Conference on System Sciences in Waikoloa, HI.

[16] B. Noveck. (2015, 10 May). Data Collaboratives: Sharing Public Data in Private Hands for Social Good. Available:
http://www.forbes.com/sites/bethsimonenoveck/2 015/09/24/private-data-sharing-for-publicgood/# 28dab08b65bb

[17] Mintzberg H. (1989) The Structuring of Organizations. In: Asch D., Bowman C. (eds) Readings in Strategic Management. Palgrave, London.

[18] Powell, W.W. (2003), "Neither market nor hierarchy", The Sociology of Organizations: Classic, Contemporary, and Critical Readings, Vol. 315, pp. 104-117.

[19] Rethemeyer, R.K. and Hatmaker, D.M. (2008), "Network management reconsidered: an inquiry into management of network structures in public sector service provision", Journal of Public Administration Policy Research and Theory, Vol. 18, pp. 617-646.

[20] Gil-Garcia, J. Ramon, Theresa Pardo and Brian Burke. (2010). Conceptualizing Inter-Organizational Information Integration in Government. (Chapter XI, pp. 179-202). In Jochen Scholl (Ed). Electronic Government: Information, Technology, and Transformation. Advances in Management Information Systems (AMIS) Series. Armonk, NY: ME Sharpe.

[21] Dawes, S. S., Cresswell, A.M., & Pardo, T. A. (2009). From'need to know' to 'need to share': Tangled problems, information boundaries, and the building of public sector knowledge networks. Public Administration Review, 69(3), 392–402.

[22] Pardo, Theresa, J. Ramon Gil-Garcia and Brian Burke. (2008). Governance Structures in Cross-Boundary Information Sharing: Lessons from State and Local Criminal Justice Initiatives. Paper presented at the 41st Hawaii International Conference on System Sciences (HICSS), organized by the College of Business, University of Hawai'i at Mãnoa, Waikoloa, Big Island, Hawaii, United States, January 7-10.

[23] Sayogo, Djoko and J. Ramon Gil-Garcia. (2015). Analyzing the Influence of Governance Structure Determinants on the Success of Inter-Organizational Information Sharing Initiatives. Paper presented at the 48th Hawaii International Conference on System Sciences (HIICS), organized by the College of Bussiness, University of Hawai'i at Mãnoa, Big Island, Hawaii, USA, January 5-8.

[24] K. Emerson, T. Nabatchi, and S. Balogh, "An Integrative Framework for Collaborative Governance", Journal of public administration research and theory 22(1), 2012, pp. 1-29.

[25] Data for Climate Action. Website. Accessed 15 June 2018 from http://www.dataforclimateaction.org

[26] Bodas Freitas, I.M., Verspagen, B. The motivations, institutions and organization of university-industry collaborations in the Netherlands (2017) Journal of Evolutionary Economics, 27 (3), pp. 379-412.

[27] OECD & The Gov Lab (2017). Access to new data sources for statistics: Business models and incentives for the corporate sector", OECD Statistics Working Papers, 2017/06, OECD Publishing, Paris.

[28] UNDP & UN Global Pulse (2016). A guide to data innovation for development: From idea to proof of concept.

[29] OCHA 2016. Building data responsibility into humanitarian action. Accessed 15 June 2018 from http://datacollaboratives.org/static/files/framework.pdf

[30] Bellini, E., Piroli, G., Pennacchio, L. Collaborative know-how and trust in university–industry collaborations: empirical evidence from ICT firms (2018) Journal of Technology Transfer, pp. 1-25. Article in Press.

[31] The Gov Lab, Data Collaboratives.org, "Designing a data collaborative", Accessed 15 June 2018 from http://datacollaboratives.org/canvas.html#phase-4

[32] The Gov Lab, "Data Stewards", Accessed 15 June 2018 from https://datastewards.net

[33] D4CA press release. Accessed 15 June 2018 from https://www.businesswire.com/news/home/20171129005218/en/Global-Pulse-Western-Digital-Announce-Winners-Data

[34] Data Makes Possible, Accessed 15 June 2018 from https://www.datamakespossible.com/data-climate-action-challenge-grand-prize-winners-interview/

[35] Susha, Iryna, Åke Grönlund, and Marijn Janssen. "Driving factors of service innovation using open government data: An exploratory study of entrepreneurs in two countries." Information Polity 20, no. 1 (2015): 19-34.

[36] Dawes, S. S. (2010). Stewardship and Usefulness: Policy Principles for Information-Based Transparency. Government Information Quarterly, 27(4), 377–383. doi:10.1016/j.giq.2010.07.001

[37] Sayogo, Djoko S., J. Ramon Gil-Garcia, Felippe Cronemberger, and Bambang Widagdo. (2017). The Mediating Role of Trust for Inter-Organizational Information Sharing (IIS) Success in the Public Sector. Paper presented at the 18th Annual International Conference on Digital Government Research, organized by the Digital Government Society. Staten Island, NY, USA, June 7-9.

[38] Bryson, John M., Barbara C. Crosby, and Melissa Middleton Stone. 2006. The design and implementation of cross-sector collaborations: Propositions form the literature. Public Administration Review 66:44–55.

[39] Dawes, Sharon S., Mohammed A. Gharawi, and G. Brian Burke (2012). Transnational public sector knowledge networks: Knowledge and information sharing in a multi-dimensional context. Government Information Quarterly 29 (Supplement 1): S112-S120.