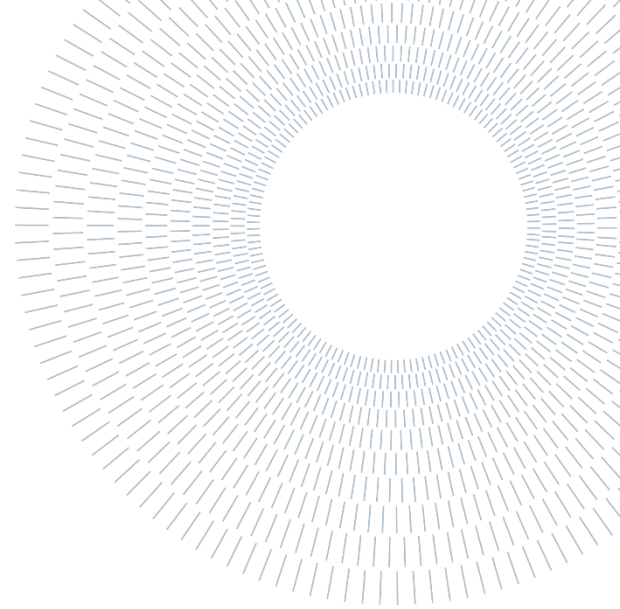




**POLITECNICO**  
MILANO 1863

SCUOLA DI INGEGNERIA INDUSTRIALE  
E DELL'INFORMAZIONE



EXECUTIVE SUMMARY OF THE THESIS

## Designing data collaboratives' governance for their long-term stability: a key factor analysis through a multiple case study

TESI MAGISTRALE IN MANAGEMENT ENGINEERING – INGEGNERIA GESTIONALE

**AUTHOR: SIMONE BARTALUCCI**

**ADVISOR: VERONICA CHIODO**

**CO-ADVISOR: FEDERICO BARTOLOMUCCI**

**ACADEMIC YEAR: 2021-2022**

---

### 1. Introduction

The complexity of social, economic and environmental challenges has made it difficult for individual central authorities to intervene. Increasingly, innovative methods involving collaboration with organizations from diverse backgrounds are required. The use of data represents an innovative approach that enables a more effective evidence-based way to address these challenges. Although the amount of data generated and collected every day through digital technologies is significant and growing, most of it is kept by private companies that consider it as a strategic source of competitive advantage. Data Collaboratives (DC) are an emergent phenomenon finalized at data sharing for social good. They are defined as “cross-sector (and public-private) collaboration initiatives aimed at data collection, sharing, or processing for the purpose of addressing a societal challenge” [1]. Differently from open data initiatives the access to data is restricted to a specific set of actors, facilitating the sharing of valuable private data. However, the socio-technical nature of this phenomenon adds

governance complexities that makes DCs struggle to evolve beyond the pilot phase or constrained to events of limited duration. This study is aimed at understanding governance components that influence the stability of these projects in order to systematize them in the long run and allowing the scaling of their impact.

### 2. Preliminary literary review

The first step of the research was a review of the existing literature on DCs and related fields in order to understand the specifics of these initiatives.

The literature reviewed provided an understanding of the specifics of the main actors involved, i.e., the motivations and critical aspects found in the private, public, and nonprofit sectors. It was then possible to understand the elements that distinguish the process of data sharing and that can determine the failure of the initiative, i.e., data quality, data competences and privacy compliance. Areas of impact and ways in which initiatives can generate social value were analyzed, also revealing a lack of impact measurement methods. The main contributions to literature are

those of two authors: S. Verhulst and I. Susha. The former conceptualized the first definition of DC and provided an earlier mapping of it [2]. He is also one of the major contributors of the open database of DC cases, accessible at [datacollaboratives.org](http://datacollaboratives.org). This database was revised by Bartolomucci et al. (2022) [3] to perform a cluster analysis, which was used to develop this research. Among the second author's publications, the paper that report the critical success factors of DCs [4] provided important inputs for this research.

The literature review revealed that this is a recent research area that still has some limitations. Gaps were found in the scarcity of empirical studies on governance of these projects [5]–[7]. In addition, the study of factors affecting long-term sustainability is mostly an unexplored field.

### 3. Research questions and methodology

To fill these gaps in the literature, research focused on establishing relationships between the elements that constitute collaborative governance and long-term stability. Specifically, two research questions arose:

1. *How governance factors influence the long-term stability of data collaboratives?*
2. *What are the governance factors distinguishing DC from other forms of cross-sector partnerships?*

To answer them, an explanatory multiple case study [8] was designed with an abductive approach. Through an in-depth review of literature, theoretical governance dimensions that influence sustainability have been deductively identified. These dimensions were used to develop semi-structured interviews with referents of DC real cases. A heterogeneous sample of nine currently active cases was selected from the database realized by Bartolomucci et al. (2022) [3], which allowed to distinguish cases with characteristics of long-term stability. The interviewees were asked to report their experiences during the interview, which lasted between sixty (60) and ninety (90) minutes. The interviews were audio-recorded and transcribed. The data gathered have been triangulated with available literature about the cases. To add further empirical validation, two cases of DCs that were no longer active but contained the same stability characteristics identified by the classification

mentioned above were studied from accessible secondary sources. The eleven cases were analyzed through the inductive coding of the text of the interviews and the documentary material. The text was coded using an open code that is mainly based on the content of the transcripts. Through an iterative process, the codes were grouped according to epistemic similarity.

Table 1 Cases analyzed

ID	DC name	Continuity	Main sector	Region	DC cluster	Data used	Data purpose	Actors involved
C1	Estima: Mobile Positioning Data for Tourism Statistics	Ongoing	Economic Development	Europe and Central Asia	3. continuous effort to improve structural responses	Observed personal data	Tertiary	Private Public
C2	California Data Collaborative	Ongoing	Infrastructure	North America	3. continuous effort to improve structural responses	Disclosed non personal data	Secondary	Private Public No profit
C3	Crivy	Ongoing	Infrastructure	Europe and Central Asia	3. continuous effort to improve structural responses	Disclosed non personal data	Primary	Private Public
C4	Salus Coop	Ongoing	Health	Europe and Central Asia	1. collaborative effort to support wide-scale research projects	Disclosed personal data Observed personal data	Secondary	No profit Civil society
C5	Impact Deal	Ongoing	Economic Development	Europe and Central Asia	5. external responses to structural problems	Disclosed non personal data	Tertiary	Private No profit
C6	Consumer Data Research Centre	Ongoing	Economic Development	Europe and Central Asia	3. continuous effort to improve structural responses	Disclosed non personal data Observed personal data	Tertiary	Private Public
C7	TID Index	Ongoing	Health	Worldwide	1. collaborative effort to support wide-scale research projects	Disclosed personal data	Secondary	Private No profit
C8	Needmap	Ongoing	Social Inclusion	Europe and Central Asia	5. external responses to structural problems	Disclosed personal data	Primary	Private No profit Civil society
C9	Act Now Coalition	Ongoing	Security/ Public safety	North America	3. continuous effort to improve structural responses	Observed non personal data	Secondary	No profit Public
C10	Siderwalk Toronto	Terminated	Infrastructure	North America	3. continuous effort to improve structural responses	Observed non personal data	Secondary	Private Public
C11	Irishboon	Terminated	Education	North America	3. continuous effort to improve structural responses	Disclosed personal data Observed personal data	Secondary	Private Public

### 4. In-depth literary review

Given the scarcity of literature about DC governance, the review comprised also a broader research field, cross-sector social partnerships. This allowed for a more robust definition of the concept of governance and its components.

Building on the set of studies analyzed by Bryson et al. (2015) [9], the components of collaborative governance were identified: processes, structure, and elements at the intersection of these.

Subsequently, the study adopted the following definition of DC governance:

*“The formal and emergent processes, structures and elements that lies at the intersection of them, of decision-making and management that engage people constructively in data-driven activities across the boundaries of public agencies, levels of government, and/or the public, private and civic spheres for a societal purpose that could not otherwise be accomplished.”*

From the combination of DC and cross-sector social partnerships literature, seven dimensions of governance relevant to the systematization of DC that can be sustainable in the long term were identified:

1. *Initiation.* Governance factors that lead to the creation of a new DC.
2. *Trust.* Governance factors that reduce efforts in risk-taking by one organization toward other organizations within the DC.
3. *Formal structure.* Governance factors deliberated and agreed by DC partners through formal procedures.
4. *Intermediation.* Governance factors related to brokering activities toward partners to achieve collaborative outcomes of the DC.
5. *Incentive system.* Governance factors that motivate partners to continue sharing data and resources to implement the DC.
6. *Business model.* Governance factors that ensure the correct match between activities to be realized by the DC and available resources.
7. *Adaptation.* Governance factors that allow DC organizational arrangements to change in response to endogenous or exogenous forces.

## 5. Results and discussion

The analysis of the cases revealed the presence of twenty governance critical factors for long-term stability of DC initiatives.

Table 2 Matrix of theoretical dimensions and critical factors

Adaptation	Business model	Incentive system	Intermediation	Formal structure	Trust	Initiation	Governance dimension/ Critical Factor	
						C1 C2 C3 C7 C9	Presence of a clear need, interdependence	1
		C3 C8			C1 C3	C1 C2 C3 C6 C7 C8 C9	Pressure and facilitation by public institutions	2
					C3 C4 C5 C6 C7 C8 C9	C1 C2 C3 C4 C5 C6 C7 C8 C9	Relationship network, reputation and previous experience, individual leadership	3
			C3 C6 C8		C1 C6 C9 C11		Competence and data expertise	4
		C1 C3	C1 C3 C5 C6			C1 C3	Innovation, technology and data infrastructure	5
					C2 C4 C6 C7 C8 C11		Vision and mission alignment	6
			C5 C9		C1 C3 C4		Privacy and regulation compliance	7
				C4 C5 C6 C7 C8			Operational agreements with partners	8
			C2 C3 C4 C6 C9				Data standardization and sharing facilitation	9
				C2 C3 C6 C7 C9			Non disclosure and data limitation agreements, presence of external committee	10
			C1 C2 C3 C4 C5 C6 C7 C8 C9				Presence of a dedicated and neutral intermediary	11
		C4 C8	C3 C8		C11	C6 C7 C10 C11	Partners and stakeholders' engagement	12
		C1 C2 C3 C6 C7 C8 C9					Interest in data outcome	13
		C4 C6 C7 C8 C9	C4 C9		C1 C2 C4 C11		Non-profit legal form	14
		C1 C4 C5 C6 C8 C9			C4 C6 C8		Social responsibility and impact measurement	15
	C1 C2 C3 C4 C5 C6 C7 C8 C9						Private donations and grants	16
	C1 C2 C3 C4 C5 C6 C7 C8 C9						Revenues generation	17
C1 C6 C8					C1 C2 C3 C4 C10 C11		Communication and transparency	18
C2 C11						C3 C4 C11	Lean development	19
C1 C2 C3 C4 C5 C6 C7 C8 C9							Openness towards new partners and partner selection process	20

The results have been interpreted according to the nature of governance component identified in the definition: processes, structures, elements the lies at the intersection between processes and structures. This led to the identification of seven categories of critical governance factors that influence long-term stability of the DC.

**Table 3** Governance critical factors for DC long-term stability

Governance component	Category	Critical factor
Process	Data Related Processes	Privacy and regulation compliance
		Data standardization and sharing facilitation
	Engagement	Pressure and facilitation by public institutions
		Vision and mission alignment
		Partners and stakeholders' engagement
		Social responsibility and impact measurement
		Communication and transparency
Openness towards new partners and partner selection process		
In between process and structure	Leadership	Presence of a clear need, Interdependence
		Relationship network, Reputation and previous Experience, Individual Leadership
	Collective intelligence	Competence and data expertise
		Interest in data outcome
	Business model	Private donations, Grants
		Revenues generation
Lean development		
Structure	Data Related structures	Innovation, technology, and data infrastructure
	Management structures	Operational agreements with partners
		Non-disclosure and Data limitation agreements, Presence of external committees
		Presence of a dedicated and neutral intermediary
		Non-profit legal form

#### *Data related processes.*

These processes are associated with the management of DC-specific data flows. Compliance with privacy and regulations is a key process in building trust between partners, especially with respect to data owners and data sharers. Other critical processes are those related to facilitation in data sharing through the use of a shared and secure technology infrastructure and data standardization, i.e., pre-processing so that the available resources can be used successfully with respect to the DC's purposes. The research found that all these processes are linked to the dimension of intermediation. This confirms what the literature argues about the need for the presence of technological intermediaries within the collaboration, also referred to as data stewards.

#### *Engagement processes.*

These processes are those aimed at engaging actors in order to stimulate collective action. Among these processes, there is the action by public agencies in providing funds and regulating their access according to specific needs in order to spur

project initiation and create incentives for actor participation. Validation of the DC by public institutions also reduces initial trust-building efforts toward new partners. Aligning the vision and mission of the collaboration around a specific social purpose helps nurture trust and prevent potential tensions. The social mission, especially when complemented by a measurement of the social impact generated, acts as an incentive for actors, with particular reference to corporate social responsibility programs of companies that hold high-value data. Involvement in decision-making processes of partners and stakeholders enables the DC to gain internal and external legitimacy and incentivizes actors to continue collaboration. Similarly, effective communication and transparency about the DC's activities nurtures trust and allows possible tensions or clashes between partners to be overcome. Finally, the results suggest the structural openness of governance to the entry of new actors. This needs to be assisted by a selection process based on ethical and technical criteria, so as to preserve integrity and efficiency while still ensuring the adaptability of the collaboration.

#### *Leadership.*

This category considers those elements at the intersection of processes and structures that motivate actors to participate in collaboration. The role of individual leadership is one of these factors. Through the network of relationships, past experiences and reputation, this person enables the acceleration of the initiation of collaborations and triggers the formation of bonds of trust among actors. This role is often emergent rather than deliberate and stands out primarily in the early stages. A second element is the presence of clear need and the recognition that this can be achieved through participation in the DC. Identifying and sharing the needs that are intended to be solved through collaboration helps create the sense of interdependence that is crucial in the decision to generate a DC.

#### *Collective Intelligence.*

This category comprises the elements at the intersection of processes and structures related to the set of knowledge and competencies available to the collaboration. A critical factor is clearly the possession of both the technical data capabilities and the skills in the relevant social domain to enable the achievement of established outcomes.

These capabilities increase trust in those who possess them, and their presence is a motivation for other actors to participate in the DC. Another critical factor, revealed to be one of the greatest incentives for actors, is the interest in the outcomes produced by data processing. Therefore, in order to ensure the sustainability of the DC, it is necessary not only to generate social value but also direct value in connection with the interests of individual actors, particularly towards the private sector.

#### *Business model.*

This category contains the critical factors at the intersection of processes and structures that enable financial support for collaboration. One element observed in all cases analyzed is the presence of public contributions or private donations. This component represents the largest source of financial sustenance for DCs, and therefore appropriate structures and processes must be considered in the organization of DCs to be able to secure this income on an ongoing basis. Secondly, the presence of alternative sources of income attributable to core or non-core revenues relative to the activity of the collaboration was also noted as critical. Many of the common revenue models related to the software industry can also be adopted by DCs. Securing a share of income from revenues allows for a higher degree of independence from possible fluctuations related to the strategies of politics or individual private parties. Finally, organizational and business models should be considered according to lean logic. Initial experimentation with small-scale models and flexible development with respect to environmental responses may prove critical to ensure long-term adaptability.

#### *Data related structures.*

This category refers to DC structures associated with the management and usage of data. The main factor that constitutes this category is the technological infrastructure that enables data collection, sharing, and processing. The presence of a performant, secure, and innovative technology solution is one of the reasons why actors take part in a DC. It is generally managed by a single organization that acts as a technology intermediary.

#### *Management structures.*

This category contains the governance structures that support collaboration among actors from a managerial perspective. Two types of formal agreements are recognized among these factors. The first is intended to clarify roles, responsibilities, and activities and to make membership in the collaboration explicit, binding the actors to continuation from a legal point of view. The second is intended to define the use of data, their limitations in dissemination, and conditions to be observed for risk reduction, such as anonymization and aggregation. The presence of an external committee is important to recognize the technical and ethical constraints associated with the undertaking of a project within the collaboration, minimizing the risks of data mismanagement. The committee can be involved in the process of selecting new partners. Also recognized among the critical structures is the presence of an organization that acts as an intermediary for collaboration among actors. There are additional advantages if this organization is an entity purposely established for the collaboration. Advantages include ensuring neutrality in case of tensions, independent management of resources and staff, gaining external accreditation and possible certifications. Finally, the study recognizes the importance of the non-profit legal form for the dedicated entity. The peculiarities of these forms generate trust among actors and confer legitimacy to the DC in pursuing the social purpose.

## 6. Conclusion

Research has identified the governance components that influence the stability of DCs and distinguish data-related components that distinguish them from other cross-sector social partnerships. The presence of a form of governance that ensures structural openness to new collaborations, the adoption of a legal form of nonprofit collaboration, and the presence of mechanisms that incentivize data donors through the direct creation of shared value are some of the most important findings that are original compared to the existing literature.

The main limitations to the study include the difficulty of accessing direct contact with case study referents and the longevity of the cases analyzed. Even though the sample analyzed reports a satisfactory number of cases and

represents one of the largest empirical contributions to the field of study, given the high internal inhomogeneity within DCs the analysis of a larger number of cases would ensure greater robustness of the results. Although the cases were selected from active ones that had passed the first pilot stages, since this is a rather recent phenomenon, the projects analyzed do not exceed a decade of activity. This opens the door for later studies that will be able to verify the research findings over a longer period.

In conclusion, the research results open up practical applications in DC systematization. The hope is that it may serve the activation of more and more projects that are sustainable over time and can generate high social impact.

## References

- [1] I. Sussha, M. Janssen, and S. Verhulst, "Data collaboratives as a new frontier of cross-sector partnerships in the age of open data: Taxonomy development," in *Proceedings of the Annual Hawaii International Conference on System Sciences*, 2017. doi: 10.24251/hicss.2017.325.
- [2] S. Verhulst, D. Sangokoya, and T. Govlab, "Data Collaboratives: Exchanging Data to Improve People's Lives," 2015.
- [3] F. Bartolomucci and G. Bresolin, "Data Collaboratives' archetypes definition and research priorities setting towards their long-term sustainability," 2022.
- [4] I. Sussha, "Establishing and implementing data collaborations for public good: A critical factor analysis to scale up the practice," *Information Polity*, vol. 25, no. 1, 2020, doi: 10.3233/IP-180117.
- [5] E. Ruijer, "Designing and implementing data collaboratives: A governance perspective," *Gov Inf Q*, vol. 38, no. 4, 2021, doi: 10.1016/j.giq.2021.101612.
- [6] I. Sussha and J. R. Gil-Garcia, "A collaborative governance approach to partnerships addressing public problems with private data," in *Proceedings of the Annual Hawaii International Conference on System Sciences*, 2019. doi: 10.24251/hicss.2019.350.
- [7] B. Klievink, H. Van Der Voort, and W. Veeneman, "Creating value through data collaboratives: Balancing innovation and control," *Information Polity*, vol. 23, no. 4, 2018, doi: 10.3233/IP-180070.
- [8] R. K. Yin, "Case study methodology R.K. Yin (2003, 3rd edition). Case Study Research design and methods. Sage, Thousand Oaks (CA)..pdf," in *Case Study Research: design and methods*, 2003.
- [9] J. Bryson, B. Crosby, and M. Stone, "Design and Implementation of Cross-Sector Collaboration Framework for Understanding Cross-Sector Collaborations," *Public Adm Rev*, vol. 75, no. 5, 2015.