POLITECNICO
MILANO 1863

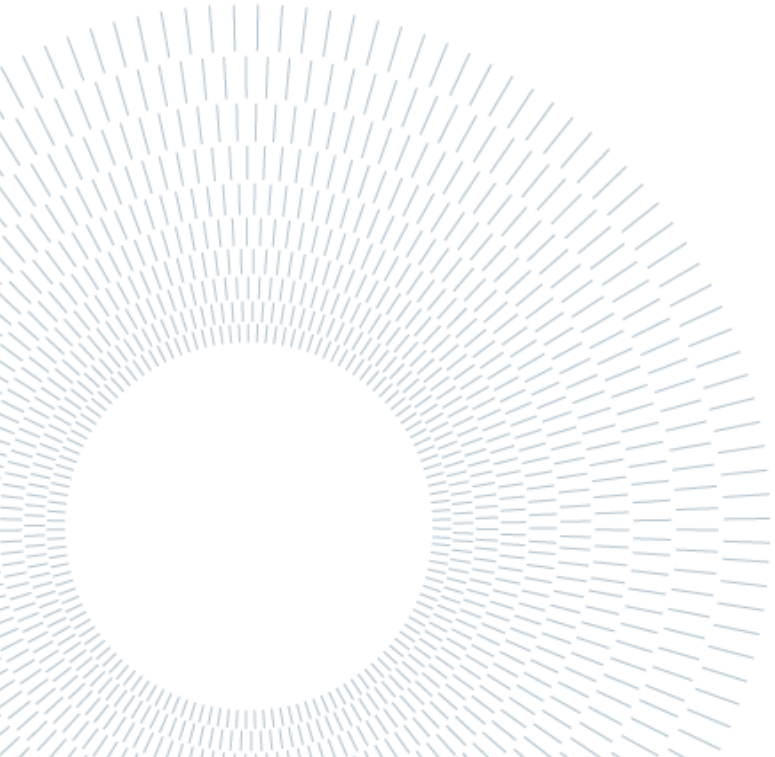
SCUOLA DI INGEGNERIA INDUSTRIALE
E DELL'INFORMAZIONE

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

TESI DI LAUREA MAGISTRALE IN
MANAGEMENT ENGINEERING
INGEGNERIA GESTIONALE

Author: **Simone Bartalucci**

Student ID: 966465
Advisor: Prof. Veronica Chiodo
Co-advisor: dr. Federico Bartolomucci
Academic Year: 2021-22



Abstract

The use of big data for the public good has been attracting increasing attention in recent years. A growing number of initiatives are demonstrating how the sharing and processing of datasets within cross-sector partnerships can generate an innovative answer to major economic, social, and environmental challenges.

However, these initiatives, referred to as Data Collaboratives, often fail to survive beyond the pilot phase or are designed to act over a limited time frame, thus having their ability to create long-term benefits and scale their impact hindered. The lack of well-defined incentive schemes for sharing data and resources, the absence of reference intermediation paradigms, and the lack of business models to economically support collaborations are among the governance-related variables that undermine the long-term stability of these types of collaborations.

To address these gaps, the study investigates the dimensions of collaborative governance in order to determine the critical factors that influence long-term stability. The first phase involved a review of the literature specific to Data Collaboratives and the broader literature on the governance of cross-sector partnerships. From this analysis, seven governance dimensions considered relevant to the long-term stability of a Data Collaborative were deduced. Then nine case studies were developed through interviews in order to inductively identify the relationships between these dimensions and sustainability. The results were compared with the analysis of two cases that ceased operations, providing further validation of the identified relationships. This enabled the identification of twenty critical factors, interpreted according to seven governance categories, to be considered in the design and implementation of a Data Collaborative.

In addition to providing useful insights for practice, the study significantly contributes in empirical terms to a field of research where the literature is still predominantly conceptual and illustrative.

Key-words: governance, collaboration, cross-sector partnership, data collaboratives, social innovation

Abstract in italiano

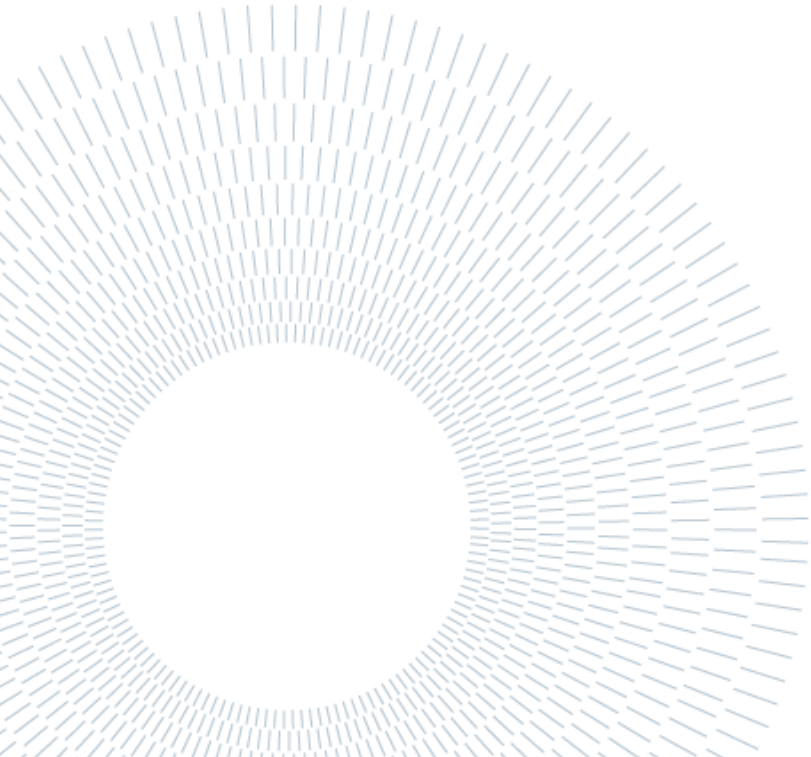
L'uso dei big data per il bene pubblico sta attirando sempre più attenzioni negli ultimi anni. Un numero crescente di iniziative sta dimostrando come la condivisione e l'elaborazione di dataset all'interno di collaborazioni cross-settoriali possa offrire una risposta innovativa alle grandi sfide economiche, sociali e ambientali.

Nonostante ciò, queste iniziative, denominate Data Collaborative, spesso non riescono a superare la fase pilota o sono progettate per agire su un arco temporale limitato, vedendo ostacolata la loro capacità di creare benefici a lungo termine e di scalare il loro impatto. La mancanza di schemi di incentivi ben definiti per la condivisione di dati e risorse, l'assenza di paradigmi di riferimento per l'intermediazione e la carenza di modelli di business per sostenere economicamente le collaborazioni sono tra le variabili legate alla governance che minano la stabilità a lungo termine di questo tipo di collaborazioni.

Per colmare queste lacune, lo studio indaga le dimensioni di governance collaborativa al fine di determinare i fattori critici che influenzano la stabilità a lungo termine. La prima fase ha previsto una revisione della letteratura specifica sui Data Collaborative e di quella più ampia sulla governance delle partnership cross-settoriali. Da questa analisi sono state dedotte sette dimensioni di governance considerate rilevanti per la stabilità a lungo termine di un Data Collaborative. Successivamente sono stati sviluppati nove casi di studio tramite interviste, allo scopo di individuare induttivamente le relazioni tra queste dimensioni e la sostenibilità. I risultati sono stati confrontati con l'analisi di due casi che hanno cessato la loro attività, fornendo un'ulteriore validazione delle relazioni individuate. Questo ha consentito di identificare venti fattori critici, interpretati secondo sette categorie di governance, da considerare nella progettazione e nello svolgimento di un Data Collaborative.

Lo studio, oltre a fornire indicazioni utili per la pratica, offre un contributo significativo in termini empirici in un campo di ricerca dove la letteratura è ancora prevalentemente concettuale e illustrativa.

Parole chiave: governance, collaborazione, partnership cross-settoriale, Data collaborative, innovazione sociale



Contents

Introduction

Background and context

The United Nations Secretary-General's Independent Expert Advisory Group (IEAG) defines the Data Revolution as an "explosion" in the volume and production of data, accompanied by an "increasing demand for data from all parts of society"[1]. The volume of data collected every day is growing at an exponential rate. The amount of data created, captured, copied, and consumed globally reached 64.2 zettabytes in 2020 and is expected to increase to more than 180 zettabytes in 2025 [2]. This process has even been accelerated by the COVID-19 pandemic, as more people worked and learned from home and used home entertainment options more often.

This data, also referred to as big data, is captured by a growing variety of digital technologies. Some examples of big data sources are satellite data, mobile network data, social media data, and Internet behavior data [3]. Most of them are increasingly collected and accessed by private companies [4]. These players control the data, financial resources and capabilities needed to harness its value and generate competitive advantage. In fact, the importance of data in the economy has become increasingly evident. More powerful computers and advances in algorithms such as machine learning have led to an explosion in the utility of data. This perception of data as a competitive and rivalry asset has led private actors to hesitate to share their data and knowledge [5], leading to a huge gap in data access and ability to exploit it between the larger economic actors, the public sector, and the non-profit sector.

Social, economic, and environmental challenges have become increasingly complex for individual organizations to solve. Many of these problems, such as climate change or poverty, are termed "wicked" because their magnitude is difficult to assess, and the cause-and-effect relationship is very complex [6]. The solutions to these problems are not clear, and individual central authorities do not have the resources or information to implement them. This means that it is becoming ever more necessary to launch collaborations with multiple organizations from different backgrounds. The United Nations Sustainability Goals are an important example of a joint effort to set the global agenda for targeted efforts to address these challenges. Central to these efforts is the recognition that solving many of the grand challenges requires collaborative and coordinated efforts in which the action of nongovernmental actors, such as business and civil society organizations, is vital. Specifically, SDG17 emphasizes the need to

create partnerships that transcend industries and sectors to comprehensively and effectively address the challenges [7].

The use of data enables a more effective evidence-based approach to address these challenges and should be considered as a strategic social resource. To unleash the full potential, this resource needs to be shared with actors who can implement high-impact programs. It is in this context that data sharing initiatives for social good have emerged.

Data sharing for social purpose

Thanks to the opportunities provided by new data sources, data analysis for social purposes is a topic that is gaining more and more attention [8].

Over the past decade, the open data movement has taken on a prominent role. The principle behind this movement is that data should be accessed, used and shared by everyone. Governments around the world have created open data portals to make data more accessible and usable by the public; other data archives have specialized in open data to foster research [9]. Managing data as a completely nonexcludable resource with no access restrictions removes barriers and potentially maximizes use and reuse [10], increasing opportunities for innovation.

However, the impact of open data is limited if they are not properly used [11] and their use is considered problematic for several reasons [9]. First, social innovation actors lack the technical capabilities to properly exploit them. Second, the quality of data is often inadequate with respect to its use, being superficial or not aligned with needs. Finally, the dissemination of high-value data, such as personal data, could instead open the possibility of their misuse which can erode human rights and delay the achievement of the SDGs [10].

To overcome these critical issues and maximize impact, new ways of data sharing are needed. Recent years have seen the development of initiatives that combine and integrate public open data with data from the private sector, NGOs, and individual citizens [12]. These programs have taken a variety of names, such as Data for good, Data donation, or Data philanthropy, but are recognized in the literature predominantly by the name Data Collaborative (DC) [12]. DCs are more structured forms of data sharing initiatives. They involve formalized partnerships between actors from different sectors [13], primarily the public, private, and non-profit sectors. Within these collaborations, high-value data, often from the private sector, are shared and processed in secure, aggregated, and anonymized form [14]. Unlike open data initiatives, the data shared are bound toward a precise social impact activity [15], which allows the data supply to be tailored to the needs of the actors in the early stages of the collaboration. Furthermore, data dissemination is restricted to members of the

collaboration, making its use controllable by the data owners, thus limiting the risks of harmful use [16].

While DC model entails clear advantages, it also raises critical issues in their management. In fact, many of these projects have difficulty lasting over time. They are often designed as occasional programs or fail to pass the first pilot phases [12], [17]. Many of these issues can be attributed to the governance of these projects [12], [17]–[19]. For example, the presence of an incentive system for data donors, a system that ensures security and minimizes the risk of use of private data, and a business model that provides financial support for the initiative are all organizational elements that can determine the success or shutdown of projects in the long run.

Scope of the research

The goal of this research is to advance theory on DC, a new field of study in which literature is still developing. In particular, the elements that make up the governance of these collaborations will be investigated to promote long-term sustainability and impact generation. In addition, this research seeks to analyze existing knowledge and complement it with new interpretations based on direct analysis of real cases. The lack of fit between theoretical research and empirical evidence is explicitly highlighted in the field of data collaborations [20].

This research included an initial phase of literature review on DCs to understand the characteristics of these types of collaborations and identify critical aspects. Given the small number of studies on the governance of DCs, the topic was further explored by analyzing the literature on cross-sector social partnerships. From these theoretical analyses, dimensions of collaborative governance that are critical for project sustainability were deductively extrapolated. A multiple-case study was then conducted through secondary source analysis and direct interviews with members of a selected sample of DCs. Finally, relationships between those dimensions and real cases were collected and critical factors within the governance of DCs were inductively identified.

This thesis work is structured as follows: in the first section, a review of the existing literature on DCs is given; in the second section, the questions this research seeks to answer are formulated in detail; in the third section, the literature review on cross-sector social partnerships is given and the theoretical dimensions of governance are presented; in the fourth section, the methodology applied to the research is presented; in the fifth section, the results extrapolated from the case study analysis are illustrated; in the sixth section, a discussion of these results is formulated; and in the seventh section, concluding statements and major implication of the study are given.

1 Preliminary literary review

Although data partnership initiatives offer real opportunities to address pressing societal challenges, these projects struggle to evolve beyond the pilot phase or are constrained to events of limited duration. The stability of such projects over time can lead to broader and ongoing benefits to society. The first phase of this research is aimed at determining the most significant aspects of these partnerships, surfacing critical issues that may undermine the sustainability of the projects through a review of the literature.

Research in the field of Data Collaboratives is scarce but has been growing in recent years [8]. Existing studies mostly focus on descriptive enrichment of a still developing literature. The research developed so far focuses on different areas and draws on notions from many neighboring fields of study: examples include open data, data partnerships, data for good, public-private partnerships, and cross-sector social partnerships. This research was done primarily through keyword searches in online databases, such as Scopus and Google Scholar. The keywords used are "data collaborative", "data for good", and "data partnership". In addition, the snowball method was used, searching for relevant document sources to expand knowledge in this field. In particular, a search was conducted on the papers of the two most relevant authors in the field of data collaborations, S. Verhulst and I. Susha.

1.1. Data Collaborative Definition

The term Data Collaborative is relatively new, though the phenomenon is not necessary novel [12], [21]. As a matter of fact, these kinds of initiatives are also referred to in the literature by other names, such as data for good, data donations, or data philanthropy [12]. However, the use of the term Data Collaborative is currently more widespread. A collaborative is an organized group of people or entities who collaborate towards a particular goal [22], so the choice of that name emphasizes collaborative processes over mere data disclosure [16].

The first definition shared in the literature is that provided by Verhulst at al. (2015) [13], who define it as "a new form of collaboration, beyond the public-private partnership model, in which participants from different sectors - including private companies, research institutions, and government agencies - can exchange data to help solve public problems."

Further progress in determining the boundaries of the phenomenon was made through the taxonomy developed by Susha et al. (2017) [15], which provided the most accepted definition at the time. They define Data Collaboratives as "cross-sector (and public-private) collaboration initiatives aimed at data collection, sharing, or processing for the purpose of addressing a societal challenge".

The key aspects of these collaborations clearly emerge from the definition: cross-sectorality, that is, the involvement of actors with different backgrounds; the centrality of data in creating value; the generation of impact in addressing social challenges. Although this definition clearly specifies the core elements of a DC, it does not exclude another field of similar initiatives, namely those based on open data.

The difference between these two fields lies in a different interpretation of the concept of data access. Open data initiatives are based on the theoretical concept that data is a public good, which implies that data is a nonrivalrous and nonexcludable commodity, so should be universally accessible to all and reusable for any purpose [12]. In contrast, DCs advocate the sharing of data among a small and well-defined set of actors and the finalization of their use to a predetermined goal, therefore belonging to the category of closed data partnership [4]. While recognizing the nonrivalrous nature of data, the notion of DC proposes that data can be excludable, with access limited to authorized members, shifting from viewing data as a digital public good to viewing data as a club good [10], [23].

1.2. Actors involved

DCs have been conceptualized as a form of cross-sector collaborations. Indeed, Susha et al. (2019) [12] recognize the three distinct spheres to which actors can belong, each with its own logic, interests, and value propositions: the state, the market, and the non-profit. DCs are not a homogeneous phenomenon and can be configured differently based on the involved parties [24]. It is possible to find all four possible combinations among sectors [12]: public-private partnerships, public-nonprofit partnerships, for-profit-nonprofit partnerships, and tripartite partnerships involving actors from all of these sectors.

In addition to the three main sectors identified, other types of actors that can play a significant role within DCs should be mentioned. The academic sector, although often under public management, is controlled by different interests, namely the generation of new knowledge rather than the results of its application. In particular, within academic institutions, researchers and universities (both as institutions and as individual researchers) are recognized. Partnerships with universities and researchers are particularly valuable to companies, which see them as valuable investments that complement in-house research and development [25]. However, problems can arise in

data partnerships related to the different goals of academics and other parties and the appropriateness of results [26]. Civil society can also be involved in these kinds of initiatives. Especially, citizens can play an important role by providing content, in the form of user-generated data [27], or by lobbying on specific societal issues to attract the attention of other sectors.

While each actor shares the goal behind the collaboration, their participation is driven by different motivations [28], making collaboration one of the main challenges faced by data initiatives [16]. Relative to each of the three main categories of actors within the DC, the value they can bring, their interests, and possible critical issues for the stability of the initiative will be described below.

1.2.1. Private sector

The private sector, also referred to as the market sector, usually plays a primary role in data sharing and sharing of technological capacity for data use [8]. Many of the data that can be used to address social challenges are owned and handled by private entities [13], [29], also referred to as Non-traditional Data [30]. Technical expertise developed in the private sector can be shared with other actors by addressing data quality and standards issues, providing and training skilled data professionals, thus building a data culture within the collaboration [31].

The main motivation for this sector to take part in these kinds of collaborations is the search for new knowledge and profits [32], [33]. This search takes the form of sharing data to improve business model, through an improvement in the company's brand image, its offerings or, more rarely, its efficiency [27]. In some cases, private companies can obtain new analytical insights and more representative information about their population of interest [34]. Another component that drives private actors to participate in DCs is that related to corporate social responsibility programs [8], [12].

Initial barriers and risks to the stability of participation of these actors are mainly related to the sharing of data with high economic potential. Their concerns regard data management in terms of privacy and security [35] and the absence of direct economic benefits or the presence of disincentives related to possible threats to business interests or maintaining competitive advantage [27], [35], [36].

1.2.2. Public sector

Public agencies can play an important role in the initiation and continuation of a DC. They can act as facilitators and data providers in data sharing projects [21], or, in most cases, stimulate and facilitate collaborative action of the other actors [27]. Susha et al. (2019) [28] recognize multiple levels of government and public agency involvement in cross-sector data partnerships: an active role in helping to define win-win scenarios that incentivize private data dissemination; a passive role in which it is primarily a

beneficiary of project outcomes, and thus facilitates it through policymaking and incentives, but also by reaping the benefits generated by data sharing and analysis; and a dual role, in which it is called upon to create favorable conditions through pressure and to play an active role in identifying benefits for other parties.

The primary incentive for this kind of actor is related to the idea of the "common good" [37]. In these projects, the goals of public bodies are to achieve economic growth, develop innovation, or solve known public problems [32].

Public bodies involvement is more common in the case of projects that address social issues that are influential in terms of public opinion [38]. This can generate some problems with relevant projects that nevertheless do not strongly influence public opinion and public perceptions of governments; this effect is exacerbated in low- and middle-income countries [38], which are constrained by lower technological literacy of the public sector.

1.2.3. Non-profit sector

The non-profit sector includes organizations with different legal nature and operational structures, including charities, philanthropic foundations, voluntary and community organizations, community groups, social enterprises, and co-operatives. Within DCs, the main role of the non-profit sector is to ensure the expertise in the social sector for generating impact from the use of data. In particular, the knowledge of beneficiaries due to their contact with communities, especially the most disadvantaged ones, serves to align data provision with real demand needs. Another role that is recognized for these types of organizations is that of project initiators, being able to exert pressure on private actors by leveraging corporate social responsibility programs and encouraging data donations [28].

Key motivations for this sector's participation include the need for funds, resources and new expertise to support high-impact activities. The progressive marketization of this sector has led individual organizations to compete for public or philanthropic funds, generating a need for accountability to funders [39]. As a result, non-profits are required to report information regarding their activities. Given the difficulty in finding the necessary data, they are often accused of overemphasizing output data that are easier to collect, as opposed to outcome or impact data [40]. Competition has also led these organizations to seek new sources of competitive advantage from data, as evidenced by the increasing investment by non-profits in data professionals and specialized data consulting [39]. While the need for information is becoming increasingly clear, major deficiencies in the ability to manage data persist. The lack of adequate investment and expertise in social data analytics leads to difficulties in adopting innovative solutions, resulting in the phenomenon referred to as the non-profit starvation cycle [41]. Participation within initiatives such as DCs allows for new data expertise and enables the implementation of more effective innovative programs.

The critical aspects in the participation of this sector lie precisely in the difficulties mentioned. The scarcity of funds, a structural problem of these realities, leads to the difficulty of acquiring figures with data skills or allowing adequate training of existing staff, necessary for the effective exploitation of data [42]. In addition, non-profit organizations suffer from cultural problems in approaching data [31] and struggle to develop a technological vision. This leads to an underestimation of the potential of technological innovations and their application to the social sphere, limiting the willingness to work toward establishing a DC.

1.3. Data sharing

The second key element identified in the definition of DC is the generation of added value from the data. As previously written, these types of initiatives differ from open data initiatives in the exclusive nature of data access, which is limited to members of the collaborative. This makes the data sharing process particularly significant for the stability of the initiative.

One of the basic prerequisites for implementing a DC project is matching information needs with possible data sources. Matching the demand for data and its supply is recognized as one of the biggest challenges [13], [16]. The demand for data depends on the purpose of the collaboration and the social field to which it applies. Examples are provided by Verhulst et al. (2017) [35], who identify five value propositions that can constrain data demand: i) situational awareness and response, ii) knowledge creation and transfer, iii) public service design and delivery, iv) prediction and forecasting, and v) impact assessment and evaluation. Supply, on the other hand, is generally the bottleneck; inaccessibility of data is cited as one of the challenges inhibiting the emergence of data sharing initiatives [43]. The private sector benefits from greater data availability, but the motivations for companies to disclose their data are often considered inadequate [13], [27], [36].

In the area of data supply, different types of data can be distinguished. A first distinction is provided by Susha et al. 2017 [15], who distinguish between primary, secondary, and tertiary data based on the reasons why the shared data were initially collected and the reasons why they are used in the partnership. Primary data are those sourced specifically with a view to the data sharing project; secondary data are collected for a purpose similar to that for which they are used in the DC; tertiary data are collected for substantially different purposes than those for which they are used in the DC. This classification may also play a role in defining collaborative governance. The use of tertiary data, given the increased risks to those who share it from its non-social nature, leads to the need for more rigorous data governance and formal agreements [21].

Among these types, the use of primary data is preferable in terms of quality, but due to higher cost and time, it is a rarely used option [15]. In addition, data sets may still

have biases, as in the case of data collection in disadvantaged environments with poor access to technology. Secondary and tertiary data are of lower quality but are the most common option. Within these categories are Non-Traditional Data (NTD). NTDs are defined as data acquired, mediated, or observed through new digital instrumentation mechanisms, often privately owned and used for purposes unrelated to their initial collection. [30]. The use of NTDs allows for increased speed of scale, considerably limiting the collection effort. Chafetz at al. (2022) [30] distinguish between four areas of interest of NTD sources, shown in the following image.

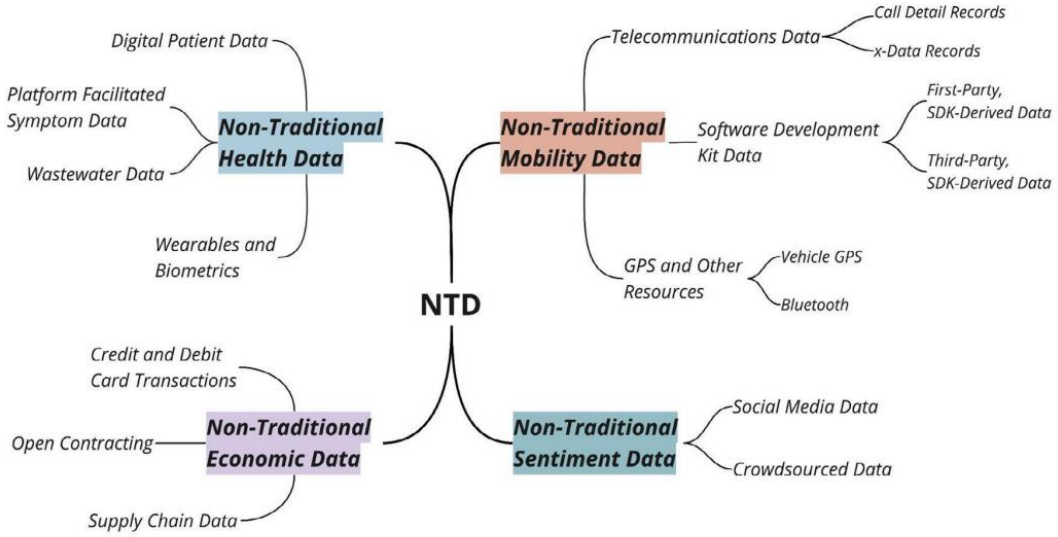


Figure 1 Non-Traditional Data sources from Chafetz at al. (2022)

Among the most critical factors within the processes related to data use is the quality of the data itself. Based on research and expert assessments, it has been recognized as the single most influential factor in the outcome of a data collaboration [8]. It is important to find the right data, with attributes and granularity suited to the specific question [16]. In addition to the quality of individual datasets, Global Partnership for Sustainable Development Data [44] considers the broader concept of dependability. Dependability encompasses two aspects: the reliability of data sharing systems and the quality and appropriateness of the information created with respect to user needs. Different approaches can be used to maintain trustworthiness, but generally these three characteristics arise: i) ensuring that technical safeguards are in place; ii) incorporating quality assurance mechanisms into data collection and analysis; and iii) adopting appropriate approaches to interoperability and data format. Although several standards and procedures for ensuring data quality have been proposed in the literature [45], it is not easy to establish a universal metric considering the high variability in DC projects. It is important to note how levels of trust among partners

can also influence the attention given to technical guarantees and quality assurance [44].

Another critical element is the presence of the right skills and technological structures within the DC. The demand for skilled professionals is high in the world of data science, far more than their supply [43]; moreover, few data scientists choose to work in the social sector [43]. The lack of adequate knowledge and tools leads to the risk of generating erroneous insights [35] that could negatively impact against the social purpose. As reported earlier, the greatest shortcomings in this area are found in the non-profit sector. To ensure the success and stability of projects, it is important that the conditions for appropriate dissemination of these skills are promoted. This can be done through collaboration with more data-savvy actors [31] or through dedicated programs to strengthen the internal capacity of these organizations.

Finally, still related to data processes, violation of privacy and security requirements can be a source of risk to DC stability. Indeed, the most common concern expressed by individuals and companies relates to the fear that information sharing may result in the disclosure of personally identifiable or demographic information, impacting privacy or creating security breaches [35].

The data may contain personal or other information that is sensitive to the data provider, putting the continued collaboration at risk if it is misused. Maintaining privacy requirements is more difficult in the case of the use of tertiary data, such as with the data provider's customer data, where consent at the data collection stage may have been provided in relation to a completely different purpose than that of use in the DC [16]. Moreover, compliance with privacy regulations poses even more complex challenges for projects that span multiple jurisdictions, as the scopes of application of privacy standards and regulations can be uncertain [45]. Even where non-sensitive data are used, a risk analysis must equally be conducted in order to identify potential harms to individuals and communities related to data use [46].

Some of the possible solutions to mitigate these risks are to use precise privacy agreements or to anonymize and aggregate data before it can be shared. In general, it is considered appropriate to include a coordination mechanism by transferring data to a designated specialist responsible for data sharing, or the so-called data steward [16].

1.4. Social Purpose

The last element included in the definition of data collaboration is the use of data for social impact generation. Given the purpose of these types of collaborations, it is important that the impact generated is clear and measurable for the long-term stability of these initiatives.

Some of the applications of data for impact generation emerge from literature. Verhulst and Sangokoya (2015) [13] recognize three main ways of generating impact from DCs.

1. *Data-driven decision making*. The use of data enables politicians and other decision makers to address major societal challenges more effectively. The benefits brought, particularly using private data to public institutions, have been explored in depth by Mcmurren et al. (2020) [47]. They recognize value in four aspects: i) increasing the scope and breadth of available evidence, ii) increasing the quality and credibility of existing datasets, iii) increasing the timeliness and of often outdated information held by statistical agencies, and iv) reducing costs and increasing the efficiency of statistical organizations.
2. *Information exchange and coordination.* DCs also bring value by creating platforms for information exchange that can accelerate research and development of innovations. One example is in the field of clinical drug trials, where through DCs scientific data from drug trials can be shared. Otherwise, this data are in the hands of private companies and inaccessible to independent researchers and citizen groups that could have added information on safety and effectiveness.
3. *Shared standards and frameworks to enable multi-actor, multi-sector participation.* These initiatives can also generate value more indirectly through synergies within the data community. The spread of DCs can enable the emergence of standards and frameworks needed to make data interoperable and useful across organizations and sectors, mitigating the difficulties of integrating data from different sources and structures.

In addition, Verhulst and Young (2017) [35] delve into the potential applications of a particularly significant subset of data held by private actors, namely those from social media. The authors associate five applications to this type of data for social purposes.

1. *Situational awareness and response.* Social media data can help the non-profit and public sectors better understand demographic trends, public sentiment, and the geographic distribution of various phenomena. This type of application is particularly useful in emergency contexts. One example of this is in the report The #Data4COVID19 Review [30].
2. *Knowledge creation and transfer.* DCs on data can unify widely dispersed datasets, thereby creating a better understanding of possible correlations and causalities, as well as which variables make a difference for a given type of problem.
3. *Public service design and delivery.* Private datasets can provide tools for more accurate modeling of public services and help guide targeted, evidence-based service delivery.
4. *Prediction and forecasting.* Richer and more comprehensive information from DC enables new forecasting capabilities for the public and non-profit sectors, allowing them to be more proactive.

5. *Impact assessment and evaluation.* DCs can contribute to monitoring, evaluation, and improvement. Public sector actors can quickly evaluate the results of their actions to iterate programs when needed. This aspect is also recognized by Farmer et al. (2023) [39] in relation to activities carried out by non-profit organizations.

Regarding scope, examples of DC application can be found in all major social challenges. For example, in the database of DC cases in datacollaboratives.org, the cases are divided into eleven areas:

1. Agriculture;
2. criminal justice;
3. crisis response;
4. digital society;
5. economic development;
6. education;
7. environment;
8. health;
9. infrastructure;
10. transportation;
11. telecommunications.

Therefore, it can be said that the use of data is an enabler for the development of innovative processes. The scope of application of these processes for social challenges may be limited by the vision capacity of the actors involved only.

Finally, even though measuring social impact is recognized as a critical element that can accelerate the diffusion of DCs [13], there appears to be no shared method in the literature for defining and measuring the impact generated by DCs and it is still a generally understudied topic [12].

1.5. Data collaborative classification

In order to identify the basic characteristics by which DCs are distinguished, with a focus on those that determine long-term projects, it is useful to look at classifications in the literature. Verhulst et al. (2015) [13] introduced a classification of DCs for the first time considering two variables: Engagement and Accessibility.

By Engagement they mean the degree to which the data supply and demand actors co-design the use of corporate data assets. Three levels of engagement are recognized among the DCs:

- *Independent use*. The analysis and the use can be completely independent, meaning that the private sector data owner has little direct involvement in data reuse.

- *Cooperative use.* Data providers and users decide the purpose of data use and analysis collaboratively.
- *Direct Use*. The data owner seeks partners to derive specific types of public and priority value from the data.

By Accessibility, on the other hand, they mean the conditionality of access to private data by external parties and distinguish between two cases:

- *Open Access*. These initiatives place very few restrictions on the exploitation of private sector data, in some cases allowing the general public to view or download certain data resources.
- *Restricted access*. In DCs with more restricted access, only preselected partners preselected partners receive access to corporate data resources.

Combinations of these variables result in six possible DC categories.

	Open Access	Restricted Access
Independent Use	**Public Interfaces**	**Trusted Intermediary**
Cooperative Use	**Data Pooling**	**Research and Analysis Partnership**
Directed Use	**Prizes & Challenges**	**Intelligence Generation**

Figure 2 Data Collaboratives Matrix of Engagement and Accessibility from Verhulst et al. (2019)

Verhulst et al. (2019) [14] applied this classification to the GovLab database, which as of September 2019 included more than 150 cases. The result is the following distribution of categories: Research and analysis 25%, Public interfaces 20%, Trusted intermediaries 20%, Intelligence generation 15%, Data pooling 12%, Prizes and challenges 8%.

While recognized as a legitimate high-level categorization, Bartolomucci et al. (2022) [48] point out its limitations. First, the classification is based on only two variables, ignoring significant others such as their duration and the intentionality of the impact generated. The categories identified are then heterogeneous in their definition. Some of them refer to the mode of data sharing (Data Pooling), others to the actors involved in the collaboration (Trusted Intermediaries, Intelligence generation), or to the scope of the collaboration (Research and Analysis), and still others to a type of incentive (Prizes Challenges). Finally, initiatives based on access to open data or initiatives driven by a single company are also included, violating some of the essential elements of the DC definition applied to this research.

Bartolomucci et al. (2022) [48] propose their own classification based on a cluster analysis applied to a dataset comprising 135 cases. The authors identify five clusters, placing greater emphasis on aspects such as long-term stability and clarity of the social impact generated.

1. *Collaborative effort to support wide-scale research projects*. To this cluster belong collaborations that aim to enhance research in different areas, with particular emphasis on international initiatives that provide new solutions to problems through the sharing and reuse of data from different sources. Generally, the academic and research sector is involved, they clearly have a societal purpose and take place over a long-term time frame.
2. *Prompt response to emergencies*. To the second cluster belong projects established to provide urgent responses to humanitarian crises caused by natural disasters or health emergencies. Their social objectives, as well as the duration of their programs, are closely linked to the emergency situation. This indicates that the initiatives in this cluster are extemporaneous, whose conception and execution depend on particular conditions and do not generate an impact beyond the emergency condition.
3. *Continuous effort to improve structural responses*. Projects in this cluster are aimed at implementing data-driven improvements to existing services and infrastructure, or in developing sophisticated systems that leverage data for effective emergency prevention (rather than recovery as in Cluster 2). If the projects in Cluster 2 are created in response to emergency conditions, those in Cluster 3 take the form of improvements to the capacity of cities, governments, and continents to respond to them, aimed at generating structural and lasting improvements.
4. *Data-driven initiatives to support innovative studies*. Projects in the fourth cluster is motivated by the availability of data rather than the resolution of a social need. These collaborations excel at accelerating data-driven innovation by identifying unexpected value in data and developing new pathways to use data. This type of initiative is often data-driven and not constrained by a specific goal. The data used are often unrelated to the problems for which they are being used, meaning that initiatives in this cluster seek to disclose data that might be of public interest in order to discover new uses for it. This shows that not only do these initiatives not have a clearly defined goal, but neither does their field of impact.
5. *External responses to structural problems*. The fifth cluster is defined by international initiatives undertaken to address structural problems especially in less developed countries. These collaborations are often based on the aspiration of foreign institutions or companies to solve structural problems through data in countries that lack the capacity and knowledge to do so. This is the case with projects aimed at improving people's living conditions, such as ensuring access

to basic services such as water and food, or ensuring adequate access to medical care, as well as addressing systemic and long-term issues that affect livability, such as infectious diseases and environmental migration.

1.6. Critical success factors

The literature review revealed the distinctive aspects of these types of collaborations and some of the elements that can determine their success or failure. Regarding these elements, it is relevant to devote space to the study conducted by Susha (2020) [8] regarding critical success factors. The study first identifies success factors reported in the literature, then assigns an order of importance in accordance with expert opinion, finally validates them by the analysis of two case studies. Fifteen key factors emerged from the study divided into three main categories.

- *Organizational factors.* The organizational dimension is related to the coordination of collaboration from the managerial point of view. The author distinguishes two stages of DC in which organizational factors affect: initial conditions and collaborative experience. Regarding the factors that affect the beginning of DC there is the presence of the incentives for the actors, the value proposition of the collaboration, the business model, the matching of data supply and demand, and the resources available to the collaboration. The presence of incentives appears to be particularly impactful on processes and the risk of collaboration failure and is recognized as the second most agreed upon factor between experts. Among the factors that, on the other hand, influence the continuation of collaboration, the author notes shared understanding, trust among actors, stakeholder participation, leadership, the presence of clear responsibilities and the use of a common terminology.
- *Technological factors*. Technological factors are those related to the use of data. According to the study, these factors mainly affect the early stages of collaboration. Among them, data quality, interoperability of technologies, and technical capabilities for data analysis are considered of greatest importance. Data quality is considered by the study to be the most critical factor, capable of both determining project failure and significantly affecting the results of the collaboration.
- *Environmental factors.* These factors are those determined by actors outside the DC. Only one element emerges from the study, namely public support, which affects in the formation phase of the collaboration. For experts, it plays an important role in the processes, but not of the first order.

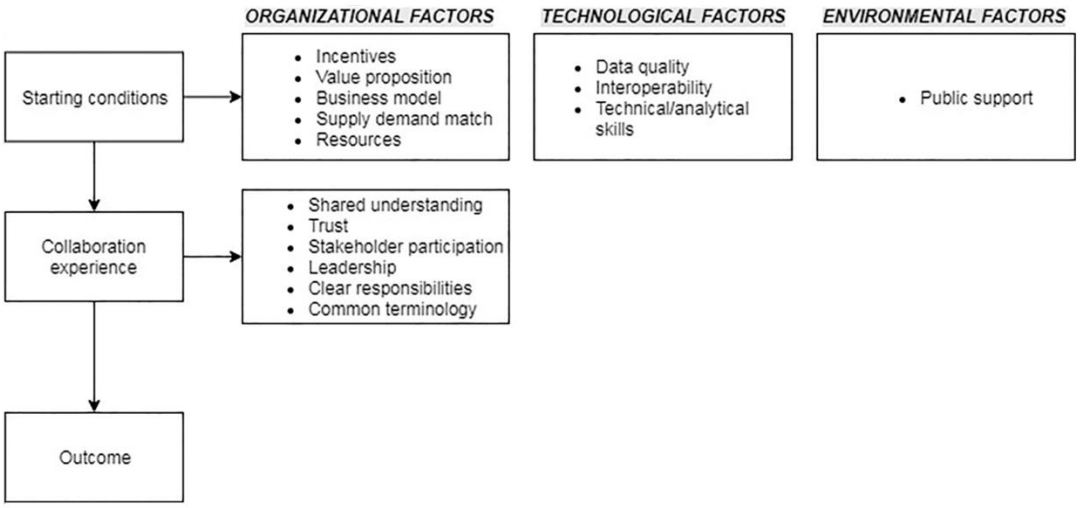


Figure 3 Top most critical factors from Susha (2020)

This study summarizes and confirms the most important elements for DC success noted in the literature. However, it should be pointed out that it does not distinguish between success stories in short-term projects and those that are sustainable in the long term. The definition of success applied to the study considers only compliance with predetermined goals and expectations of stakeholders. Although useful information is provided for the systematization of DC training, it lacks the temporal dimension of carrying out the initiative.

Another aspect that emerges from this study is the relevance of organizational factors. DCs are a socio-technical phenomenon, in which the technological and organizational dimensions are linked by mutually influential relationships. Although data quality is the critical factor that finds most consensus among experts, eleven of the top fifteen factors belong to the organizational dimension. Furthermore, the literature shows the influence of trust on the quality of information shared [44], thus placing organizational aspects on a preponderant level in determining the success of a DC.

2 Literature gap and research questions

2.1. Literature gap

After having reviewed existing literature in the field of data collaboratives, a clear picture of the extant knowledge on the topic was defined. Data Collaboratives are a

fairly new field of research, and many questions are still open for research. Although there are studies emphasizing the critical factors that can lead to collaboration failure, there is a lack of research geared toward the long-term sustainability of DCs. Since projects that offer long-term continuity are able to generate significantly greater impact than one-off initiatives, the concept of success should also be redefined on the basis of durability.

It is clear from the literature that governance and organizational components are of primary importance to the sustainability of DCs. The lack of well-defined incentive schemes for public and private actors [21], [28], the absence of effective referral schemes for intermediation [21], [26], [49], and the lack of business models that can make these collaborations economically viable [8], [17], are among the governance-related variables that undermine the long-term stability of these types of collaborations.

Unlike generic cross-sector collaborations, DCs have complexities related to their socio-technical nature [12] and therefore studies in this area are insufficient. However, research on this topic is still limited. There are currently three studies that explicitly address the issue of governance for DCs [20], [21], [37]. Each of these studies approaches the problem in a deductive way: they start from a specific framework developed for cross-sector social partnerships, apply modifications based on theoretical knowledge of DCs and compare it with the analysis of a single case study. Klievink et al. (2018) [21] draw inspiration from Ansell et al. (2008) [50] model, Ruijer (2021) refers to the model of Bryson et al. (2015) [51], Susha et al. (2019) [37] adopts the model of Emerson et al. (2012) [52]. Ruijer (2021) [20] himself points out that the existing literature is predominantly limited to conceptual or illustrative research compared to the number of empirical studies. An analysis of the three studies cited above confirms this shortcoming and opens the space for research that tests theoretical assumptions on a larger number of real cases.

2.2. Research questions

The purpose of this work is to advance knowledge on data collaboratives, with the final goal to foster long-term sustainability of these projects and to facilitate impact generation; in doing so, special care was taken to base each in-depth study on a sufficiently large selection of real cases, using interviews with direct witnesses as the main method of data collection.

To fill the identified gap, the research seeks to answer the following questions.

- *How governance factors influence the long-term stability of data collaboratives?* The first question concerns the lack of links in the literature between the elements that constitute the governance of DCs and long-term stability. The studies analyzed focus attention on the issue of governance and confirm its relevance

to the success of initiatives [18], [19], [37]. None of them, however, explicitly considers sustainability as an interpretive key to the success of DCs. In particular, there is no framework that identifies the governance components that play a critical role in the long-term continuity of projects. Therefore, the research aims to identify these components in order to illustrate the connections between governance and sustainability and provide useful guidance for practice.

- *What are the governance factors distinguishing DC from other forms of cross-sector partnerships?* The second question is aimed at providing evidence that can advance theory on DC governance differentiating it from theory about cross-sector social partnerships. Although the former initiatives can be considered a subset of latter, the technological and legal aspects associated with DCs add factors of complexity that require specific arrangements in governance. This study also aims to identify such factors and arrangements.

3 In-depth literary review

To answer the research questions, a literature review specifically about governance of Data Collaboratives was necessary.

The purpose of the review was twofold: to understand the meaning of governance of DC and to identify what dimensions of governance can be reasonably linked to the long-term sustainability of the collaborative. The starting point was a review of literature about DC. At the moment, the amount of information about the topic is rather scarce: just three papers are specifically about DC and governance [20], [21], [37]. For this reason, it was necessary to explore a broader topic. DC 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 [15]. From this definition, three elements distinguish a data collaborative: i) cross-sector collaboration, ii) data centrality and its implications, iii) social purpose. Considering these elements, Data Collaboratives are a subset of cross-sector social partnership, which includes the element i) and the element iii). Given the fact that several important frameworks have been developed about the topic [50]–[57], the review started from it. How the element ii) influences the collaborative governance is part of the research question.

The first part of the chapter illustrates the definition of governance of DC; the second part shows the main forms of collaborative governance; the third part is about the governance dimensions identified for lung-term sustainability.

3.1. Definition of DC governance

As for all collaboration, also for DC to be successful some form of governance is necessary [56].

Understanding what constitutes the governance of collaborations is not straightforward. While collaborations are horizontal systems, the concept of governance implies hierarchy and control, which is problematic [56].

Ansell et al. (2008) [50] define collaborative governance as follows:

A governing arrangement where one or more public agencies directly engage non-state stakeholders in a collective decision-making process that is formal, consensus-oriented, and deliberative and that aims to make or implement public policy or manage public programs or assets.

This definition suggests that collective governance involves formal processes and that there is active participation by all members. However, it emphasizes the predominant role of public entities in initiating collaborations and focuses just on activities involving public programs or public assets. For these reasons, it does not fit perfectly with the characteristics of DC, which may arise and develop independently from the participation of public entities.

Emerson et al. (2012) [52] formulated another definition:

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.

This definition better captures the proper aspects of DC, as it places the role of public, private, non-profit, and community entities on an equal level. Indeed, Ruijer (2021) [20] adapted it to DC as follows:

The processes and structures 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.

Another aspect that emerges from the definition is the nature of the elements that constitute governance, namely processes and structures. These components are recognized also in the definition of Vangen et al. (2014) [58]:

The governance of a collaborative entity entails the design and use of a structure and processes that enable actors to direct, coordinate, and allocate resources for the collaboration as a whole and to account for its activities.

Bryson et al. (2015) [51], in their analysis of the main collaborative governance frameworks, add that a common feature of these frameworks is the fact that governance lies at the intersection of processes and structures. In accordance with the model developed in the mentioned study, not only structures and processes are identified as constituents of governance, but also those elements that lie at the intersection between them. The following paragraphs will explain what is meant by these three categories.

Ultimately, the collaborative governance is not static but is itself a process in which interaction and conflicts between partners generate new rules that determine the collaborative environment [54].

Considering all the elements discussed above, in this research, DC governance will be interpreted as follows:

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.

3.1.1. Collaborative processes

A process is a series of actions that organizations take to achieve a result. In the case of cross-sector social partnerships, collaborative governance processes are about facilitating interaction between partners in order to accomplish the activities to pursue a socially relevant objective. Bryson et al. (2015) [51], recognise as collaborative processes those aimed at trust-building, at inclusive participation, at the development of a common understanding of the problem, at the establishment of a commitment towards goals, and formal and emergent planning.

In addition to these processes, DCs have those related to data usage. These include standardized data collection or acquisition, data sharing and access, and data integration and exchange, but also the co-creation of value and data-driven insights for public problems [20].

3.1.2. Collaborative structures

With structure is intended the way in which the parts of a system are organized, or the system itself arranged in this way. In cross-sector social partnerships structures are the arrangements and resources that characterize the formation and the development of a partnership [57]. Structures include norms and rules that emerge to promote collaborative processes as well as to reach agreement on collaborative goals and

actions. The structures of individual organizations are a topic widely addressed in the literature (e.g., division of labor, procedures, rules, authority relations), whereas collaborative structures are less discussed, even with respect to collaborative processes. Collaborative structures are a relevant element to focus on, as they can cause structural tensions when organizations within collaborations simultaneously maintain differentiated and integrated structures [53].

Considering the peculiar case of DC, are examples of relevant structures the technical infrastructure, the data sharing and use agreements, the presence of a legal dedicated entity for the collaboration.

3.1.3. Intersection of processes and structures

The boundary between collaborative structures and collaborative processes can be very blurred. Structures can be the result of a process, as in the case of the creation of the initial agreement of the partnership, or they can be a fundamental element for the execution of the process, as in the case of a leadership role where the process is strictly related to the definition of the authority. Indeed, some frameworks about collaborative governance does not consider structure as a separated component with respect to process dynamics[52]. Bryson et al. (2015) [51] recognise as at the intersection of the two components the leadership role, the technology, and the collaborative capacity and competences.

In the case of DCs an example of a governance element which is at the intersection of processes and structure is data stewardship. Its definition includes both the creation of defined procedures and responsibilities and the execution of the process.

3.2. Forms of collaborative governance

Based on the review of the literature by Provan et al. (2008) [56] three forms of collaborative governance were identified.

The forms are classified according to two dimensions:

- the presence of an intermediary organization;
- the position of that organisation in relation to the collaboration.

The first dimension refers to the presence of an organization whose role is to facilitate management by limiting direct interaction between other actors in the collaboration. The first form, called *Participant-Governed Network*, envisages the absence of this entity and is the simplest and most common form in collaborations. Decision-making power may be more or less symmetrical, but there is no separate formal administrative entity.

In the presence of an intermediary, a distinction can be made according to the second dimension. The second form is the *Lead Organization-Governed Network*. In this case,

there is a lead organization among the actors taking part in the collaboration that assumes the role of coordinating and administering the activities.

The third form, the *Network Administrative Organization* model, involves the presence of an organisation with the role of intermediary, but it is an entity external to the actors in the collaboration specifically created to govern the collaboration. Usually, this organisation is a governmental entity or a non-profit organisation, depending on the sectorial composition of the collaboration.

All three forms can be identified in the DCs. An example of the first form is the Consumer Data Research Centre project, in which several academic institutions share the same decision-making power; an example of the second form is the Insights for impact Zika project, in which UNICEF took the role of lead organization; an example of the third form is the California Data Collaborative, in which a new dedicated entity was created and participated in by all members of the collaborative.

3.3. Dimensions

One of the objectives of the literature review was to identify what dimensions of governance can be reasonably linked to the long-term sustainability of the Data collaborative. To do this, seven cross-sector social partnership governance models were initially analysed. Subsequently, the detected dimensions were compared with the DC literature.

The starting point of the research was the review carried out by Bryson et al. (2015) [51], as it considers the characteristics of all the major frameworks developed about cross-sector social partnerships [50], [52]–[57]. Table 1 shows the comparison reported in the aforementioned paper highlighting the most important aspects of each framework.

From all the elements analysed, seven dimensions of DC governance were isolated. The dimensions are shown in Tabel 2 and are discussed in detail in the following sections. Those dimensions consider both the characteristics of the collaborative governance and the peculiarities of data processes and structures. Those dimensions are not independent, but often interrelated. For example, elements that determine the initiation of the collaboration also are fundamental to maintain the trust among the partners during the collaboration.

Table 1 Comparison of cross-sector social partnership frameworks from Bryson et a. (2015)

Publication	Bryson, Crosby, and Stone (2006)	Thomson and Perry (2006)	Ansell and Gash (2008)	Agranoff (2007, 2012)	Provan and Kenis (2008)	Emerson, Nabatchi, and Balogh (2011)	Koschmann, Khun, and Pfarrer (2012)
Theory base	Diverse, including organization theory, public administration theory, leadership theory, strategic management theory	Diverse, including organization theory, public administration theory, strategic management theory	Diverse, including organization theory, public administration theory, policy studies, Planning and environmental management studies	Diverse, including organization theory, public administration theory, strategic management theory	Network theory	Diverse, including organization theory, public administration theory, conflict management theory, planning and environmental management studies	Communication theory
Major components	Initial conditions Formal and informal processes • Agreements • Leadership • Legitimacy • Trust • Conflict management • Planning Formal and informal structures • Membership • Structural configurations • Governance structures Contingencies and constraints • Type of collaboration • Power imbalances • Competing institutional logics Outcomes and accountabilities	Antecedent Processes • Governance • Administration • Organizational autonomy • Mutuality • Norms of trust and reciprocity Outcomes	Starting conditions Collaborative processes • Face-to-face dialogue • Trust building • Commitment to process • Shared understanding • Intermediate outcomes • Facilitative leadership Outcomes Plus core contingencies: time, trust, and interdependence	Decision networks versus nondecision networks Processes • Activation • Framing • Mobilizing • Synthesizing	Ideal types of network governance: • Participant governed. • Lead organization • Network administration organization Critical contingencies: • Degree of trust, number of members, goal consensus, need for network-level competences (nature of the task and kinds of external demands) Persistent tensions: • Efficiency versus inclusion • Internal versus external legitimacy • Flexibility versus stability Evolution of these governance systems over time	System context • Drivers Collaborative governance regime Collaboration dynamics • Principled engagement • Capacity for joint action • Shared motivation Actions Impacts Adaptation	Communication practices • Increasing meaningful communication • Managing centripetal and centrifugal forces • Creating a distinct and stable identity Development of authoritative texts Trajectory of authoritative texts Communication practices to assess overall cross-sector partnership value • External intertextual influence Accounts of capital transformation
Emphasis	Cross-sector collaboration, institutional logics, planning, contingencies, power and the importance of remedying power imbalances, the need for alignment across components	Learning, organizational autonomy, leadership, administration	Face-to-face dialogue, incentives and disincentives, the importance of remedying power imbalances	Leadership through a whole range of roles, processes, and structures, public value, capacity building, and learning	Governance structures	Collaborative regimes what makes collaborations work, capacity building Pulling out collaborative actions from overall impact/outcomes	Authoritative texts and their effects on activities and partners

Table 2 Theoretical dimensions and sources

Dimension	Sub-dimension	Cross-sector social partnership publications	Data collaborative publications
Initiation	• Leadership • Interdependence	Bryson, Crosby, and Stone (2006) Thomson and Perry (2006) Ansell and Gash (2008) Agranoff (2007, 2012) Emerson, Nabatchi, and Balogh (2011)	
Trust	• Type of trust • Commitment	Bryson, Crosby, and Stone (2006) Thomson and Perry (2006) Ansell and Gash (2008) Provan and Kenis (2008) Emerson, Nabatchi, and Balogh (2011)	Stalla-Bourdillon et al. (2021) Klievink et al. (2018) Crane (2020)
Formal Structures	• Formal agreements • Roles and responsibilities	Bryson, Crosby, and Stone (2006) Koschmann, Khun, and Pfarrer (2012) Emerson, Nabatchi, and Balogh (2011)	Stalla-Bourdillon et al. (2021) Donge, Bharosa and Janssen (2022)
Intermediation	• Facilitative leadership • Intermediation model	Ansell and Gash (2008) Provan and Kenis (2008) Bryson, Crosby, and Stone (2006) Thomson and Perry (2006)	Gil-Garcia and Sayogo (2016) Perkmann and Schildt (2015)
Incentive system	• Incentive model	Ansell and Gash (2008) Provan and Kenis (2008) Emerson, Nabatchi, and Balogh (2011)	Susha et al. (2019) Susha (2020) Gil Garcia and Sayogo (2016) Perkmann and Schildt (2015) Susha, Janssen, Verhulst (2017)
Business model	• Value proposition • Revenue model		Carballa Smichowski, (2019) Susha (2020) Robin et al. (2016) GSM Association (2018) Barhosa (2022)
Adaptation	• Inclusivity versus efficiency • Stability versus flexibility	Bryson, Crosby, and Stone (2006) Thomson and Perry (2006) Agranoff (2007, 2012) Ansell and Gash (2008) Provan and Kenis (2008) Emerson, Nabatchi, and Balogh (2011) Lockwood (2009)	Stalla-Bourdillon et al. (2021) Page et al (2021)

3.3.1. Initiation

The elements influencing the initiation of cross-sector social partnerships are widely discussed in the literature.

Bryson et al. (2015) [51], distinguish it in *General antecedent conditions* and *Initial conditions*. For the purposes of this research, these categories overlap in the search for success factors for the formation of Data Collaboratives.

This dimension is based on focusing on the environment in which collaborations are embedded and the immediate preconditions that influence the formation of collaborations [53].

In relation to the initiation of the Data Collaborative, two main sub-dimensions have been identified: *Interdependence* and *Leadership*. The former is understood as the motivation that prompted the actors to join the collaboration, and the latter as the agent that involved and facilitated the parties during the formation phase.

The concept of interdependence is based on several elements. Firstly, there is the identification of a pressing social need. Second, there is the realisation that this need cannot be solved by a single agent in the field. This is what Bryson et al. (2006) [53] call sectoral failure, referring to the observed situation where the efforts of a single sector to solve a public problem are insufficient before attempting cross-sector efforts. Third, interdependencies are based on both different and shared interests [54]. This means that there must be alignment both in the common social mission and in the benefits expected from each partner based on its mission. We therefore consider a direct consequence of interdependence to be the shared motivation to join together and achieve the result of collaboration, based on the intertwining of individual interests.

Interdependence in Data Collaboratives is verifiable in the fact that the different resources and competences driving the collaboration are distributed among different actors and sectors, as data, data competences and social implementation competences are never in the hands of a single actor.

The concepts of leadership and collaborative leadership are abundantly described in the literature. Leadership refers to the presence of an identified leader who is able to initiate and help secure resources and support for collaboration [52]. Although it is also a fundamental element during the implementation of the collaboration, in this research it was given particular importance within the initiation dimension, as it helps to identify which actor assumed a leadership position in the formation phase with the aim of systematizing the creation of Data Collaboratives.

Different types of actors and leadership roles can be found in collaboration. The leader may be a member of one of the parties or there may be several leaders, formally and informally defined [50]. The leadership roles can be political, managerial, or technical. Leaders can be sponsors or champions, where the former have the authority and

resources to initiate collaboration but are not employed in day-to-day activities, while the latter are directly involved in collaboration processes [53].

Agranoff (2006) [55] identifies four dimensions of collaboration leadership. Firstly, there is a champion who may be the head of a party who organizes and supports the collaboration; secondly, there is a political core with representatives of the parties; thirdly, there is a technical core with members of think thanks and committees; fourthly, there is the staff who is involved in each stage of operations. This structure is more suited to large partnerships with public bodies, whereas in other cases these dimensions often overlap. Nevertheless, it represents the depth and layering of the concept of leadership within cross-sector partnerships.

3.3.2. Trust

Trust is a critical dimension for the formation, the success, and the long-term sustainability of collaborations [52]–[54], [56]. The presence of it reduces complexity and transaction costs and increase the active participation, while the reduction can cause tensions or even the failure of the entire project.

In the case of Data Collaboratives, the concept of trust is strictly linked to the core data activities, especially when sensitive information is involved, where trusted data structures and processes are necessary [59].

In general, trust can be defined as "the willingness to accept vulnerability based on positive expectations about another's intentions or behaviours".

In Data Collaboratives, this vulnerability may be represented by the risk of others opportunistically using the data, contrary to the interests of the data sharer, endangering the continuation of the collaboration [21].

Scholars have identified three components of trust: ability, benevolence, and integrity [60]. Ability refers to the capacity to perform as expected; benevolence is related to the fact that the organization shows good intentions and seeks to satisfy the needs of stakeholders; integrity is the idea that the organization’s actions are correct, that the organization keeps its promises and operates according to moral principles [60]. Moreover, scholars also identified three levels of trust: information-based trust (rational decisions about whom to trust), identity-based trust (familiarity between stakeholders) and institution-based trust (social norms and structures that define acceptable behaviour) [61].

Inside collaborations, the intensity of trust is not static and not always reciprocal among partners. Collaborations can start with different degrees of trust and trust is built during the collaborative timeframe. Bryson et al. (2015) [51] identify as trust-building activities the sharing of information and knowledge and the demonstration of competence, good intentions, and commitment. Ansell et al. (2008) [50] recognise that building trust is a process that requires time and emphasise the role of

collaborative leadership in doing so. In addition, they suggest that trust should be built not only within the collaboration, but also with external actors who can influence stakeholders, such as past adversaries.

In their framework, Emerson et al. (2012) [52] identify the development of trust as a key element for the shared motivation to achieve the result of the collaboration. Trust generates mutual understanding, which in turn generates legitimacy and ultimately commitment. The commitment generated by trust is a fundamental element in the case of cross-sector partnerships, as it allows overcoming the limitations and cultural differences resulting from the structural diversity of the actors.

Trust is also reflected in the governance model. Provan et al. (2008) [56] argue that the density of trust between the parties influences the formation of the governance structure. Specifically, shared governance is more likely when trust is distributed, while intermediary organisation arises when trust is less or more concentrated in the hands of a single member.

3.3.3. Formal structures

Compared to the previous ones, this dimension has less prominence in the analyzed literature. This dimension refers to the formal rules agreed upon between the partners. Among the dimensions identified, this is the only one that reflects the regulatory function of collaborative governance.

Koschmann et al. (2012) [57], in their framework on the communicative constitution of cross-sector partnerships, emphasize the role of texts. Concrete texts, such as formal agreements, together with figurative texts, contribute to the formation of the authoritative text, which according to them is the key element in achieving collective agency. In particular, the authoritative text makes it possible to influence efforts to attract the necessary resources and voluntary participation of the parties.

During the initial stages of collaboration, formal agreements are useful for building the bond between actors, as they declare interests in participating in the collaboration and solving the social problem, thus sharing the connections and interdependence between actors [53].

Such agreements also make it possible to define the resources that each member must make available to the collaboration and the limitations on the use of these resources [53]. This is particularly relevant when it comes to sharing data and data infrastructures, as this dimension plays a key role in defining the data governance model [59].

Through formal agreements, it is possible to clarify the roles of actors in both operational and decision-making terms [53]. This dimension therefore has a close link with the formal definition of the collaboration's leadership.

Agreements can also influence the flexibility and adaptability of governance by limiting or incentivizing a priori the entry of new actors. Another important function is the sharing and allocation of responsibilities of each member and their accountability. This aspect is important in Data Collaboratives because it allows defining the data stewardship role and the responsibilities of each actor towards it [62].

3.3.4. Intermediation

As mentioned in the section on governance forms of collaborations, in cross-sector social partnerships there is often an organization that facilitates their development. In the literature, this role is referred to in various ways: intermediary, boundary spanning organization, convener, etc. In this research we refer to it by the term intermediary, highlighting its relative position within the collaboration.

This dimension is deeply linked to the start-up phase. The presence of an intermediary organisation can facilitate the formation of the collaboration by drawing attention to a social problem and giving it legitimacy within the parties involved in the problem [53].

The intermediary function is also crucial during the development of collaboration. Ansell et al. (2008) [50] emphasise the importance of facilitative leadership and identify the main competencies of this type of leader. Intermediaries must promote the participation of actors, guarantee control and influence over them, facilitate processes, and extend the scope of the partnership.

Perkmann et al. (2015) [26] identify two mechanisms enabled by intermediary organizations in data partnerships: mediated disclosure and multiple targeting. The first mechanism refers to the data aggregation and anonymization function performed by these organizations. In this way, they can leverage valuable private data while mitigating risks for data providers. This role is only possible if the intermediary enjoys the trust of all parties, and this can be achieved by limiting its mission to a specific social objective. The second mechanism is to consider the different interests of each party and to manage the activities of the collaboration by considering both overlapping and non-overlapping objectives. This is possible because intermediary organizations have a prerogative over the allocation of resources and control of operations.

The task of intermediaries is not only to ensure that collaboration requirements are met, but also to manage and build inter-organizational relationships. Capabilities recognized for an intermediary include the ability to build and sustain effective interpersonal relationships between partners, the ability to act as an honest broker between contested power structures and the ability to manage interdependence between actors [54].

Susha et al. (2023) [49] explored the role of intermediaries in enhancing the sustainability of data collaboratives. Specifically, they identified five functions performed by these organisations in the creation of data ecosystems: i) the *connecting*

role consists of engaging relevant actors in the ecosystem and matching supply and demand within the collaborations; ii) the *mediating role* consists of mitigating conflicts of interest, building trust, managing expectations, creating a shared vision and providing governance structures and processes; iii) the *stimulating role* is to stimulate innovation processes, mobilize the necessary support, funding and organizational resources and ensure openness and transparency; iv) the *learning catalyst role* is to facilitate the creation and dissemination of shared knowledge and skills; v) *the infrastructure provision role* is to provide an initial set of resources, leverage resources and capabilities across the network and make the creation of the collaborative outcome more efficient.

3.3.5. Incentive system

Despite this dimension not being present as a main point in cross-sector social partnership governance frameworks, the existence of an incentive system for sharing resources and especially data is key in data collaboratives. In the analysis of critical success factors carried out by Susha (2020) [8], the incentive system appears as a prominent element both from literature research and from empirical research, where it is recognized by experts as the second most important element after data quality. This gives further strength to the critic of Susha et al. (2019) [37] of the adoption of the framework by Emerson et al. (2012) [52] for Data Collaboratives. In fact, they recognize incentive systems for data providers to donate commercial data for a social purpose as an important element for the governance of this type of collaboration.

The alignment of incentives for data providers with the DC's objectives is considered one of the main coordination issues by Susha et al. (2017) [63]. They identify three possible coordination mechanisms: coordination by feedback, transfer of knowledge between parties and third-party coordination. In particular, in industry-academia partnerships, companies are motivated to engage for the following reasons: i) support for product development; ii) access to public funding for research; iii) solutions to technological problems; iv) pursuit of research opportunities. More in general, incentives for companies to share data include the 5Rs: reciprocity; research, recruitment and insights; reputation and public relations; and revenue enhancement. Even if there is no immediate financial gain, private sector data providers need to clearly understand the benefits of their participation in DCs [37].

For collaborations mainly involving public sector organizations incentivization for data sharing is less problematic. In these cases, the incentives for data exchange are linked to the idea of the 'common good', but also to legal requirements and political will [37].

Finally, when it comes to incentives, these should be able to stimulate all actors. For example, Perkmann et al. (2015) [26] emphasize the problem of encouraging

individuals within academic institutions, where the monetary incentive is not the main driver compared to the standing in their academic community.

3.3.6. Business model

One of the problems of these types of collaborations is their one-off nature. This aspect is linked to a lack of financial independence [17], [18] and this is why this dimension has been emphasised with respect to the aim of systematising long-lasting Data Collaboratives. Although this dimension has little relevance in the literature on the governance of cross-sector partnerships, it has resonance in the literature on data-driven social initiatives. Susha (2020) [8] identified this aspect as one of the critical success factors among the organisational elements of DCs. Robin et al. (2016) [34] recognized the business model as one of the four main challenges of data collaborations. In particular, they found that the business model of this type of initiative can be negatively affected by the uncertainty of supply and demand for alternative data sources. Demand uncertainty is related to the fact that data users may not be able to effectively leverage on this type of data, as they usually did not do so in the past, thus raising doubts about the extent of the market. Supply uncertainty has to do with the benefits of data sharing, which might not be immediate or straightforward, and even if they are short-term, they might generate unproductive costs in the long run. Carballa Smichowski (2019) [64] in his analysis of alternative data governance models identified the business model as a challenge for the sustainability of the initiatives. This is particularly relevant in the case of crowdsourced data commons, as all of them rely on pro bono work and research grants or donations and most are struggling to ensure financial stability over time.

Another aspect to be considered in the choice of business model is to consider not only sustainability but also responsibility for the consequences of such models. Bharosa (2022) [65] acknowledges that among the outputs of governance design, with particular reference to the use of new technologies in collaboration with the public sector, there must be the co-development of responsible business models as there is a lack of precise guidelines on the management of such solutions given the degree of innovativeness. Business models must consider characteristics of the actors involved, privacy and data protection regulations, and ethical concerns. For instance, business models offering free services in exchange for the resale of personal data or based on advertisements may be problematic.

Furthermore, GSM Association (2018) [17] identifies five main types of business models for initiatives using mobile big data for social purposes: i) philanthropy, ii) donor-subsidized, iii) cross-subsidized, iv) tiered, v) commercial. The models are differentiated according to the level of resources coming from donations and the level of revenues generated by the commercial activity. The first model involves the pro bono donation of data and processing expertise by the data owner. The second model envisages a third-party entity acting as a donor of funds to cover the costs of the data

and processing. The third model involves the data owner subsidizing the project with part of the revenues associated with other services provided with the same data. The fourth model considers multiple levels of service, with basic services offered free of charge and more onerous services paid for. The fifth model assumes that all costs are borne by the revenues generated by social services and is considered the best chance for long-term sustainability.

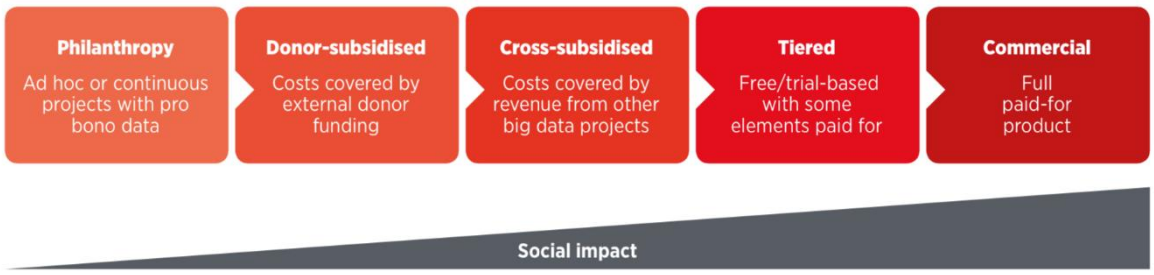


Figure 4 Business models for big data for social good activities from GSM Association (2018)

3.3.7. Adaptation

This dimension refers to the ability of partnerships to mutate in response to the need for change and to mitigate internal and external tensions. In the literature on the governance of cross-sector social partnerships, there is consensus about tensions of structural origin generated by the composition and the objective of the collaboration.

A first category relates to conflicts over the goals and strategies to be implemented within the collaboration given the difference in interests among the actors involved. In these cases, the mission of the collaboration can influence the levels of conflict [53], but more importantly, the actors must leverage the perception of mutuality as a win-win problem solving technique [54]. These conflicts are related to another source of structural tension which is about the trade-off between autonomy and interdependence between the actors. During the duration of initiatives, individual partners maintain their own distinct governance entities that overlap with the collaboration. This creates a tension between individual and collective interest. Since collaboration is voluntary, partners must justify their involvement in terms of contributing to their own objectives. When objectives conflict if the specific problem is not sufficiently urgent for all partners, individual missions are likely to prevail over collaborative missions [54].

Another category of tensions arises from power differences between actors. This is the most discussed source of conflict in literature. Different actors may occupy different role positions and have different weights, creating unequal opportunity contexts and filling structural holes, while others may be less willing or able actors [55]. This can be

exacerbated when they differ in status (in size, funding, or reputation) [53]. Ansell et al. (2008) [50], state that these kinds of tensions surface when important actors do not have the organizational infrastructure to be represented in collaborative governance processes. The authors acknowledge that this problem emerges more in cases where the level of actor diffusion is high and the problem horizon wide. They identified three factors that diminish the power of stakeholders: i) lack of organized groups to collectively represent individual stakeholders; ii) lack of expertise to engage in discussions on highly technical issues; and iii) lack of time, energy, or freedom to engage in time-consuming collaborative processes. This category of tension applies to different areas, depending on the life stages of the collaboration. Initially, when actors try to agree on the nature of the problem to be addressed, issues are likely to revolve around convening and inclusion; then, when they discuss the direction to be taken to address the problem, issues revolve around setting the agenda for collaboration and sharing relevant information; finally, once implementation has begun, issues of power revolve around the exercise of influence, authorization of action and control of resources [53].

A structural tension recognised by Provan et al (2008) [56] is related to the level of flexibility in spite of stability. On the one hand, flexibility makes it possible to better adapt to competition, environmental threats, and opportunities, providing for leaner procedures and giving the possibility to easily involve new partners; on the other hand, stability becomes important in the efficiency of continuous processes of long-term projects and to maintain legitimacy and relationships within and outside the collaboration. Both characteristics confer advantages over the sustainability of the collaboration, and it is important to be able to mediate between the two extremes. In order to be able to do this, a formal hierarchy must be introduced accompanied by frequent reassessment of structural mechanisms and procedures in the light of new developments.

Another governance tension discussed by Provan et al. (2008) [56] is caused by inclusiveness in collaborative decision making and administrative efficiency. If on one side inclusiveness is necessary for legitimacy and effectiveness of the collaboration, on the other "the more that organizational participants are involved in the network decision process, the more time consuming and resource intensive that process will tend to be".

Regarding these two sources of tension Stalla-Bourdillon et al. (2021) [59], in their framework for the creation of a data governance system within the Data Collaborative, emphasise the importance of establishing an inclusive decision-making body and a flexible membership process to join the collaboration.

These governance compromises agree with the interpretation of Bryson et al. (2015) [51], which argue the importance of structural ambidexterity inside cross-sector collaborations. Ambidexterity allows effective tension management and can be

achieved through the separation of elements of tensions in time or in space. This concept is also deepened in the work of Page et al. (2021) [66], which defined and demonstrated the collaborative ambidexterity, that consist in linking knowledge exploitation and exploration to create innovative solutions to social problems.

In addition to structural stresses, it is possible for a partnership to experience shocks due to exogenous or endogenous factors that undermine its continuity. These events may depend on various factors, such as a change in cash flow, a change in client demographics, involvement in scandals by one or more members, changes in political direction, and the exit or entry of members. Bryson et al. (2015) [51] suggest that strategic planning and scenario development can help partnerships anticipate future developments in order to manage them effectively.

4 Methodology

An interpretive approach was used to answer the research question. In particular, a multiple case study was conducted. This is a qualitative research methodology, a type of research that produces results not obtained by statistical procedures or other means of quantification [67]. A case study is "an empirical inquiry that investigates a contemporary phenomenon within its real-life context, especially when the boundaries between phenomenon and context are not clearly evident" [68]. Since this is a rather recent phenomenon and only a few empirical cases are available, an explanatory research approach was adopted in order to search for the connecting elements between theory and evidence from real cases. The choice of the multiple case study over the single case study is due to the fact that multiple cases provide greater robustness to the results. Another aspect is due to endogenous variability of DCs [15]. Wanting to identify factors that have general validity for any type of DC, it was important to replicate the study on a significant number of heterogeneous cases.

4.1. Problem and objective

Data collaboratives represent one of the frontiers of social innovation. Part of these collaborations are of a one-off nature, thus with limited duration and impact generated [17], [37] and for this reason there is a recognized need to create methods and tools for the systematization of them [69]. Examples of such cases are DCs organized for a competition or set up in response to emergency events such as epidemics or natural disasters, which terminate their activities once the competition is over or the urgency of the problem has diminished. Being a phenomenon that has been studied for less

than a decade, literature is still scarce and indications for building sustainable models do not abound. This study assumes that the problems that undermine the continuation of DCs are not just of technical nature in data processes, but can rely on governance components of collaborations [17]–[19], [37]. Directly observing cases in which the condition of sustainability is manifested can significantly enrich the literature on DC governance.

The aim of this study is to identify relevant governance factors in order to systematize, accelerate and facilitate the creation and the preservation of DCs in the long run, thus generating a continuous positive impact for society.

4.2. Frame and unit of analysis

The analysis was conducted according to an abductive reasoning, in which theory defines the initial framework for observing data and identifying patterns. For this research, the theory used refers to two strands of literature: Data Collaborative and governance of cross-sector social partnerships. Both are described in the previous chapters. This theoretical framework is reflected in the dimensions used to identify the key elements of governance.

The units of analysis of the DCs analyzed are the elements of governance. Specifically, the elements of governance were initially divided according to the theoretical dimensions identified in the first literature review, i.e., initiation, trust, intermediation, formal structures, incentive system, business model, adaptation. The study was designed following a multiple holistic case model, according to Yin's classification [68]. Eleven DC cases were analyzed.

4.3. Cases selection

The cases were selected starting from the DC database built to support a previous study by Bartolomucci et al. (2022) [48]. In particular, an initial selection was made considering the classification produced in the aforementioned study. This classification considers five clusters of DCs: i) collaborative effort to support wide-scale research projects, ii) prompt response to emergencies, iii) continuous effort to improve structural responses, iv) data-driven initiatives to support innovative studies, v) external responses to structural problems. The aim of the research is to find the governance elements for the long-term sustainability of DCs with high social impact. For this reason, clusters ii) and iv) were excluded a priori, as the former includes projects that are structurally limited in continuity and the latter does not consider strictly social initiatives. Considering that the initial database included 135 DCs, the exclusion of these clusters led to a reduction to 80. From the set of 80 DCs, a sampling of 30 cases was made, seeking to observe as much heterogeneity as possible with respect to the following dimensions: DC cluster according to Bartolomucci et al. (2022)

[48], DC type according to Verhulst et al. (2019) [14], presence of a clear profitability model, presence of a clear social purpose, macro social sector. The heterogeneity of the sample is shown in Table 3.

Table 3 Heterogeneity of DC sample

DC type	
Data Pooling	5
Intelligence Generation	5
Public Interfaces	5
Research and Analysis Partnerships	10
Trusted Intermediary	5
Clear profitability model	
1	12
0	18
Clearly social good	
1	22
0	8
Macro social sector	
Cities' smart management	7
Education	1
Environment	4
Gender equality	1
Health	7
Humanitarian	2
Multiple social purposes	3
Poverty & Underserved Economies Empowerment	5
DC cluster	
collaborative effort to support wide-scale research projects	7
continuous effort to improve structural responses	16
external responses to structural problems	7

Via e-mail or social media contacts of the exponents, all cases in the sample were approached. Only nine cases were available for interviews, but the proportion with respect to the mentioned dimensions was maintained, making the analysis representative of the sample. More information on the final sample can be found in Chapter 5.

4.4. Data collection

The exploratory case studies were studied by conducting interviews with experts who participated in the DCs and analyzing documents. The theoretical governance dimensions were used to develop a semi-structured interview protocol. Semi-structured interviews are suitable for exploratory studies, as they allow for dialogue with the respondents during the interviews [70]. This approach allows interviewees to reflect on their experiences, bringing out new insights [71]. The interviewees were selected on the basis that they worked directly in a DC and in most cases were the director of one of the organizations involved. 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. A total of nine semi-structured interviews were conducted. In two of the eleven cases, it was not possible to conduct the interviews, but as these were cases that had large resonance, it was possible to find more documentary material than the other nine. The data collected from the interviews were triangulated with data from the documentary analysis. Publicly available documentation on each of the cases was searched. The documentation included scientific documents, official reports, content from actors' websites and newspaper articles.

4.5. Data analysis and interpretation

The cases were analyzed through the inductive coding of the text of the interviews and the documentary material. The interviews were transcribed and organized together with the text extracted from the documents, subdividing the contents by theoretical dimension. The contents were reported in a synoptic table to facilitate cross-reading, as in template shown by Table 4.

Table 4 Template of the synoptic table

Case / Dimension	D1	D2	D3	D4	D5	D6	D7
C1							
C2							
...							
C11							

To analyze the interview data, the initial level of analysis was from the individual answers to single questions. Each question is associated with a theoretical dimension of governance identified. Then, among the answers from the same respondent, the response items to a question that referred to aspects of other dimensions were rearranged according to conceptual consistency. The text was coded using an open code that is mainly based on the content of the transcripts. This means that, initially, codes were identified inductively, in line with the exploratory nature of this case. The codes were first individually identified by the author and the co-advisor. Subsequently, the analyses were compared and the codes on which there was agreement were selected. Through an iterative process, the codes were grouped according to epistemic similarity [72]. These groups were then combined to create categories referring to elements of governance, following the methodology introduced by Hatch (2002) [72] and Gioia et al. (2010) [73]. The process was iterative and abductive, as the logical scheme of the categories was reviewed and checked against the theory several times before arriving at the final framework.

5 Results

As described in the previous chapter, the research was first carried out by identifying case studies that had the desired characteristics of social impact and long-term sustainability; then semi-structured interviews were conducted based on the governance dimensions deductively identified in the literature; finally, concepts expressed by the interviewees as critical success factors were inductively extrapolated and summarized. This chapter presents the cases analyzed and reports critical factors according to the related governance dimensions identified in the literature.

5.1. Data Collaborative cases analyzed

Eleven case studies of Data Collaboratives have been identified. Nine of these are cases that we define as 'ongoing' indeed they have the organizational characteristics of long-term sustainability and reached their initial objective. For each of these cases, it was possible to directly interview a contact person, in most cases the director, who testified about his or her experience. Confirmations and additions to the oral testimony were obtained through a search in scientific and official literature. The other two cases, referred to as 'terminated', are projects that did not achieve full implementation and were therefore ceased prematurely. These cases provided useful information in conducting an analysis of failure factors, i.e., the missing factors that led to the failure were identified as critical factors. For these, it was not possible to collect direct testimonies through interviews with the participants, but given the relevance of the projects, a significant number of secondary sources were available, i.e. papers, official reports, contracts, newspaper articles, etc. Details about the cases are given in Table 5.

The macro sectors most represented by the cases analyzed are economic development and infrastructure, but health, education, social inclusion, and public safety are also present, providing enough heterogeneity in the impact areas.

Regarding geographical location, more than half of the cases operate in Europe and Central Asia. This number can be explained by the high attention in Europe to the issue of data use and protection, as demonstrated by EU legislation, but also by the higher ease of access to direct contacts due to the geographical proximity to the research location. The remainder of the cases are projects from North America, where there is a higher concentration of investment in digital technologies. One case involved partners and staff from several parts of the world: North America, Asia, and Oceania.

Also visible from the table is the cluster associated with each case according to the classification of Bartolomucci et al. (2022) [48]. As reported in Chapter 4, those that belonged to Clusters 2 and 4 were excluded from the sample in which the cases were extracted. Among the cases analyzed, it can be seen that there is a strong prevalence of cases belonging to Cluster 3. This is because the third cluster is the most heterogeneous and numerous, and although an attempt was also made to include cases belonging to Clusters 1 and 5, the group of cases analyzed is representative of the numerosity of the clusters in the source database.

Another characteristic observed in the cases is the type of data used, in accordance with the categorization applied by GovLab in its dataset. This attribute has a high level of heterogeneity between cases. All possible types of data are used in the cases and there is almost a balance in the distinction between observed and disclosed data and between personal and non-personal data. There is only a slight preponderance of disclosed non-personal data and observed personal data over the other two types.

The sector most involved is the private sector (eight cases), reflecting the great importance of non-traditional private data for solving social problems. The presence of the public and non-profit sectors is also significant (seven and five cases respectively). The contribution of civil society appears in only two cases, in fact in some DC classifications it is not even considered, as in that of Susha et al. (2019) [12].

Finally, a further distinction was made regarding the purpose for which the data were initially collected, based on the classification of Susha et al. (2017) [15]. The least recurring data type is primary data, while secondary and primary data share the rest, with the former being slightly predominant. This is in accordance with the concepts of non-traditional data and data reuse peculiar to DCs, as highlighted in the literature [12], [14], [15], [59], [74]–[76].

Where possible, useful documentation was identified to support the information gained from the interviews. In the cases interviewed, reference was made to more authoritative sources, such as official reports and scientific articles. For the cases analyzed using only secondary sources, newspaper articles and websites were also analyzed in addition to the sources mentioned for the other cases. Table 6 shows the secondary sources used associated with each case.

Table 5 Cases analyzed

ID	DC name	Continuity	Main sector	Region	DC cluster	Data used	Data purpose	Acrtors involved
C1	Estonia: 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	Civity	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	Discolsed non personal data Observed personal data Observed non personal data	Tertiary	Private Public
C7	T1D Index	Ongoing	Health	Worldwide	1. collaborative effort to support wide-scale research projects	Disclosed personal data	Secondary	Private No profit
C8	Needsmap	Ongoing	Social Inclusion	Europe and Central Asia	5. external responses to structural problems	Disclosed personal data Disclosed non 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	Sidewalk Toronto	Terminated	Infrastructure	North America	3. continuous effort to improve structural responses	Observed non personal data	Secondary	Private Public
C11	InBloom	Terminated	Education	North America	3. continuous effort to improve structural responses	Disclosed personal data Observed personal data	Secondary	Private Public

Table 6 Secondary sources by case

Case	Secondary sources
C1	Robin et al. (2016) [34], Tiru et al. (2010) [77] Ahas et al. (2011) [78], Ahas et al. (2007) [79] Esko et al. (2015) [80], Saluveer et al. (2020) [81]
C3	Susha et al. (2020) [75], FIWARE Foundation (2019) [82], Civity (2022) [83]–[85]
C4	Calzada (2021) [86], Farmer et al. (2023) [39], SalusCoop et al. (2018) [87]
C6	Vij (2016) [88]
C8	Klimczuk (2021) [89]
C10	Dawson (2018) [90], Wyilie (2018) [91], [92], Waterfront Toronto (2017) [93], Barth (2018) [94], Sidewalk Toronto (2018) [95] [96], Doctoroff (2016) [97], Carter et al. (2020) [98], Sidewalks Labs (2017) [99], Donovan (2020) [100], Johnson (2019) [101]
C11	Lingard (2019) [102], Levine et al. (2014) [103], Bulger et al. (2017) [104]

5.2. Governance Critical Factors for DC sustainability

The aim of the research was to identify in real cases critical factors for the creation of impactful and long-term DCs. Table 7 shows all twenty factors identified in the cases associated with the governance dimensions derived from the literature.

Table 7 Matrix of critical factors and theoretical dimensions

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

The following paragraphs illustrate the factors ordered according to the seven dimensions of governance identified in the literature, i.e. initiation, trust, formal structures, intermediation, incentive system, business model, adaptation. Each paragraph presents the factors by also considering their nature with respect to the DC governance definition used in the study. This means that within each governance dimension, the factors are described as elements of process, structure or at the intersection of process and structure.

5.2.1. Initiation: critical factors

5.2.1.1. Initiation process critical factors

Pressure and facilitation by public institutions

Among the governance processes that influence the successful initiation of DCs, it was found in three cases that the activities of public bodies have a significant impact. The public sector plays an important role in the choice of organizations participating in a DC. This can be explained by the fact that the public sector has significant resources for social projects and because it has the regulatory power to influence the behavior of organizations.

D1C6 *"We started out in 2014 with a large tranche of funding from [national council name]. They gave us the money to enable us to build an infrastructure. When I say infrastructure I mean data infrastructure, so secure data warehousing plus all of the things that you see now on the website and a team for professional services, so people who deal with contracts, people who deal with relationship management, people who deal with everything else that you need to do for a data center."*

D1C2 *"There was also a lot of pressure on water suppliers from the state to save water to report data on how much water they're using. There was a combination of this big driving force from the drought and then the state responding to the drought on local water suppliers."*

This factor is also influenced by the so-called "sector failure" of public sector, recognized in literature.

D1C1 *"If we talk about this [partner name] use case, it started because they had a problem. The problem was that in 2008, there was a recession. There were a huge budget cuts in Estonia, in public sphere and one of the places where this budget cut was hit was the Statistics Office Estonia, who used to conduct border service in order to get the tourist numbers on the borders. They decided that they will discontinue this survey, meaning that the [partner name] who used to rely on that data in order to calculate their tourist expenditure numbers just didn't have any more data source. "*

Among the cases interviewed, there was no prevalence of a single type of public body, but bodies of different types emerged: a national bank, a local government, a national council.

5.2.1.2. Initiation structure critical factors

Innovation, technology, and data infrastructure

Among the structures that determine the initiation of DCs, it can be seen that the presence of innovative and safe technology can positively influence this dimension. Organizations are attracted by the use of better performing technologies owned by partners. The demonstration of the effectiveness of new technologies, as an objective and measurable element, contributes to increasing the initial trust in the collaboration.

D1C8 *"The main point is using the [platform name]. If we don't have this kind of platform or this kind of organization, it's not possible to work with this kind of organizations."*

5.2.1.3. Initiation "in-between" critical factors

Presence of a clear need, interdependence

Among the elements that lie at the intersection of processes and structures and that significantly impact the initiation of a DC is the presence of a clear need on the part of one or more actors. The need may be of a different nature. It can be of an exogenous origin, such as an economic or environmental crisis, or of an endogenous nature, such as ineffective communication of information by an actor.

D1C1*"They had a need; we had a solution. I must say that even with other kind of projects that we've done and collaborations, the leading factor for them is always that there has to be some sort of pain that needs to be solved."*

D1C2 *'There was also a drought at the time. So, there was a lot of pressure on water suppliers."*

D1C7 *"The collaboration actually came out of a strategic review that the [partner name] had been doing. So, they've been trying to understand what the gaps were at the global level in understanding the impact of type one diabetes because even in 2020 people didn't have any idea. We couldn't tell you basic things like: How many people live with type one diabetes around the world? Where do they live? What kind of care do they get? So really basic questions about the condition. There was no research that had pulled that together."*

Another aspect noted that is answered in the literature is interdependence. This has been noted as the need on the part of an organization for the presence of another organization in order to accomplish its goal.

D1C5 IT *"Noi da soli non possiamo fare questo: abbiamo solo le competenze e le conoscenze scientifiche rispetto ad aspetti fondamentali. Quindi si trattava di capire che tipo di partnership si cercare di individuare per poter fare il progetto al meglio."*

D1C5 ENG *"We by ourselves cannot do this; we only have the expertise and scientific knowledge with respect to fundamental aspects. So, it was a matter of figuring out what kind of partnership you try to identify in order to do the project the best way possible."*

Relationship network, Reputation and previous Experience, Individual Leadership

The element on which the greatest consensus was found in this dimension is leadership and its relational and reputational aspects. The presence of pre-existing relationships reduces initial trust-building effort and so compresses start-up timeframe. Among the relational links that emerged from the testimonies, in addition to family and friends relationships, are those created in universities and from other work relationships. The latter highlights the importance of the presence of competence clusters in the creation of DC.

D1C2 *"When I was at [university name] doing my Master's program, a friend of mine there, named [friend name], has his family in the water industry. His brother worked at a water district, and he was able to get an introduction to his brother's boss, who was the general manager of a water district."*

D1C9 *"Then some of his friends and then other people were interested in this tracker. So, it just organically grew. Then in the way it works, I feel like in that Silicon Valley bubble, you talk to your friend, you get other people involved and you get the one guy who knows how to do the data, you get the other one, who has some of the connections."*

Links between actors are often facilitated by their reputation and previous experiences. Organizations that have participated in social or data-driven collaborations in the past are more likely to do so again.

D1C5 IT: "*La [nome del partner] da diverso tempo stava investendo sull'ambito dei dati, licenza dei dati. Poi al tempo stesso abbiamo portato avanti anche collaborazioni con soggetti che si occupavano di imprenditoria d'impatto: [nomi di altri partner]. Poi abbiamo anche realizzato le cinque edizioni di un programma destinato agli imprenditori sociali del Mediterraneo. Quindi c'era questo background."*

D1C5 ENG: "*The [partner name] had been investing in data, licensing data, for some time. Then at the same time, we also pursued collaborations with entities that dealt with impact entrepreneurship: [names of other partners]. Then we also did the five editions of a program for social entrepreneurs in the Mediterranean. So, there was this background."*

D1C7 *"So, one of the first things we wanted to do was understand how many people actually live with type one around the world and the collaboration came through because the [partner name] had already been doing some work on the [program name]."*

Reputation accelerates the creation of trust in the leadership of collaboration for actors who are new to this type of initiative. Having the opportunity to demonstrate that an organization has generated great value in the past with similar projects may lead new actors to join the collaboration.

D1C8 *"If the people that work in the organizations know us, our works, and our effort, they can trust us. We are sharing with the people and organizations our projects, our efforts in the field and we are saying that we create this impact."*

Another aspect that emerged related to the initial leadership is the role of individuals. The start-up phase of the collaboration is often led by a single person or a few

individuals. From the interviews, the initial intention as a top-down strategy of an organization emerges less frequently than the bottom-up approach of individuals.

D1C9 *"One of them happened to be studying, going back to [school name] to study pulmonary disease and happened to be looking at COVID and his work and was risingly noticed. So, he actually just built up a kind of quick, like Excel type, dashboard, just to track cases in San Francisco because that's where we saw them in the first US cases. Then some of his friends and then other people were interested in this tracker."*

Lean development

The last factor that emerged from the cases analyzed is the importance of a lean development of the initiatives. DCs should start with a minimum viable product, test and develop over time. This aspect has been detected in two interviews but emerges particularly in the material related to a failure case, where the absence of it is considered as one of the main elements that determined the unsuccess.

D1C3 *"So we made it as a showcase, but the showcase was so successful that more municipalities came to us, and they wanted to use this application."*

D1C4 *"Our first attempt of a Minimal Viable product was starting in 2021, the second one, the version we have right now is from 2022."*

D1C11 *"[project name] has to scale fast. It's a big, heavy, large-scale project, with a large initial investment. It builds on the promises and commitments made by [project name] predecessor, the [predecessor project name], to early adopting states such as New York, guaranteeing them very high security and profound changes in data management in a compressed timeframe. The belief is that it can't be designed small, so we have to go big. A big visible launch is planned to solidify the commitments previously made and send messages to the industry players that things are changing and exciting times are ahead."*

5.2.2. Trust: critical factors

5.2.2.1. Trust process critical factors

Pressure and facilitation by a public institution

The presence of a public institution was found to be a relevant factor for trust building in two cases. If there is the participation of a public institution or if it acknowledges and validates the collaboration, it becomes easier for actors to trust the initiative. This was mainly found in the initial phases, where actors have little knowledge of each other and therefore sponsorship by institutions acts as a kind of certification.

D2C8 *"In Turkey, they don't understand what social cooperative is, what social enterprise is. So we are saying that we are officially registered in the Ministry of Trade, so they can trust us. [...] To overcome this problem we work with the Chamber of Commerce in the provinces and with municipalities. We signed an agreement between [company name] and Chamber of*

Commerce, so when we visit local SMEs, we say that we are working with Chamber of Commerce, we share the letter and they trust us."

D2C1 *"There are things that help you to build trust to things like you've done it before. The university might have helped us. "*

Vision and mission alignment

Actors have different interests and cultures that can lead to divergences in the realization of the social mission. On one hand, in a negative sense, lack of mission and vision alignment can diminish trust and lead to partnership failure.

D2C2 "*I mentioned some of the agencies that departed early. The fundamental underlying issue was a difference in vision and in particular there was some disagreement around our role of policy analysis."*

On the other hand, from a positive point of view, an alignment on vision and social mission can lead to an increase in initial trust in the collaboration.

D2C4 "*People donate data for something, and we need to build these research projects. To build trust, when we have to go to street and try to convince people we obviously think that explaining the vision and the mission of the project is very very important."*

D2C8 "*The third one [important element to build trust] is the mission, the fact that we are working for the community social benefit."*

Privacy and regulation compliance

Respect for rules is clearly a fundamental aspect of trust. Where this is lacking, trust in collaboration collapses.

D2C1 "*You have to show how it all is in accordance with legislation, basically you have to prove their lawyers that everything is according to the local laws".*

As far as DCs are concerned, this factor is strongly linked to data management. Compliance with the constraints imposed by local legislation is a key element for the continuation of collaboration. Compliance with privacy becomes paramount when personal data is involved. To build legitimacy, collaborations need to be capable not only of defining secure data processes but also of communicating them effectively with stakeholders, given the relevance of the topic in the public debate.

D2C4 "*To build trust, when we have to go to street and try to convince people we obviously think that explaining the vision and the mission of the project is very very important. Explain that one of our values is keeping privacy at the very high level, because we know that this is a very sensitive thing for citizens."*

Compliance can also be an opportunity to reduce trust-building efforts when it concerns obtaining certifications on the security and quality of data processes.

D2C3 "*We also comply with highest ISO standards and with the GDPR standards, which are really difficult sometimes. So, we have all the certificates to manipulate the data."*

Partners and stakeholders' engagement

Another aspect noted is that of the engagement of the parties in the collaboration. Effectively involving partners and stakeholders in processes helps to increase trust and legitimacy of the DC. This aspect is particularly evident in the failure cases analyzed. Among the elements that have been recognized in both as contributing to failure is the failure to involve stakeholders, in particular the beneficiaries of services.

D2C11 "*A key takeaway from this study is the critical importance of achieving buy-in with the population served."*

In case C11, a form of beneficiary involvement in the design of innovative services was completely absent.

D2C11 "*The original [founded program name] model here in New York was created without one ounce of input from the public-school districts so if you're going to build whatever the initiative is, you have to build trust."*

In the C10 case, however, there was a system of stakeholder and community involvement in the co-design of services. but with substantial deficiencies. There is evidence of workshops that were organized with the original purpose of gathering information but turned out to pursue other communication purposes.

D2C10 *"Waterfront Toronto and Sidewalk Labs will devote the next year to extensive community and stakeholder consultation and long-range planning.*

[…]

Over the last year, Sidewalk Toronto has hosted a series of public consultations focused on urban life staples, from mobility to housing.

[…]

On May 3, [media company name] attended a town hall-style meeting for the project, one of many public relations events that have been held since October, at which residents were invited to give input. In previous meetings, the presentation had focused on cutting-edge technology but now seemed more about convincing attendees that Toronto would not become a "dystopian technocapitalist hellscape," as [newspaper name] columnist [journalist name] recently described big tech's pivot to city-building. A tone of defensiveness, if not annoyance, pervaded the question-and-answer session that followed, during which attendees raised thorny questions about how their data would be used."

A tool that can help in the management of party engagement is what was defined by C6 case respondent as a 'ladder of engagement'. According to this model, the level of engagement must be commensurate with the initial level of knowledge and trust. The intensity of engagement in the party's processes increases with increasing trust based on previous experience. This generates a virtuous circle in which trust and engagement gradually evolve.

D2C6 *"We've also relied very heavily on a process that we call the ladder of engagement. So, if you can imagine a ladder that has various rungs that you climb up to the top, and if you start at the bottom, you might think of something like a bilateral arrangement. So, you go to a data provider or whoever, and you say "we'd really like to do something with you" and you sign up to do something which is kind of bilateral. One to one and that works. That can work really well. But then the next rung of the ladder might be that you say "well, actually, we'd like to do something a bit more collaborative and engaging with other people or we might like to joint sponsor a PhD student", and it effectively works up this ladder of engagement. You have the top tiers where it's partners depositing data within our repository to be made available. So that's quite a big thing, gaining the trust to get data into the repository to be shared. We've always relied on this kind of ladder to demonstrate that we're engaging more and more with partners."*

Case C7 presents a way to achieve partner trust through a participation to a joint pilot project. The involvement of all actors in the pilot project, in this case paper writing, increased mutual trust before undertaking the next phase of collaboration.

D2C7 *"Obviously one of the main things was the paper and obviously everyone had varying amounts of time that they could contribute to it. So, [partner name] took the lead in writing the paper, but everybody else was also authors on the paper. So, we ensured that no one was left behind even if they didn't have the capacity to do as much work on the paper. [...] So it was important for us to keep all the partners engaged."*

Social responsibility and impact measurement

Among the cases analyzed, a positive effect of the organizations' social purpose on the creation of trust was found. It is easier for organizations that have produced value in the social sphere to generate trust in other actors.

D2C6 *"There's definitely something around shared benefit, so it was that we demonstrated there was value in what we were doing, so everybody would be able to benefit whether that be from better insight into behaviors within the data sets or being visible in terms of corporate social responsibility or whatever else."*

This factor can become more relevant when combined with social value measurement tools in order to provide an objective assessment of the impact on trust-building.

D2C8 *"The first touch generally is with the people not the organizations. If the people that work in the organizations know us, our works, and our effort, they can trust us. But if they don't know us, they want to see some kind of information. We are sharing with the people and organizations our projects, our efforts in the field and we are saying that we create this impact."*

Communication and transparency

Communication plays a key role in the trust-building process. In a DC, the activities carried out, the allocation of resources, the use of data, etc. must be communicated in a transparent manner.

D2C1 *"The concern of the statistical authority is whether the definitions have been followed. The client is concerned about if we are able to deliver on time, if we could actually put together the the data collaborative and working in it etc. It is very subjective. It's between people and I think it's always been like that. So the technology doesn't solve here anything, it's all communication."*

D2C2 *"A lot of it is communication that goes on. We try to be very communicative about what we're working on at any given time. In terms of trust, we're very transparent about our financials, how we're spending money, what projects we're working on. There were some disagreements in the past. At the time maybe we weren't communicating properly around the actions that we were taking, but essentially few agencies decided that can't really support this anymore."*

On the one hand, communication is important to build trust between the members within the collaboration, on the other hand, communication to the outside world is crucial to foster legitimacy towards beneficiaries and stakeholders.

D2C8 *"We always try to be open in anything, in any project. [company name] is a social cooperative, which is an actor of social entrepreneurship. We should be transparent with our partners, including volunteers, about organizational, administrative, and financial information. [company name] is not just the employees, the management team, or the founders, we are also the partners, the volunteers, and the whole community, because in the long term we want to create a solidarity among the supporters, the people in need and the organizations. We are working for all the community. [company name] will not work until it will be operational in the long term. The first thing is being transparent and open with partners."*

D2C9 *"And so when we made that big switch on our website, obviously the whole map changed and we did everything in our power to communicate this to our users and explain why we made this decision, why we felt it was right, and of course you just still get people who are totally pissed off and say, like, "ohh, you're matching the [public agency name], they clearly like were in your ear forcing you to do this". When reality, no, it was like actually a thoughtful decision on our part. Since then there was people really angry at the moment and I think it's this kind of calm down and there's more understanding for why we made that switch."*

Ineffective communication that lacks transparency can reduce trust to the point of seriously jeopardizing the project.

D2C10 *"Who will own the data streaming from sensors in every park bench, lamppost and dumpster in Quayside? No one at [company name], nor in local government, has given a straight answer to that question yet.*

[…]

When the Quayside project was announced last year, the terms of the contract between the company and [public agency name], a government-created agency that has partnered with [company name] to develop Quayside, were not made public. This was because they included "commercially sensitive provisions," reporters were told at the time.

[…]

"[company name] faces a number of political minefields, as I see it," said city councilor when The [media company] reached her by phone. "I don't want to shut the project down, but the more that they don't answer critical questions ... [the more] they're going to erode their social currency.""

D2C11 *"The lack of information coming from official channels likely increased skepticism.*

[…]

inBloom representatives and partners repeatedly failed to address questions and concerns about student data privacy and security. When present at meetings, references to brief, often legalistic, statements and assurances about FERPA compliance only hardened opponents' view that inBloom and their partners were aloof, secretive, and condescending.

[…]

inBloom representatives had difficulty in articulating what the concrete value would be for individual students. With public doubts about safety, and no clear success case to show, proponents of inBloom lost control of the narrative. Regardless of whether inBloom could have fulfilled its promise, without a clear narrative of its potential benefits for education, it failed.

[…]

Trust between states and districts, between districts and teachers, between parents and states, and teachers and technology was not adequately addressed by inBloom's communication strategy"

5.2.2.2. Trust structure critical factors

Non-profit legal form

Among the governance factors that influence trust-building is also the legal form. It has been found in four cases that the presence of an organization representing the collaboration that is non-profit in nature provides a perception of greater security of the DC.

D2C2 "*By being a non-profit, it really does feel build trust, especially in the early days when no one really knew who we were. It's like a badge of honor that build trust."*

D2C4 *"We chose to be a cooperative because we felt that it was the kind of organization that represented the most our idea that data keeps in propriety of people. [cooperative name] is a group of citizens, you can join [cooperative name] and take part in our assembly. This is a way to express trust in the sense that is a project made by everyone."*

Since DCs rely on the use of data, including sensitive data, it is important that a structure is in place that discourages the pursuit of profit at the expense of social benefit.

D2C11 "*There were also real parental and community concerns about the possibility of the on-selling of student data to third parties for profit. Research interviews with [stakeholders associations] indicated their deep worry that the data infrastructures underpinning schooling systems would potentially be controlled by private for-profit interests, including the world's largest edu-business, [company name].*"

It was also found that the trust generated by the legal form depends on the context in which the DC operates. For example, in some countries where the non-profit culture is historically less widespread, this type of legal form may even have a negative effect on trust.

D2C8 "*Another one can depend on the national or the internal level. [company name] is a social cooperative. When we say it at the international level is easier to create trust, because they know the ecosystem and what we can do. In Turkey, they don't understand what social cooperative is, what social enterprise is.*"

5.2.2.3. Trust "in-between" critical factors

Relationship network, reputation and previous experience, individual leadership

This factor, as well as for the initiation dimension, is also widespread in the trust dimension. Such a strong presence in these two dimensions is also due to the relations that exist between them. Trust is a component that influences the initiation, particularly when it is due to the presence of personal relationships or previous experience.

D2C2 "*The initial startup trust was just trust in [person name]. The other water suppliers trusted what [person name] was saying, they trusted her vision and so they went along with it.*"

In this factor emerges the importance of the leadership of individuals and their ability to create personal bonds of trust.

D2C1 "*There's no trust between two organizations, there is trust between people in those organizations. It is very subjective. It's between people and I think it's always been like that.*"

D2C4 *"We know each other, all of us. It's like friendship connections and we are all around the health research sector in Catalonia. So, we more or less know what we all are doing."*

D2C5 IT *"La fiducia è anche reti di relazioni costruite negli anni."*

D2C5 ENG *"Trust is also networks of relationships built over the years."*

In cases where trust is not generated by personal ties, the organization's history may help reduce the trust-building effort.

D2C3 "*They saw that we were doing business with the bigger municipalities in the Netherlands. So, if you do business with bigger municipalities, smaller ones think that they do it right because the bigger municipalities also do business with them.*"

Competence and data expertise

Another factor impacting trust between the actors in the collaboration is the demonstration of the right competence. This aspect was found to belong more to the actor in charge of managing and processing the data. This factor is also recognized in the literature as one of the fundamental components of trust, referred to as trust in ability.

D2C1 "*You have to show your technical capabilities that you actually know what you are doing.*"

D2C6 "*Then going forwards we have got very well-established processes and protocols for doing lots of things, which mean that we are hopefully very trustworthy.*"

5.2.3. Formal structure: critical factors

Compared to the other dimensions investigated, this one has a greater homogeneity of results. All the factors identified are governance structures and there is great agreement between the cases.

Operational agreements with partners

The first type of formal structures concerns operational agreements. This includes agreements in which the interest in participating in the collaboration is defined and in which responsibilities are assigned to each actor. In most cases, these are simple agreements, often in the form of a memorandum of understanding.

D3C4 "*The agreements are not technical at all. They are very simple.*"

This does not exclude the possibility that in the continuation of the collaboration these agreements may be reworked to define the activities and responsibilities of the members more precisely.

D3C5 IT "*Il concetto è: definisco un agreement sugli interessi e una disponibilità a fare quello per cui sono stati ingaggiati, ossia concedere i dati. Irrobustito perché potenzialmente potresti regolamentare meglio da una parte che tipo di dati, come li dai, qual è il livello di collaborazione che chiedi, dall'altra parte magare dare anche delle opportunità concrete di ritorno. Sicuramente si può dettagliare meglio questo tipo di collaborazione***."**

D3C5 ENG "*The concept is: I define an interest agreement and a willingness to do what they are contracted to do, which is to give the data. I strengthen it because you could potentially regulate better on the one hand what kind of data, how you give it, what level of collaboration you ask for, on the other hand maybe give concrete opportunities for return. Surely you could better detail this kind of collaboration.*"

These kinds of structures also have utility in maintaining roles over time, especially in contexts where individual organizations are changeable.

D3C7 "*So the MOU [memorandum of understanding] definitely helped keep people committed because there's change of leadership and all the organizations. Starting up [first partner name],*

but [second partner name], [third partner name] have new presidents every year. So it was it was good to have an MOU to keep continuity"

Arrangements must adapt to the characteristics of the partner. Therefore, it is possible to have different contracts depending on the sector partners belong to, i.e., public, private, and non-profit.

D3C8 *"Agreements change depending on the partners. If the companies want to run a project we sign agreements that show all the rules, a real contract (written amount and date, etc.). It's a long-term contract. For the public authorities generally we use a memorandum of understanding, with no official responsibilities. We ask the public and local authorities just non-monetary support, in kind support. For example, for a project one of the activities was providing capacity building activities. We are working with municipalities for the venue, but we don't get any money for political issues. We have different kinds of agreements with companies, public authorities, NGOs. If the partner gives or donates money the contact change."*

Non-disclosure and Data limitation agreements, Presence of external committees

This factor is linked to formal structures that, with the aim of safeguarding them, place limitations on the scope of DCs. Non-disclosure agreements and specific agreements that determine the constraints in the use and sharing of data are part of this factor.

D3C6 *"Very clear ways that we sign agreements for data sharing, clear tiring of who is able to access what data sets at what point and what we do with those data, then we have a kind of sign off process with partners, which is that they have oversight of a lot of the things that go on. So, there's a shared trust there in the sort of outputs that we're able to produce as a research center. So, I think that's very important."*

D3C2 *"In addition to the membership fee, each agency signs a nondisclosure agreement, so again, going back to trust and the nondisclosure agreement basically says that we're not gonna share their data or use it for anything beyond apart from the data collaborative, unless they give us written permission."*

D3C3 *"Public bodies sign an agreement that they don't give us the data. They let us manage their data, but it's always their data."*

A second type of structure associated with this factor is the presence of committees formed by members from outside the DC actors. These committees may be more technical or more ethical, but in any case, they fulfil the function of deciding which projects the data should be allocated to and screening the entry of any new actors.

D3C6 *"So if you decided that you would like to use some of the data that we warehoused within the [project name], you'd come to us and you'd submit a provisional request. Then we would ask you to fill out a comprehensive request, which details your proposal, your ethics, and whether you've got ethics, whether you've had this level of secure training that you need to access that particular data set. Then we take that to review. We have some expert reviewers who*

look at all of those proposals and then we take it to the data provider who has ultimate sign off on that proposal."

D3C9 *"We decided we are just going to match the [public agency name], we did a lot of research. We talked to our Board of advisors who include serious medical experts and they agreed too."*

5.2.4. Intermediation: critical factors

5.2.4.1. Intermediation process critical factors

Privacy and regulation compliance

The presence of an intermediary representing the collaboration is useful for compliance with certain legal requirements. For example, if a DC needs a specific certification in order to access development opportunities, it becomes important to identify an organization within or create a new entity.

D4C5 IT "*Un altro beneficio potrebbe essere essere riconosciuto in determinati ambienti, ad esempio vuoi diventare un acceleratore certificato."*

D4C5 ENG "*Another benefit could be recognised in certain environments, e.g., you want to become a certified accelerator."*

The presence of a recognizable and certified entity also guarantees the possibility of collaboration with certain actors, such as with certain public sector actors, and represents a barrier to entry for new emerging DCs.

D4C9 "*We're not big, but you need to be willing to navigate all that government bureaucracy to be recognized and trusted. And I think that's why these bigger orgs like [company name] or some of these orgs who can put tons of resources behind getting [certification name] certification kind of like going through these processes to be officially recognized. That's that was a huge barrier."*

Partners and stakeholders' engagement

Organizations acting as intermediaries, with their boundary spanning role, help to enable greater involvement of DC partners. Among the cases analyzed, not only the ability to engage members but also to interface with stakeholders and create communities emerges.

D4C8 "*They can implement the projects but it's about knowing what disadvantaged and disabled people need. So, you need to connect with the community. What they need? What is the bottleneck of this community of or region? Or if we can create this kind of platform, how we can sustain this platform. So, we are always thinking about these kind of problems. To do it we need to connect with panthers, users, experts, we are creating an ecosystem."*

D4C3 *"In the sensor project [partner name] has the responsibility to find people who were actually going to cycle. They decided to make a community and appoint a community manager, so we had like a project leader that only talk to the people."*

D4C1 *"I have some self interest, but still I would consider this is a good sign when there is a intermediary between of the different stakeholders."*

Data standardization and sharing facilitation

One of the main functions performed by intermediaries in DCs is to facilitate the exchange of data between actors in order to pursue the purpose of collaboration. Not all actors are equally willing to share their data, especially if they are private companies for whom data can be a source of competitive advantage. The presence of a trusted intermediary enables easier data sharing.

D4C8 "*For us it is easier to work with third parties, but for the companies it is not easy to collaborate with other companies, to share their data."*

D4C4 *"It legitimizes citizens' rights to control their health records while facilitating data sharing to accelerate public research innovation in healthcare."*

One of the principles on which DCs are based is the re-use of data. When data come from different sources and have inconsistencies or deficiencies, operational tensions arise within the DC. It is the role of the intermediary to overcome these technical problems by standardizing the data.

D4C2 "*That's a lot of diversity in billing systems, and so every data export that we get from an agency is in a different format. There's not a lot of standardization. And so, we have to sort of impose that standardization. So, we've built some data pipelines and software. What I did for my first three years with the collaborative is building out our data infrastructure to pull in this data, standardize it across agencies so that it's easy to do comfortable analysis across these different agencies. We tried as much as possible to make this transparent to the user. A lot of it is work on our end. I'd say it could generate tensions."*

D4C3 *"In the beginning there weren't any data standards within the Netherlands, so we tried to make logical metadata standards for them to comply and send the data to the national data. Nowadays, we don't do that anymore, but we try to be helpful in creating better metadata. So, if you want good data standards you need some obligated fields they have to send with the data."*

5.2.4.2. Intermediation structure critical factors

Innovation, technology, and data infrastructure

Among the governance structures of DCs are certainly data management technologies. The technology infrastructure is primarily administered by the intermediary. Cases have shown that the infrastructure may also be shared between more than one actor in the collaboration.

D4 C6 "*We have shared knowledge across teams in doing all of those things and I think there's a real benefit to doing that rather than having a single way of doing it, which is kind of "don't buy one single person". It also makes our secure environments more accessible, so we have got secure labs in [university name], we've got secure labs in [university name], which means that people can go to London and look at data or they can go to Leeds to look at data."*

It was also found that the governance model of collaboration may be influenced by the type of technology infrastructure, since with certain solutions, the formation of a dedicated entity is required.

D4 C5 IT "*Anche il tema infrastrutturale, come può evolvere il data club? Una strada può essere quella di farlo evolvere come data space, seguendo anche i trend europei. Allora il data space è un soggetto che necessita di una governance. Se il data space è un soggetto neutrale, allora deve essere neutrale anche la governance, non può essere un soggetto singolo, allora potrebbe essere una necessità di un soggetto terzo indipendente."*

D4 C5 ENG "*Also the infrastructure theme, how can the data club evolve? One way may be to evolve it as a data space, also following European trends. Then the data space is a subject that needs governance. If the data space is a neutral subject, then the governance must also be neutral, it cannot be a single subject, then it could be a need for an independent third party."*

Non-profit legal form

In cases where an entity representing the DC has been identified, this is almost always configured with the non-profit legal form. In fact, four of the successful cases have the presence of this structure and only two cases showing a specific data intermediary do not take this form. The forms found are the 501c3 non-profit of the American system and the social cooperative.

D4C4 *"Salus Coop is a non-profit data co-operative for health research"*

D4C9 *"We're a 501C3 nonprofit."*

Presence of a dedicated and neutral intermediary

In each of the cases, the presence of an organization acting as an intermediary was noted, particularly regarding data exchange. The presence of an intermediary brings numerous advantages within the collaboration. Firstly, the neutrality of the intermediary mitigates tensions due to different political views and due to competition between the actors themselves and alleviates the risks of power imbalances.

D4C2 *"Being able to move things to a third party made it so that the members and the participants could be more equal. [...] I mentioned some of the agencies that departed early. The fundamental underlying issue was a difference in vision and in particular there was some disagreement around our role of policy analysis. There is a law in California about water use targets for supply agencies. Before this law was approved, there were some real differences in vision around this. That was a source of tension. Since that moment, we established the principles to not explicitly endorse legislation, but just analyze policy impartially."*

Secondly, the presence of an organization representing the DC facilitates the management of support activities for the initiative, such as human and monetary resource management and process monitoring. Thirdly, it allows the organic growth of the DC by being able to deal with the search for new funds and the hiring of new staff entirely dedicated to the initiative.

D4C2 "*So I think in terms of handling money, hiring staff, getting work done, there are sort of concrete benefits of a third party. Maybe if there was no money, if we didn't have full time staff, if it was a more sort of volunteer driven process, then a third party wouldn't be necessary."*

D4C6 *"We do things like hiring data scientists on to a training program in order to undertake our own research, we support Masters students and dissertation students, we have an in house data science team as well."*

The most common form of intermediation detected is one in which there is an organization that is entirely dedicated to the initiative, generally created specifically for the DC.

D4C4 *"I can't imagine this project without a legal entity that represents the project. It could be a cooperative, a foundation, whatever, but you need something. You need an ID, you need a legal form to sign a contract, you need some statutes."*

However, the creation of a dedicated entity, despite the aforementioned advantages, also requires a considerable initial bureaucratic effort that violates the principles of a lean development. Therefore, this type of governance form is suggested at a mature stage of the initiative.

D4C5 IT *"In questa fase è prematuro parlare di un ente terzo. Magari ci saranno anche delle esigenze specifiche, ad esempio gestire il funding. Le singole strutture che partecipano al progetto non hanno la conformazione adatta a gestirlo. Però adesso è prematuro. Non deve venire tra 10 anni, ma nel momento giusto, quando c'è la necessità puntuale di attivarlo, altrimenti per soggetti di questo calibro... Noi siamo piccoli strutturalmente e abbastanza agili, però fondare un oggetto nuovo implica tanta burocrazia."*

D4C5 ENG "*At this stage it is premature to talk about a third party. Perhaps there will also be specific requirements, for example to manage the funding. The individual structures participating in the project do not have the conformation to manage it. But it is premature now. It doesn't have to come in 10 years, but at the right time, when there is a specific need to activate it, otherwise for subjects of this caliber... We are small structurally and quite agile, but setting up a new object involves a lot of bureaucracy."*

In one case the presence of an intermediary is found to have no formal independence but is formed by a collaboration agreement between universities where processes and responsibilities are fully shared between them.

D4C6 "*But we are based at different universities and there are teams in each university. There is some level of autonomy there as well. Because we've been going for so long it's, there's a lot of activity that sits underneath that umbrella, basically. It works quite well. There is lots of*

shared effort in things that are difficult, so ways of working or building trust with partners, etc. We have shared knowledge across teams in doing all of those things and I think there's a real benefit to doing that rather than having a single way of doing it, which is kind of "don't buy one single person".

In another case the intermediary role was not neutral but played by a branch of a partner organization. In this situation, the key element in the sustainability of the project was that the branch had been created specifically for the DC and had staff entirely dedicated to those activities.

D4C7 *"So the fact that [partner name] invested in a team which was only looking at global, only focused on this project, that was really important to the ongoing sustainability of the work. Otherwise over a 3-year period of time if someone was doing another day job, the project would have probably fallen especially with COVID and everything."*

5.2.4.3. Intermediation "in-between" critical factors

Competence and data expertise

As mentioned in the description of the previous factor, DC intermediaries are characterized by high capabilities in data management processes. The presence of specific skills for the secure use of data is a key element for this type of organization.

D4C6 *"We do things like hiring data scientists on to a training program in order to undertake our own research, we support master's students and dissertation students, we have an in-house data science team as well. So, the intermediary data supply piece is part of it."*

However, key competences can also relate to other aspects of the DC, such as the social problem to be solved.

D4C8 "*Off course, for companies it is possible to interact with beneficiaries without the intermediation of [company name], but they don't have expertise. They have resources, but not expertise, only some of them have it. They can implement the projects but it's about knowing what disadvantaged and disabled people need."*

The absence of an intermediary who possesses the ability to interact effectively with the target beneficiary can be a cause of project failure.

D4C11 *"Even the people that districts sent were their data people. They weren't their instructional people, they weren't their curriculum people. It operated strictly in the data world."*

5.2.5. Incentive system critical factors

5.2.5.1. Incentive system process critical factors

Pressure and facilitation by public institutions

Similar to what was observed in the initiation dimension, the incentive dimension for actors to participate in the DC is also linked to the action of public institutions. Public authorities can exert the right pressure on private and other public entities to share data.

D5C3 *"I guess governmental pressure of making the system perform better was the major interest in giving us the data."*

Partners and stakeholders' engagement

One aspect that emerged among the incentives for DC partners was the level of involvement. Actors are more willing to participate in the collaboration if they can be involved in the planning and designing of activities. In this sense, the concept of co-creation of initiatives is prominent among the cases.

D5C6 *"It's called cocreation. You build your research questions in collaboration and generally speaking, that's what we do. […] It ties back into what I was talking about at the beginning, which is this shared vision, this buy-in, this co-creation and trust building I think it's really important."*

One element that can foster the level of involvement of partners is the planning of long-term interventions. It could be argued that partner engagement is a success factor for long-term sustainability and that this is positively influenced by long-term projects, thus activating a virtuous circle.

D5C8 *"A second element is that we are always working with companies in a long term. It is always a long-term program; we are not asking money to do things. We are proposing a project which includes at least 12 months program, and the project is able to involve the employees of this organization too. Active participation of the employees in the program is really important."*

Social responsibility and impact measurement

The social responsibility of partners is one of the incentive factors on which there is most consensus. Thus, large companies decide to participate in social initiatives such as DCs mainly through Corporate Social Responsibility (CSR) programs.

D5C1 "*Many mobile network operators want to do good for the society, so that's why they participate. They have some social responsibilities, and they would do those data collaboratives to fill them."*

D5C4 *"We made an experiment in 2018 in which we face people with several scenarios. What we understood from this game was that people were willing to donate without any money compensation in the short term. Those are the five conditions in which we are now relating to citizens. Something that matters to society, open access to result, non-for-profit and public institutions, anonymization and you can join but you can revoke."*

D5C6 "*Then there is the definitely the kind of corporate social responsibility angle that I pointed to. We are very clear that the research that we do is for the public good."*

D5C8 *"Some companies just want to confess their sins. They are doing many damages for the environment and the community, so they just want to create some social responsibility project and we always propose high impact ones."*

5.2.5.2. Incentive system structure critical factors

Innovation, technology, and data infrastructure

Being able to exploit new technologies and techniques to use the data held by an actor is an incentive to participate in collaboration. With respect to the generic case of cross-sector social partnerships in DCs, the component of technological innovation is relevant and becomes an attraction for organizations interested in enhancing the use of their data.

D5C1 "*When we talk about data collaboratives, then of course innovation is something that is throughout the different stakeholders, so innovation it would be for everybody incentive just to do new things and that is kind of the goal. For the operators, it might as well be that they want to learn more about their data because those kinds of initiatives could reveal for them something that they didn't know before. Bringing out some information that might be for them useful in their operations."*

D5C3 *"It [technological platform] was new and they found it very interesting to just upload their data.*

[…]

What usually happens is that if you buy a sensor, you also get a dedicated like vertical data solution. But if you buy 50 different sensors, you have 50 different data housing solutions. If you use ours, you only have one, so you can buy combine all these data from all these verticals on our platform. That's a big advantage."

Non-profit legal form

Since the incentive dimension is linked to the trust dimension, where greater trust implies greater incentive to share data, there is a connection between the legal form of the receiving entity and the propensity to share.

The non-profit legal form induces an incentive for data owners. This is found both for data sharing by private companies and by individuals from civil society.

D5C4 "*They were fine giving their data for social good, for open research, for some disease that might be of their interest to public institutions on the higher level of anonymization, so maximum respect for privacy, and ensuring that it was something that they could revoke the participation. Those are the 5 conditions in which we are now relating to citizens. Something that matters to society, open access to result, non-for-profit and public institutions, anonymization and you can join but you can revoke."*

D5C6 *"Because we've got this kind of nice middle ground as being academics, we can potentially do more by talking to a range of people that you couldn't do otherwise. So, I think there clearly are benefits because we do get good engagement."*

5.2.5.3. Incentive system "in-between" critical factors

Interest in data outcome

Within a collaboration, the interests of the actors are different, but intertwined. While on the one hand, the output of processes has a positive impact in relation to a social problem, on the other hand, actors may benefit from the processing of data. Indeed, the most widespread incentive for data sharing is simply the possibility to obtain a service in which there is an interest aligned with the main mission of the actors.

D5C1 *"Although they have this kind of fraud detection systems already in place, sometimes they can't discover them, because the product and people also start to get better and better. We were able to discover it because those SIM cards didn't behave like phones and they were very concentrated into one area and then that came out when we tried to make sense of the data and that kind of things are interesting for the mobile network operators as well, but that there are of course more use cases."*

D5C2 "*The incentive is in terms of the benefits that we can provide back to them. We can have all this data on hand that when an important policy question comes up, we can conduct some policy analysis and respond rapidly. So, it's sort of this beneficial circle where member agencies fund us, they provide, they govern us. We really have to phrase it in terms of the benefits that they're getting in order to incentivize them to share data because there are folks who are skeptical about sharing data, they want to protect their data. By phrasing it in terms of "if you join the collaborative, you know you get access to this software that we built as a collaborative". You know you're contributing; you get the option to share your data to support research. You know to help respond quickly."*

D5C3 *"The people are very curious about using the sensor, because they want to know how the air quality is on their route."*

D5C6 *"There's definite benefit from providing insight. It might be that they can better understand their customer's behaviors or it might be that they just see something that's going on that they didn't see before."*

D5C7 *"I think for partners like [partner name], they have a better data set now."*

D5C9 *"I think really just their motivation was they had a bunch of pain points and our product or our tools just resolved a lot of their pain points and kind of cut out all the noise and and just reduce their anxiety or reduce the sense of being overwhelmed by all this information."*

5.2.6. Business model critical factors

As in the description of the factors of formal structures, also in the case of the business model dimension the results are homogeneous and belong to the same type of governance element, i.e., the intersection between processes and structures.

Revenues generation

Although most of the cases surveyed claim to be generating revenue or are in the process of doing so, the types of services offered, and revenue stream models vary widely. In the cases analyzed, revenue generation is more related to the organization acting as an intermediary than to the collaboration as a whole.

Among revenue-generating activities, a first distinction can be made between 'core' services, i.e., those that are very closely related to the purpose of the collaborative, and 'non-core' services, which are useful for survival during the early stages. Core services and products include the provision of technology infrastructure, access to software tools, platforms, APIs, and datasets. These monetary flows are mostly internal to the DC, i.e., they come from an actor and generally flow to the technological intermediary.

D6C2 *"Most of our funding comes from annual membership revenue. That's probably 70% of our revenue."*

D6C3 *"We currently have 4 products: open data platform, data catalogue for municipalities, reporting application for companies, sensor project. The revenue model is always the monthly fee, because it's hard to make money by selling data from public."*

On the other hand, non-core services are more frequently directed towards actors outside the collaboration. The most common of these is the provision of consulting services, but there are also cases of strategic orders such as software development. More rarely there are revenues from activities less related to technological competences, such as the organization of events.

D6C1 *"In general, we offer a consultancy and for the calculation of those statistics [company name] has its own software to do that that will drive down the cost."*

D6C8 *"The fourth one [income stream] is that we are developing some map-based platforms, like Ukraine platform. We are developing this platform for international organizations and public authorities to collect data according to their needs."*

D6C9 "*That's us, that's a client, it's a partnership. But at the end of the day is a kind of a client relationship where they're asking us to build this dashboard. And so that's really the one I was saying about what we've been trying to explore was that we were trying to find our new revenue model and that new revenue model we thought could be making these like, you know, 100 to 50K dashboards where we would give them a little extra like design work, and we'd go through, you know, a few weeks or maybe a month of sort of customer or user research or whatever."*

Revenue stream models are equally varied but related to those typical of the software industry. There are freemium models, one-off license payments, recurring payments, or transaction fees, depending on the type of service/product offered.

D6C8 *"The fifth one is social marketplace. As a supporter, if you want to support something you purchase some materials from some e-commerce sites, like Amazon. You can purchase the material from local businesses. We get a commission for all the transaction on the social marketplace. "*

D6C9 *"We actually have been able to kind of have a funding model where we can license access to the API and the data to these businesses. And so that was actually a model we used. So that would be the one case of like having official agreements in place where we're, you know, it's like commercial licensing."*

Private donations, grants

One factor found among all the cases analyzed is the presence of funds from private donations or public subsidies. These two income lines are the main ones for most of the cases analyzed.

Public funds can be local, national, or international and support initiatives mainly by financing the initial stages. Private donations may come from foundations, individuals or companies that provide funds through CSR programs.

D6C1 *"They organized an open tender for everyone to participate, so we were one of the companies to participate in the tender. We want the tender and that's the model."*

D6C2 *"We also sometimes receive grants."*

D6C3 *"We get some EU funding for the European project."*

D6C4 *"At this moment we are financed 100% by public administration."*

D6C6 "*We are almost entirely [national council name] funded, which comes from [larger national council name]. We have grants of determinate length."*

D6C8 "*We are collecting donations; this is the first income generating activity of [company name]."*

In addition to these traditional forms of donations, DCs present a further modality in the form of the donation of data for the implementation of the initiative.

D6C5 IT "*Chi ha messo a disposizione i dati ha portato comunque un valore."*

D6C5 ENG "*Those who made the data available still brought value."*

D6C7 *"[partner name] contributed financially, but we didn't pay any of the other organizations [which shared data]."*

Although these funds are mostly the primary source of revenue for these projects, they are subject to fluctuations that are difficult to predict and can cause financial strain.

D6C9 *"Donations are so unpredictable and so uncertain and that's why these projects like what we did with [project name] that was a funded project."*

5.2.7. Adaptation: critical factors

5.2.7.1. Adaptation process critical factors

Communication and transparency

Among the factors that consent the adaptation of the DC by overcoming tensions between the actors is communication. As described above, communication allows for an increase in trust between actors and this is true even when trust is damaged.

Transparent communication allows to overcome tensions related to the entry of new actors into the collaboration.

D7C8 *"Some donors reached us to work on this platform and we see that we have different kind of organizations. So, we asked them "we are working with you because you contacted us before, but we also have another partner who wants to be involved, do you accept it?". Some organizations have restrictions in collaborations with other companies. For us, the easiest way was to directly ask them."*

It also enables to clarify and manage the expectations of the actors regarding the outcome of the collaboration.

D7C6 "*Then there are sometimes tensions with the users of the data or people who want to collaborate, because quite often the understanding of what consumer data are is not that great across the board, so there will often be an idea that you can do anything with the data because it's so big or so good, but actually all of these data are messy or incomplete or not quite in the format that you might like. Trying to manage the expectations of collaborators is really important in that severe, and again it comes down to the communication and demonstrating the value, or at least working through a research question that might be as good as the original one."*

Another aspect of communication is related to external accountability with respect to collaboration. As demonstrated by the failure cases analyzed but also by the tensions experienced by one case interviewed, it is crucial not to underestimate transparency towards stakeholders. When communication is unclear or delayed, tensions can arise that undermine continuity of the DC.

D7C1 *"One very important part of those data initiatives is public as well. It has killed already quite many data collaboratives. All the different parties that we talked previously might agree on data collaboratives and then there is a journalist who will dig up the thing and publishes it in a way that there is a public backlash. This is why my recommendation would be to always engage with the public as well, especially when you use public data."*

Openness toward new partners and partner selection process

The DC's openness to new actors is a factor noted in all the cases analyzed. New partners can be crucial for sustaining growth and durability of the initiative, as they can bring new skills, resources and data to respond to environmental changes.

D7C1 *"It depends on the kind of how long term the projects are. If they are longer term, and for most of the data collaboratives I would say that the longer activity is a goal, environment changes, so you need to change as well. Different partners in the project or you might change as well the goals of the project."*

D7C5 IT "*C'è la necessità di avere nuovi partner. In primis, sulla parte dati."*

D7C5 ENG "*There is a need for new partners. First of all, on the data side."*

D7C6 *"We tend to onboard data providers quite often, through all of the models that we've discussed where we're continually building those relationships."*

In some cases, new actors are seen as new customers for the intermediary, so there is a high level of openness.

D7C2 "*New agencies are absolutely welcome to join and we're actually trying to grow right now. Just last week a new agency decided to join."*

D7C3 *"We sell them the product."*

D7C9 "*I guess partner for us at this point would be like new potential opportunities for sustainability like a new, we don't want to say customer, but that's really what they are."*

In others, there may be criteria for selecting new partners, but they are limited to an initial screening of economic activity.

D7C8 "*We apply some criteria to select who to work with. If the company is involved in controversial activities, such as about terrorism, arms dealers, etc., we cannot work with it. Otherwise, we can work with all companies if they accept our mission and the rules of the project."*

More rarely, there may be more structured selection processes that go through external committees. On the one hand, this slows down the growth of the DC, but on the other hand, it can prevent internal tensions and shifts in the mission of the collaboration.

D7C4 *"We have not experienced any tensions because we have been very conservative on partnership. Of course, we want to sign agreements with any research institutions, but as I said we are growing in terms of people participating, people being members of the cooperative, very slowly. […] Project that we host on our settings need to have the consent of an ethical committee. You cannot reach us and say, "I want to make research on blablabla". You go through an ethical committee and when you the approval you can collaborate with us. It's a way in which we can filter quality and critical mass."*

Finally, the C6 case offers an interesting model of flexible collaborative governance. The structure involves a stable core of partners constituting the main entity and several partners who can join or leave the DC with greater freedom.

D7C6 *"We have the core actors as they are stated. So, the center is comprised. But we collaborate with a huge number of universities and other organizations. So, the model works well there, we can crowd in a huge number of different organizations, but they don't become organizations who are part of the CDRC."*

5.2.7.2. Adaptation "in-between" critical factors

Lean development

As generally accepted for innovative projects, lean development is also recommended in DCs. The increase in the volume and complexity of activities must be commensurate with the stage of evolution of skills and technologies within the collaboration. The adoption of an incremental growth model can prevent unforeseen tensions.

D7C11 "*Above all, [project name] needed to be an enterprise scale system in order to process the volume of data anticipated across the country from state and district partners with an expectation of continued growth. And as [participant person name] said, it needed to offer high security at scale on day one, adding to the development complexity. [...] The sheer number of actors at the state and district levels was overwhelming."*

6 Discussion

The previous sections described the research methods used in this project and the results subsequently obtained.

The results enabled the research questions to be answered. To answer the first question, twenty critical DC governance factors that influence the long-term sustainability of projects were recognized and will be reported below. To answer the second question, critical factors were identified that are not found in the literature on cross-sector social partnerships: compliance processes with respect to privacy and security regulations, processes for standardizing data and facilitating data sharing, the presence of specific expertise in data management and analysis, incentives based on the creation of shared value for actors through the use of data, a technology infrastructure capable of managing data, and specific agreements for protection in the use of data.

This chapter discusses the main aspects found in the research. Initially, attention is focused on the concept of governance adopted in the research. Next, an interpretation of the twenty factors is provided based on synthetic categories of governance, making explicit the links with the theoretical dimensions identified in the literature and the

definition of governance adopted. After that, a comparison is made between the analyzed literature and the results, bringing out common elements, elements that are not reflected in the cases, and original elements from the research. Some of the possible applications of the results to the practical development of DCs are then reported. Finally, the limitations of the study and possible future additions are discussed.

6.1. Governance concept

First of all, the definition of governance of collaborations is not yet an unambiguous concept. As explained in Chapter 3, there are several definitions. For example, some definitions refer just to processes [50], [54] or others consider structures only [56]. Compared to other definitions of governance in the literature, a less restrictive definition was used for this research. The reason was not to overlook any element that might be useful in identifying key factors for the long-term stability of DCs. This has led to the need to adopt an interpretive framework that can capture differences in the nature of governance components. In fact, critical factors were divided according to three categories: processes, structures and elements that lie at the intersection between them. A further complexity in the definition of governance is related to the socio-technical nature of DCs, where governance at the managerial level has some overlap with the more technical concept of data governance.

6.2. Synthetic categories

As explained above, the twenty critical factors extrapolated from the cases were further studied to provide an easier interpretation of the results. An initial classification was made based on the components identified according to the DC's definition of governance. Three types of elements that constitute the governance system emerge from the definition adopted: i) processes, ii) structures, and iii) elements that lie at the intersection of processes and structures. Then, critical factors were grouped according to epistemic affinity in seven synthetic categories. Table 8 shows the results sorted according to the nature of the governance components and according to the seven synthetic categories.

In order to provide a comprehensive view for the theoretical discussion of the research findings, a synthetic framework was then created that includes all the relationships among the concepts used in the study. Specifically, Figure 5 shows a graph where:

- on the left are the dimensions of governance for the long-term stability extracted from the literature,
- in the middle are the critical factors identified from the case study analysis,
- and on the right are the synthetic categories to which these factors belong according to the definition of governance.

In addition to the relationships between the critical factors and the theoretical dimensions, the graph also makes explicit the relationships between the dimensions themselves, as revealed by the analysis of the data collected. These are represented as arrows, in which the arrowhead is directed toward the influenced dimension.

Table 8 Synthetic categories and critical factors

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

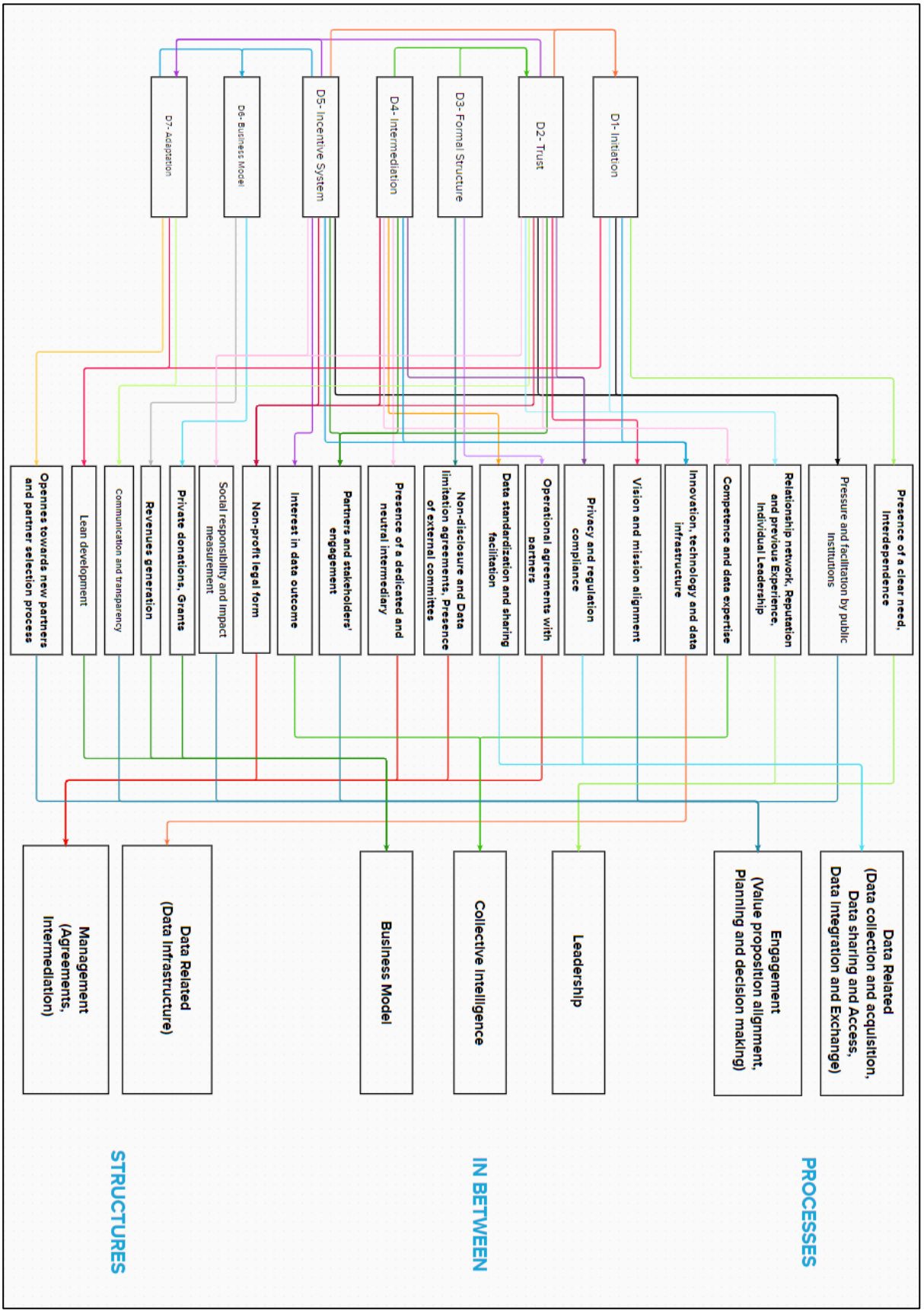


Figure 5 Relationships among theoretical dimensions, critical factors, and synthetic categories

6.3. Theoretical dimensions

Identifying the governance dimensions of the collaborations to be analyzed in the case studies was one of the preliminary stages of the research. It should be pointed out that these dimensions are not only related to the success of the collaboration in terms of the achievement of results but have a specific relevance in the long-term stability of the DC. This means that some dimensions have been privileged over others because they are more relevant for the systematization of DCs that can last over time. Examples of dimensions with a greater weight than in the literature of cross-sector social partnerships are the business model, the incentive system and adaptability. All these dimensions are closely linked to the continuation of the DC over time. A business model that grants financial autonomy to the DC makes it possible to overcome difficulties in finding resources for activities after the initial stages. An incentive system that guarantees each actor the right involvement and the right reward allows the partners, especially those who own the data, not to abandon the project after the initial interest linked to short-term objectives. The adaptability of the governance structure allows for an appropriate and efficient transformation of the DC in response to endogenous or exogenous tensions that may occur over the long term. Other dimensions less related to long-term stability, although important for DCs governance, were not considered so as not to burden the analysis of the cases, especially when collecting data from interviews. An example of these is represented by the general antecedent conditions to the formation of the collaboration.

6.4. Critical factors and literature

In the analysis of the results, the critical factors identified were compared with the literature on cross-sector social partnerships and DC.

In general, it can be seen that the different nature of the factors manifests itself in a different representation in the literature. The factors most related to managerial and organizational aspects of collaborations are predominantly found in the literature on cross-sector social partnerships. Factors that have to do with the exchange and management of data are only related to the literature on DCs.

Although there are several points of contact with the literature, the empirical analysis also found new elements that may prove useful in advancing research in the field. What follows is a comparison for each synthetic category detected with knowledge found in the literature.

6.4.1. Processes

Data related processes

This category summarizes those collaborative governance processes that are strictly concerned with the use of data. In general, high consensus with literature is noted

among these processes. With respect to the literature, privacy compliance has already been noted as a critical issue [17], [18], [44], [62], [105], [106], as well as data standardization and sharing facilitation [8], [18], [26], [44], [62], [63]. A point to note is that these processes related to data use are predominantly carried out by a single technology intermediary and rarely involve all partners.

The research also adds the importance of compliance with regulations and standards. Having certain certifications on data use qualifies collaborative data processes and adds legitimacy to the DC. This was noted as an important aspect at the project initiation stage. There may be certifications that are barriers to collaboration with certain public sector actors. Certifications can also be useful when the DC is already implemented. For example, they allow to access specific benefits, such as those for certified incubators.

Engagement

Instead, this category includes all the processes involved in involving partners from a managerial perspective. These include many critical factors already recognized in the literature on both DC and cross-sector social partnerships. Communication and transparency are aspects recognized as fundamental in this type of collaboration reviewed [20], [44], [50], [51], [54], [57], [62], [105]–[107]. Regarding this factor, the study adds empirical evidence of how its mismanagement has been a cause of project stability failure and how it is a critical aspect even in successful cases. The same goes for the facilitative action by public institutions [5], [17], [18], [21], [50], [51], [53] and for partner and stakeholder engagement [8], [39], [44], [50], [52], [53], [62], [105]–[107]. For both factors the research provides empirical evidence in the majority of the cases studied.

Regarding vision and mission alignment, it should be noted that this element, which has great consensus in the literature on cross-sector social partnerships [50], [51], [53], [54], [69], is little studied for DCs. The research has found that the absence of alignment on vision can cause a partner to leave the DC. Alignment on a data-driven mission can be a determining factor in bringing together different actors who would otherwise not have participated in a similar partnership.

Within the Engagement category there are also aspects less studied by scholars. One of these is the need to demonstrate the social impact of collaboration. While this is acknowledged by a theoretical study about DCs [13], there is not acknowledgement in the literature analyzed on cross-sector social partnership. In addition to the findings of the preliminary literature review, the presence of this element among the critical factors empirically observed further highlights how the connection between social impact generation and data collaborations needs to be investigated further in the literature. The research also focuses on the positive effects of social impact measurement in gaining legitimacy from collaboration and in building trust among actors.

Another aspect in contrast with the literature concerns the structural openness to new partners in the governance of these types of collaborations. This discrepancy seems quite surprising, given that the totality of cases interviewed stated that they met this requirement. Openness to new partners is a strategic element for these collaborations in acquiring new resources and data. Moreover, the study highlights that the inclusion of new members within the DC should be preceded by a selection process based on technical and ethical criteria.

6.4.2. Elements in between processes and structures

Leadership

Regarding the factors that make up this category there is generally high consensus in literature. The cases analyzed provide empirical evidence regarding the need for a clear social need for the initiation of collaboration in agreement with the position in literature of both cross-sector social partnerships and DCs [8], [20], [50], [51], [53], [56]. Similarly, the relevance in the trust formation and building phase of elements widely recognized in the literature such as relationship networks, reputation and previous experience [18], [20], [21], [37], [39], [44], [50]–[53] has been demonstrated.

What appears slightly at odds, however, is the role of interdependence. From the analysis of direct evidence, the presence of this factor in explicit form is rarely found. Literature recognizes the interdependence among actors in pursuing individual and collective interests as a necessary element for the formation and stability of collaboration. Its presence as a critical factor emerges from the study, but it is pointed out that given the small number of cases in which it was identified it is placed on a lower level than in the literature.

Another aspect that deserves consideration is the role of the leader. Compared to the different forms of leadership discussed in the literature, the cases analyzed show little formal leadership as opposed to leadership that emerges from the bottom up. In fact, the leader's role is often individual and internal to one of the organizations participating in the collaboration. The leader's personal connections with other members, past experience and reputation are key to the initial building of trust in the project.

Collective intelligence

This category considers the system of knowledge and expertise shared within the collaboration. Among the factors that make up collective intelligence, skills and knowledge possessed by actors as antecedent conditions are distinguished from those instead generated as a result of the collaboration.

On the former, consensus is found in the DC-specific literature [39], [62], particularly regarding skills on the use of data, so-called "data literacy" [1], [40], [44]. Possession of the appropriate data skills appears to be a prerequisite for these kinds of initiatives,

which are otherwise impossible to implement. The study also confirms the importance of possessing adequate knowledge of the target beneficiary, emphasizing the role of community relations and non-profit organizations. Compared to literature, the need for collaborative skills, meaning the ability of actor to collaborate with other actors as highlighted by Bryson et al. (2015) [51], does not directly appear from the case study analysis. However, it can be argued that this aspect emerges because of other factors, such as communication and partner engagement processes.

Regarding the knowledge generated by collaboration, some considerations can be made in relation to the incentive system. Among the research findings that have more empirical evidence in the cases analyzed is the interest in the outcome of data processing. This means that those who share data, even in the private sector, do so in addition to a social responsibility because they consider that the DC's activity can generate direct value according to their individual interests. Examples of this are intentionally developed services for data owners or positive externalities from newly developed knowledge. This aspect is almost totally ignored in DC literature. Only a recent study by the Global Partnership from Sustainable Development Data (2023) [44] highlights the need to create shared value for partners. Given the relevance of this aspect to long-term sustainability, more research in this direction is suggested.

Business model

Overall, the business model category is novel compared to the literature on the governance of cross-sector social partnerships. One reason for this may be that cross-sector partnerships often are financed by a single partner or sector [108]. In DC literature, however, the presence of this component among the critical factors has been noted with some significance, as reported in Chapter 3. Another reason may be the more application-oriented nature of certain studies analyzed about DCs.

The importance of using revenue sources from private donations or public contributions noted in literature [17], [39], [44] is reflected in the empirical study. The totality of cases analyzed recurs to one or the other or both sources. Revenue generation assumes greater importance in relation to the long-term sustainability of projects for the literature [17], [44]. This aspect is also in agreement with the results, although it was found in fewer cases than grants and donations.

Like the other critical factors belonging to this category, lean development was only found in the literature concerning DCs [17], [21], [44]. Again, the research confirms the criticality in the development of collaboration from the transition from pilot project to enduring initiative. The cases analyzed offer empirical evidence of how the failure or success of a DC can be determined by the predisposition to change and experiment with new ways to achieve the desired social impact.

6.4.3. Structures

Data related structures

This category is related to the intensity of technology that distinguishes DCs from generic cross-sector social partnerships. Indeed, in the literature on cross-sector social partnerships, the concept of technology emerges only in the framework of Bryson et al. (2015) [51] as a tool to facilitate collaboration. In contrast, there is more evidence in the DC literature [18], [44], [49], where technology infrastructure is considered a foundational element of collaboration. What also emerges in a novel way from the study is how the use of technologies new to partners is an incentive for participation as a process or product innovation. For the systematization of DCs, it is relevant to effectively recognize and communicate the features of the technology infrastructure not only in terms of security and reliability in data processing but also as capable of offering technological improvements over the partners' current state.

Management structures

This category consists of the structures that regulate the relationships between the actors inside collaboration and with the outside world.

In accordance with the literature on cross-sector social partnerships [51], [53], [56], the study confirmed the importance of the presence of a neutral entity that facilitates collaborative action through facilitative leadership. Also, in agreement with that stream of literature [53], the case analysis highlighted numerous advantages in establishing an entity dedicated entirely to collaboration from a management and legitimacy perspective. The DC literature [17], [26], [34], [39], [59], [109], on the other hand, focuses on the importance of the role of a technology intermediary to act as a trusted third-party actor for receiving and processing data. The research offers empirical evidence for these claims as well.

What appears novel compared to previous studies is the legal form that this dedicated intermediary organization should take. Although the literature reviewed does not address this aspect, it has been shown that the non-profit legal form allows for increased trust of actors in sharing critical resources and confers legitimacy toward stakeholders in pursuing social goals.

Other discordant aspects with respect to the reviewed literature concern formal agreements. The research found that there are two main types of agreements: operational agreements and data protection agreements.

On the former, feedback can be found in much of the literature, but although there is broad agreement on the problem they are intended to solve [20], [51]–[53], namely that of clarity of roles and responsibilities, they scarcely appears to be a major issue. Only Koschmann (2012) [57] assigns a first-order role of these agreements in the construction of an authoritative text that triggers collective action. Furthermore, the study reveals their importance not only for clarity in day-to-day activities but also for DC stability. The presence of such formal agreements protects the collaboration by binding partner organizations where there are unforeseeable changes in management and thus ensuring continuity in support to the initiative.

Regarding the second type of agreement, again the study raises an issue that is less considered in literature. Only a study by Susha et al. (2017) [16] specifies the need for formal documents among the control mechanisms that protect data owners in sharing. In contrast, Klievink et al. (2018) [21] argue that the presence of conditions that overly delimit the scope of the dataset can be a factor in the failure of DCs. From the empirical study, legal protection against the risk of harmful use of data seems to help increase trust between partners and eliminate possible adversities to data sharing, a widely recognized problem in the formation phase of DCs [16], [44], [106].

In conclusion, one aspect of which no trace was found in the reviewed literature is related to the presence of external committees. The role of these committees was noted in the cases as important in the selection of new partner organizations, as well as in the choice of single projects to be undertaken by the collaboration. The relationship between this aspect and the sustainability of DCs should be explored further by future studies.

6.5. Factors in the literature not revealed by the cases

It is important to note that not all the elements highlighted in the literature find correspondence in the critical factors extrapolated from the cases. As mentioned in the previous section, the concepts of leadership and interdependence promoted in the literature found less space in the case study testimonies, but they are not the only ones.

Prominent among these factors is the construction of internal and external legitimacy [51], [52], [56], [107]. Although not specifically identified within the cases, it is possible to associate it with the trust-building process widely explored within the research. The concepts, though strongly connected, do not overlap, and trust may be only one of the components that constitute legitimacy.

The literature also shows the importance of the planning process [8], [53], [107]. A dual origin of this process is identified: deliberate and emergent planning. Related to this process is performance measurement and goal setting [8], [20], [50], [51], [54], [106], [107]. According to the literature, planning can be deliberate or emergent, and goals should provide a path for achieving small but frequent successes. Also, part of this process is reevaluation [51], [107] and identification and mitigation of risks, particularly regarding data [37]. From the cases this process is not detected as a critical factor. The absence may be justified by the research question, which does not aim to describe the fundamental processes of DC governance but wants to identify factors that have specific relevance to long-term stability.

Another element on which there is high consensus in the literature, but which does not emerge as a critical factor from empirical research is the accountability system [19], [44], [50], [51], [53], [62], [106], [107]. However, the presence of formal agreements on

the activities of DC partners among the critical factors may indirectly address this need.

While among the factors identified in the research there is recognition of the importance of the knowledge and competence of individual actors, particularly with regard to the use of data, among the cases there is no direct expression of the generation of new knowledge and competence at the collective level. Identified in the literature [39], [44], [49], [52], [106] as a key process is that of new knowledge creation, as well as collective training about specific new skills and even the creation of new professional figures and functions, with particular reference to data management. The study recognizes the generation of shared value for partners that acts as an incentive for their participation as a critical factor for sustainability. While it is possible to see a link between the generation of new knowledge and the creation of shared value, the cases show no link between the generation of new competences and the creation of shared value.

Regarding the theoretical dimension of incentives, it should be noted that the literature also identifies other forms of incentives. Gil-Garcia et al (2016) [18] argue for the importance of financial incentives. With respect to the evidence gathered, it is interesting to note that this type of incentive is regarded by some of the cases, who consider it important for future DC developments. These incentives range from simple monetary rewards for the data collected to more original forms such as data provision in exchange for equity. However, no financial incentive mechanism is currently implemented by the DCs analyzed. One of the main reasons lies in the difficulty in assigning a monetary value to the data.

Another incentive identified in the literature is the exclusivity of collaboration [50], [51], i.e., the absence of alternative options for achieving the desired results. There is no empirical evidence of this element in the cases. Alternatively, it has been noted how the participation of prominent actors in non-competitive sectors can attract other actors in the same sector, as if they were conditioned by a bandwagon effect. This incentive, in addition to the fact that it is short-term, was found in only one case.

To conclude, it is appropriate to dwell on the factor on which there is most consensus within the DC literature, namely *data stewardship* [10], [14], [37], [44], [59], [62], [75], [106]. Within the research, this factor was considered by analytically studying its components. From the cases it was possible to identify four components of data stewardship:

- *Data expertise*. The technical ability to conduct data analysis and interpret the results.
- *Data infrastructure*. The set of hardware and software equipment and their connections to be able to transfer, process and store data in a secure way.
- *Data standardization.* The process of modifying the raw data shared by partners for the purpose of making it useful in order to conduct the desired analyses.

- *Data sharing facilitation.* The process of supporting partners in sharing their data by connecting with trusted data infrastructure, while also ensuring compliance with privacy, regulations, and data restriction agreements.

These components have been divided into three different critical factors in accordance with their nature with respect to the definition of governance. The process components form the critical factor *data standardization and sharing facilitation*; the structure component forms the critical factor *innovation, technology, and data infrastructure*; and the competence component falls under the critical factor *competence and data expertise*.

6.6. Implications for practitioners

The results of the study explicitly identified governance factors that can influence the long-term sustainability of projects. The study of these factors leads to clear implications for practice for practitioners from different sectors (private, public non-profit).

The following is non-exhaustive set of suggestions that can support the work of regulators in systematizing the creation of long-term projects:

- stimulate the generation of knowledge networks among members of organizations from different sectors.
- encourage the creation of professionals with competencies in both data analysis and its application to social work.
- create economic incentives that ensure not only the creation but continuity for the development of DCs over time.
- provide certifications or accreditations that make the public endorsement of such initiatives recognizable.
- encourage the development of expertise on data and the acquisition of technological infrastructure for nonprofit organizations.
- make available standard forms of agreements for the management and restriction in the use of private data that protect all parties involved in the DC.
- encourage the development of technology intermediaries who can mediate between organizations and make the use of data safe and effective.
- incentivize the measurement of the social impact generated by these collaborations.

The following is non-exhaustive set of the practical implications for supporting organizations considering creating a new DC or those currently involved in the first phases of a DC:

- make explicit the needs and interests of each actor to participate in the collaboration, build the collective vision and mission based on these and make them manifest.

- solicit the support of public institutions and local authorities to increase trust in new actors.
- declare past experiences in the field and adopt techniques for measuring the impact generated.
- use the presence of innovative technologies as leverage for attracting new partners.
- leverage the social responsibility of private actors to require their participation, in terms of economics and data sharing.
- create an entity dedicated to the project that has a nonprofit legal form and can handle the management of funds and personnel.
- adopt forms of governance that encourage flexibility in collaborating with different actors, such as having a stable core linked by more stringent bond and accompanied by collaborations of a more occasional nature.
- create mechanisms to generate direct value for actors involved in data donation, such as the production of new services for partners through data processing.
- provide sources of revenue related to the use of data that are alternative to public contributions or private donations.
- create moments of discussion with the target beneficiary and partners for effective involvement and transparent communication of activities and results.
- experiment with different business models starting with small projects and adapting them according to environmental responses.
- establish formal agreements governing the actions and responsibilities of each partner with particular reference to data management.

6.7. Limitations and future developments

Among the major difficulties in this research was the availability of case studies. In fact, only a low percentage of the cases contacted were open to cooperate with the research. Nevertheless, the analysis of interviews with members from nine distinct cases represents a satisfactory result and one of the largest empirical contributions to DC literature. Given the high degree of internal inhomogeneity among DCs, however, a larger number of cases could add more robustness to the results.

Data collection was then mainly focused on the testimony of a single member within the collaboration. Only in one case was it possible to analyze joint responses from two organizations within the same DC. This issue was overcome by analysis of documentation from secondary sources, although it was not always possible to access authoritative sources regarding the cases and rarely focused specifically on governance.

Finally, one aspect to note within the sample of cases analyzed is the longevity of the cases. As DCs are an emerging phenomenon, the duration of activity of these case

studies does not exceed a decade. Regardless, cases were selected that had passed the pilot phases and were active and designed to be durable.

Future developments will be possible by continuing to observe the cases analyzed over a longer time frame and adding new case study analyses.

This study may offer several implications for practice. An advance in the research may come from implementing the governance components identified in the establishment of a new DC and from monitoring their consequences.

7 Conclusion

Data Collaboratives are an emerging form of social innovation with high potential impact. However, these initiatives are often occasional or hardly pass the first pilot phase. The literature is developing in recent years but still has limitations. In particular, deficiencies have been found in the knowledge about factors that may influence the long-term sustainability of these projects. Given the critical aspects of managing and coordinating these types of collaborations, research has focused on establishing relationships between the elements that make up collaborative governance and long-term stability.

To do this, the literature on DCs and related fields of study was first reviewed in order to gain a thorough understanding of the specifics of these projects and to identify the organizational aspects that can determine their success and failure. Next, the literature on cross-sector partnerships, a broader field of study with more contributions on the topic of governance, was reviewed. This allowed for a more robust definition of the concept of governance and its components. Specifically, this is the definition of DC governance adopted for the study:

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 these two fields of study, 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 resources to carry out the DC.
6. *Business model.* Governance factors that ensure the correct match between activities to be carried out by the DC and available resources.
7. *Adaptation.* Governance factors that allow DC organizational arrangements to change in response to endogenous or exogenous forces.

A multiple case study was designed with an abductive approach aimed at identifying the key governance factors that define the relationship between the theoretical dimensions and the sustainability of initiatives. The analysis was based on the study of a heterogeneous sample of nine currently active cases that had the characteristics of long-term stability as classified by Bartolomucci et al. (2022). Information was collected by conducting semi-structured interviews based on the theoretical dimensions. To add further empirical input, 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. Analysis of the data collected identified twenty critical governance factors that can contribute to the systematization of these types of collaborations.

An aspect to consider in relation to the theoretical dimensions are their partial dependency relationships. The dimensions, although relevant in different aspects, have some points of contact. This became particularly evident during the analysis of the data, where some information gathered for one dimension was more relevant for another dimension. Below are some of the relationships identified between the dimensions.

- The initiation of the DC depends on the mutual trust present between the actors during the preliminary establishment phases. Initiation also depends on the interests of each actor to participate, so it is also linked to the incentive system that is used to satisfy individual and collective interests.
- Trust may depend on specific constraints between the responsibilities of actors that are specified within formal agreements. Trust relationships between actors may be catalysed by the presence of an intermediary acting as a boundary spanning organisation.
- The collaborative business model, which depends on the sharing of key resources and internal fund-raising within the collaboration, is influenced by the ability to incentivise key partners to participate. If there are breaks in the relationship with some of the members, the DC has to be able to adapt its structure so that it can continue with the same business model or modify it according to the new resources.

- The adaptability of collaboration is linked to the level of trust between existing and new actors. Adaptability also depends on the level of interest in existing and new entrants that the DC is able to generate and maintain over time.

The results of the study revealed processes, structures and governance elements lying at their intersection that can determine the long-term continuity of these collaborations. To answer the research questions, the study identified seven categories of critical governance factors that influence long-term stability of the DC and that distinguish it from other kinds of cross-sector social partnerships:

- *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 one notes the action by public agencies in providing funds and regulating 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 formal 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.

The results of the study explicitly identified governance factors that influence the long-term sustainability of DC projects and can offer concrete insights to practitioners. The hope is that this comprehensive framework may facilitate the emergence and development of new projects that can be durable and generate high value for the society.

References

[1] IEAG, "A World that Counts," 2014. [Online]. Available: www.undatarevolution.org

[2] P. Taylor, "Volume of data/information created, captured, copied, and consumed worldwide from 2010 to 2020, with forecasts from 2021 to 2025," 2022. https://www.statista.com/statistics/871513/worldwide-data-created/ (accessed Apr. 03, 2023).

[3] D. Blazquez and J. Domenech, "Big Data sources and methods for social and economic analyses," *Technol Forecast Soc Change*, vol. 130, 2018, doi: 10.1016/j.techfore.2017.07.027.

[4] A. Rasche, M. Morsing, and E. Wetter, "Assessing the Legitimacy of 'Open' and 'Closed' Data Partnerships for Sustainable Development," *Bus Soc*, vol. 60, no. 3, 2021, doi: 10.1177/0007650319825876.

[5] J. Mercille, "Inclusive smart cities: Beyond voluntary corporate data sharing," *Sustainability (Switzerland)*, vol. 13, no. 15, 2021, doi: 10.3390/su13158135.

[6] S. Manning, U. Boston, and J. Reinecke, "We're failing to solve the world's 'wicked problems.' Here's a better approach," *The Conversation*, 2016. Accessed: Apr. 03, 2023. [Online]. Available: https://theconversation.com/were-failing-to-solve-the-worlds-wicked-problems-heres-a-better-approach-64949

[7] United Nations, "Agenda 2030." https://unric.org/en/sdg-17/ (accessed Apr. 03, 2023).

[8] I. Susha, "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.

[9] G. Santos-Hermosa, A. Quarati, E. Loría-Soriano, and J. E. Raffaghelli, "Why Does Open Data Get Underused? A Focus on the Role of (Open) Data Literacy," 2023, pp. 145–177. doi: 10.1007/978-3-031-24193-2_6.

[10] Digital Public Goods Alliance, Global Partnership for Sustainable Development Data, Jain Family Institute, UN Global Pulse, and UNICEF, "Exploring Data as and in Service of the Public Good," 2023.

[11] R. Máchová and M. Lněnička, "Evaluating the quality of open data portals on the national level," *Journal of Theoretical and Applied Electronic Commerce Research*, vol. 12, no. 1, 2017, doi: 10.4067/S0718-18762017000100003.

[12] I. Susha, Å. Grönlund, and R. Van Tulder, "Data driven social partnerships: Exploring an emergent trend in search of research challenges and questions," *Gov Inf Q*, vol. 36, no. 1, 2019, doi: 10.1016/j.giq.2018.11.002.

[13] S. Verhulst, D. Sangokoya, and T. Govlab, "Data Collaboratives: Exchanging Data to Improve People's Lives," 2015.

[14] S. G. Verhulst, A. Young, M. Winowatan, and A. J. Zahuranec, "LEVERAGING PRIVATE DATA FOR PUBLIC GOOD A Descriptive Analysis and Typology of Existing Practices," 2019.

[15] I. Susha, 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.

[16] I. Susha, M. Janssen, and S. Verhulst, "Data collaboratives as 'bazaars'?: A review of coordination problems and mechanisms to match demand for data

with supply," *Transforming Government: People, Process and Policy*, vol. 11, no. 1, 2017, doi: 10.1108/TG-01-2017-0007.

[17] GSM Association, "Scaling Big Data for Social Good: The need for sustainable business models," 2018. [Online]. Available: www.gsma.com/betterfuture/bd4sg

[18] J. R. Gil-Garcia and D. S. Sayogo, "Government inter-organizational information sharing initiatives: Understanding the main determinants of success," *Gov Inf Q*, vol. 33, no. 3, 2016, doi: 10.1016/j.giq.2016.01.006.

[19] S. Stalla-Bourdillon, G. Thuermer, J. Walker, L. Carmichael, and E. Simperl, "Data protection by design: Building the foundations of trustworthy data sharing," *Data Policy*, vol. 2, 2020, doi: 10.1017/dap.2020.1.

[20] 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.

[21] 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.

[22] "Wikitionary." https://en.wiktionary.org/wiki/collaborative (accessed Apr. 01, 2023).

[23] M. Savona, "Governance of Data Value," 2020.

[24] I. Susha, S. Verhulst, M. Janssen, and T. Pardo, "Data Collaboratives: How to create value from data for public problem solving? Panel," in *ACM International Conference Proceeding Series*, 2017. doi: 10.1145/3085228.3085309.

[25] B. H. Hall, A. N. Link, and J. T. Scott, "Barriers inhibiting industry from partnering with universities: Evidence from the Advanced Technology Program," *Journal of Technology Transfer*, vol. 26, no. 1–2, 2001, doi: 10.1023/a:1007888312792.

[26] M. Perkmann and H. Schildt, "Open data partnerships between firms and universities: The role of boundary organizations," in *Research Policy*, 2015. doi: 10.1016/j.respol.2014.12.006.

[27] M. Janssen and A. Zuiderwijk, "Infomediary Business Models for Connecting Open Data Providers and Users," *Soc Sci Comput Rev*, vol. 32, no. 5, 2014, doi: 10.1177/0894439314525902.

[28] I. Susha, B. Rukanova, J. Ramon Gil-Garcia, Y. H. Tan, and M. G. Hernandez, "Identifying mechanisms for achieving voluntary data sharing in cross-sector partnerships for public good," in *ACM International Conference Proceeding Series*, 2019. doi: 10.1145/3325112.3325265.

[29] B. S. Noveck, "Data Collaboratives: Sharing Public Data in Private Hands for Social Good," *Forbes*, 2015. https://www.forbes.com/sites/bethsimonenoveck/2015/09/24/private-data-sharing-for-public-good/?sh=7e37eb3251cd (accessed Mar. 28, 2023).

[30] H. Chafetz, A. J. Zahuranec, S. Marcucci, B. Davletov, and S. Verhulst, "The #Data4COVID19 Review Assessing the Use of Non-Traditional Data During a Pandemic Crisis Assessing the Use of Non-Traditional Data During a Pandemic Crisis Authors Acknowledgements," 2022.

[31] N. Martin, "How Tech Companies Can Advance Data Science for Social Good," 2020. https://ssir.org/articles/entry/how_tech_companies_can_advance_data_science_for_social_good (accessed Mar. 28, 2023).

[32] M. Micheli, M. Ponti, M. Craglia, and A. Berti Suman, "Emerging models of data governance in the age of datafication," *Big Data Soc*, vol. 7, no. 2, 2020, doi: 10.1177/2053951720948087.

[33] M. Le Pennec and E. Raufflet, "Value Creation in Inter-Organizational Collaboration: An Empirical Study," *Journal of Business Ethics*, vol. 148, no. 4, 2018, doi: 10.1007/s10551-015-3012-7.

[34] N. Robin, T. Klein, and J. Jutting, "Public-Private Partnerships for Statistics : Lessons Learned , Future Steps : A focus on the use of non-official data sources for national statistics," *OECD Development Co-Operation*, no. February, 2016.

[35] S. G. Verhulst and A. Young, "THE POTENTIAL OF SOCIAL MEDIA INTELLIGENCE TO IMPROVE PEOPLE'S LIVES Social Media Data for Good," 2017. [Online]. Available: https://ssrn.com/abstract=3141457Electroniccopyavailableat:https://ssrn.com/abstract=3141457Electroniccopyavailableat:https://ssrn.com/abstract=3141457

[36] S. Xuan *et al.*, "An incentive mechanism for data sharing based on blockchain with smart contracts," *Computers and Electrical Engineering*, vol. 83, 2020, doi: 10.1016/j.compeleceng.2020.106587.

[37] I. Susha 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.

[38] M. Gutierrez, A. Daniels, and G. Jobbins, "Briefing note Fishing for data The role of private data platforms in addressing illegal, unreported and unregulated fishing and overfishing," 2018. Accessed: Mar. 28, 2023. [Online]. Available: https://odi.org/en/publications/fishing-for-data-the-role-of-private-data-platforms-in-addressing-illegal-unreported-and-unregulated-fishing-and-overfishing/

[39] J. Farmer, A. McCosker, K. Albury, and A. Aryani, *Data for Social Good*. Singapore: Springer Nature Singapore, 2023. doi: 10.1007/978-981-19-5554-9.

[40] Lisa. Lalande and Joanne. Cave, *Measuring Outcomes in Practice : Fostering an Enabling Environment for Measurement in Canada*. 2018. Accessed: Apr. 01, 2023. [Online]. Available: https://mowatcentre.munkschool.utoronto.ca/

[41] B. Ann Goggins Gregory and D. Howard, "The Nonprofit Starvation Cycle," 2009. [Online]. Available: www.ssireview.com

[42] Salesforce, "Nonprofit Trends Report," 2020.

[43] M. Chui *et al.*, "NOTES FROM THE AI FRONTIER -APPLYING AI FOR SOCIAL GOOD," 2018.

[44] Global Partnership for Sustainable Development Data, "Effective and Ethical Data Sharing at Scale A cookbook for data producers, donors, policymakers, and other development practitioners With support from Google.org," 2023.

[45] M. Brack and T. Castillo, "Data Sharing for Public Health - Key Lessons from Other Sectors," 2015.

[46] W. Hoffman, A. Boral, and D. Olukoya, "Insight Report Data Collaboration for the Common Good Enabling Trust and Innovation Through Public-Private Partnerships," 2019. [Online]. Available: www.weforum.org

[47] J. Mcmurren and S. G. Verhulst, "Data to Go: The Value of Data Portability as a Means to Data Liquidity," 2020.

[48] F. Bartolomucci and G. Bresolin, "Data Collaboratives' archetypes definition and research priorities setting towards their long-term sustainability," 2022.

[49] I. Susha, T. van den Broek, A.-F. van Veenstra, and J. Linåker, "An ecosystem perspective on developing data collaboratives for addressing societal issues: The role of conveners," *Gov Inf Q*, vol. 40, no. 1, p. 101763, Jan. 2023, doi: 10.1016/j.giq.2022.101763.

[50] C. Ansell and A. Gash, "Collaborative governance in theory and practice," *Journal of Public Administration Research and Theory*, vol. 18, no. 4, 2008, doi: 10.1093/jopart/mum032.

[51] 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.

[52] K. Emerson, T. Nabatchi, and S. Balogh, "An integrative framework for collaborative governance," *Journal of Public Administration Research and Theory*, vol. 22, no. 1, 2012, doi: 10.1093/jopart/mur011.

[53] J. M. Bryson, B. C. Crosby, and M. Middleton Stone, "Design and Implementation of Cross-Sector Collaboration 45 Figure 1 A Framework for

Understanding Cross-Sector Collaborations," *Public Adm Rev*, no. December, 2006.

[54] A. M. Thomson and J. L. Perry, "Collaboration processes: Inside the black box," *Public Administration Review*, vol. 66, no. SUPPL. 1. 2006. doi: 10.1111/j.1540-6210.2006.00663.x.

[55] R. Agranoff, "Inside collaborative networks: Ten lessons for public managers," *Public Administration Review*, vol. 66, no. SUPPL. 1. 2006. doi: 10.1111/j.1540-6210.2006.00666.x.

[56] K. G. Provan and P. Kenis, "Modes of network governance: Structure, management, and effectiveness," *Journal of Public Administration Research and Theory*, vol. 18, no. 2, 2008, doi: 10.1093/jopart/mum015.

[57] M. A. Koschmann, T. R. Kuhn, and M. D. Pfarrer, "A communicative framework of value in cross-sector partnerships," *Academy of Management Review*, vol. 37, no. 3. 2012. doi: 10.5465/amr.2010.0314.

[58] S. Vangen and N. Winchester, "Managing Cultural Diversity in Collaborations: A focus on management tensions," *Public Management Review*, vol. 16, no. 5, 2014, doi: 10.1080/14719037.2012.743579.

[59] S. Stalla-Bourdillon, L. Carmichael, and A. Wintour, "Fostering trustworthy data sharing: Establishing data foundations in practice," *Data Policy*, vol. 3, 2021, doi: 10.1017/dap.2020.24.

[60] B. Crane, "Revisiting Who, When, and Why Stakeholders Matter: Trust and Stakeholder Connectedness," *Bus Soc*, vol. 59, no. 2, 2020, doi: 10.1177/0007650318756983.

[61] S. S. Dawes, A. M. Cresswell, and T. A. Pardo, "From 'need to know' to 'need to share': Tangled problems, information boundaries, and the building of public sector knowledge networks," *Public Adm Rev*, vol. 69, no. 3, 2009, doi: 10.1111/j.1540-6210.2009.01987_2.x.

[62] W. van Donge, N. Bharosa, and M. F. W. H. A. Janssen, "Data-driven government: Cross-case comparison of data stewardship in data ecosystems," *Gov Inf Q*, vol. 39, no. 2, 2022, doi: 10.1016/j.giq.2021.101642.

[63] I. Susha, M. Janssen, and S. Verhulst, "Data collaboratives as 'bazaars'?: A review of coordination problems and mechanisms to match demand for data with supply," *Transforming Government: People, Process and Policy*, vol. 11, no. 1, 2017, doi: 10.1108/TG-01-2017-0007.

[64] B. Carballa Smichowski, "Alternative Data Governance Models: Moving Beyond One-Size-Fits-All Solutions," *Intereconomics*, vol. 54, no. 4, 2019, doi: 10.1007/s10272-019-0828-x.

[65] N. Bharosa, "The rise of GovTech: Trojan horse or blessing in disguise? A research agenda," *Gov Inf Q*, vol. 39, no. 3, p. 101692, Jul. 2022, doi: 10.1016/j.giq.2022.101692.

[66] S. B. Page, J. M. Bryson, B. C. Crosby, D. Seo, and M. M. Stone, "Ambidexterity in Cross-Sector Collaborations Involving Public Organizations," *Public Performance & Management Review*, vol. 44, no. 6, pp. 1161–1190, Nov. 2021, doi: 10.1080/15309576.2021.1937243.

[67] "Basics of Qualitative Research: Grounded Theory Procedures and Techniques," *The Modern Language Journal*, vol. 77, no. 2, 1993, doi: 10.2307/328955.

[68] 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.

[69] GovLab, "The #Data4COVID19 Review: Learnings From COVID-19 and Recommendations For Future Crises," 2022.

[70] L. S. Whiting, "Semi-structured interviews: guidance for novice researchers.," *Nurs Stand*, vol. 22, no. 23, 2008, doi: 10.7748/ns2008.02.22.23.35.c6420.

[71] C. Dearnley, "A reflection on the use of semi-structured interviews.," *Nurse researcher*, vol. 13, no. 1. 2005. doi: 10.7748/nr2005.07.13.1.19.c5997.

[72] J. A. Hatch, *Doing qualitative research in education settings.* . Suny Press, 2002.

[73] D. A. Gioia, K. N. Price, A. L. Hamilton, and J. B. Thomas, "Forging an identity: An insider-outsider study of processes involved in the formation of organizational identity," *Adm Sci Q*, vol. 55, no. 1, 2010, doi: 10.2189/asqu.2010.55.1.1.

[74] N. Zingales, "Data Collaboratives, Competition Law and the Governance of EU Data Spaces," *SSRN Electronic Journal*, 2021, doi: 10.2139/ssrn.3897051.

[75] I. Susha, M. Flipsen, W. Agahari, and M. de Reuver, "Towards Generic Business Models of Intermediaries in Data Collaboratives: From Gatekeeping to Data Control," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2020. doi: 10.1007/978-3-030-57599-1_23.

[76] GovLab, "The #Data4COVID19 Review:What is non-traditional data? How was it used during COVID-19?," 2022.

[77] M. Tiru, E. Saluveer, R. Ahas, and A. Aasa, "The positium Barometer: A web-based tool for monitoring the mobility of tourists," *Journal of Urban Technology*, vol. 17, no. 1, 2010, doi: 10.1080/10630731003597348.

[78] R. Ahas, M. Tiru, E. Saluveer, and C. Demunter, "Mobile telephones and mobile positioning data as source for statistics : Estonian experiences," in *Presentation for NTTS*, 2011.

[79] R. Ahas, A. Aasa, S. Silm, and M. Tiru, "Mobile Positioning Data in Tourism Studies and Monitoring: Case Study in Tartu, Estonia," in *Information and Communication Technologies in Tourism 2007*, 2007. doi: 10.1007/978-3-211-69566-1_12.

[80] S. Esko and E. Saluveer, "Crunching big data to produce statistics on large international events," 2015.

[81] E. Saluveer *et al.*, "Methodological framework for producing national tourism statistics from mobile positioning data," *Ann Tour Res*, vol. 81, 2020, doi: 10.1016/j.annals.2020.102895.

[82] FIWARE Foundation, "FIGHTING CLIMATE CHANGE WITH FIWARE," 2019.

[83] Civity B.V., "SnifferBike," 2022. [Online]. Available: www.civity.nl

[84] Civity B.V., "Smart reporting," 2022.

[85] Civity B.V., "Data Platform," 2022. [Online]. Available: www.civity.nl

[86] I. Calzada, "Data co-operatives through data sovereignty," *Smart Cities*, vol. 4, no. 3, 2021, doi: 10.3390/smartcities4030062.

[87] SalusCoop and IdeasForChange, "TRIEM," 2018.

[88] N. Vij, "Opinion piece introducing the Consumer Data Research Centre (CDRC)," *Journal of Direct, Data and Digital Marketing Practice*, vol. 17, no. 4. 2016. doi: 10.1057/s41263-016-0007-8.

[89] Andrzej. Klimczuk, *The collaborative economy in action European perspectives*. University of Limerick, 2021.

[90] A. H. Dawson, "An Update on Data Governance for Sidewalk Toronto," *Medium*, 2018. [Online]. Available: https://medium.com/sidewalk

[91] B. Wylie, "Searching for the Smart City's Democratic Future," 2018.

[92] B. Wylie, "Debrief on Sidewalk Toronto Public Meeting #2 — Time to Start Over, Extend the Process," *Medium*, 2018.

[93] Waterfront Toronto, "New District in Toronto Will Tackle the Challenges of Urban Growth," 2017. [Online]. Available: www.sidewalktoronto.ca.

[94] B. Barth, "Opinion The fight against Google's smart city," *The Washington Post*, 2018.

[95] "PLAN DEVELOPMENT AGREEMENT BETWEEN TORONTO WATERFRONT REVITALIZATION CORPORATION AND SIDEWALK LABS LLC." 2018.

[96] Sidewalk Toronto, "Sidewalk Toronto Public Engagement Plan," 2018.

[97] D. L. Doctoroff, "Reimagining cities from the internet up," *Medium*, 2016. [Online]. Available: https://medium.com/sidewalk-talk/reimagining-cities-from-the-internet-up-5923d6be63ba

[98] A. Carter and J. Rieti, "Sidewalk Labs cancels plan to build high-tech neighbourhood in Toronto amid COVID-19," *CBC News*, 2020.

[99] Sidewalks Labs, "Project Vision," 2017.

[100] V. Donovan, "Waterfront Toronto advisory panelstill has concerns about Sidewalk Labs' data collection, new reportsays," *TORONTO STAR*, 2020. [Online]. Available: https://www.thestar.com/news/gta/2020/02/26/waterfront-toronto-advisory-panel-still-has-concerns-about-sidewalk-labs-data-collection-new-report

[101] R. Johnson, "Waterfront Toronto won't sell user data, new planning document says," *STATESCOOP*, 2019.

[102] B. Lingard, "The Global Education Industry, Data Infrastructures, and the Restructuring of Government School Systems," in *Researching the Global Education Industry: Commodification, the Market and Business Involvement*, 2019. doi: 10.1007/978-3-030-04236-3_7.

[103] M. Levine and A. Levine, "Charters and foundations: Are we losing control of our public schools?," *American Journal of Orthopsychiatry*, vol. 84, no. 1, 2014, doi: 10.1037/h0098942.

[104] M. Bulger, P. McCormick, and M. Pitcan, "The Legacy of InBloom," *Data and Society*, 2017.

[105] B. J. Evans and H. M. Krumholz, "People-powered data collaboratives: Fueling data science with the health-related experiences of individuals," *Journal of the American Medical Informatics Association*, vol. 26, no. 2. 2019. doi: 10.1093/jamia/ocy159.

[106] S. Marcucci, N. G. Alarcón, S. G. Verhulst, and E. Wüllhorst, "Mapping and Comparing Data Governance Frameworks A benchmarking exercise to inform global data governance deliberations," 2023.

[107] M. Lockwood, "Good governance for terrestrial protected areas: A framework, principles and performance outcomes," *J Environ Manage*, vol. 91, no. 3, 2010, doi: 10.1016/j.jenvman.2009.10.005.

[108] X. Yan, H. Lin, and A. Clarke, "Cross-Sector Social Partnerships for Social Change: The Roles of Non-Governmental Organizations," *Sustainability*, vol. 10, no. 2, p. 558, Feb. 2018, doi: 10.3390/su10020558.

[109] M. Levi and B. Rajala, "Alternatives to Social Science One," *PS Polit Sci Polit*, vol. 53, no. 4, 2020, doi: 10.1017/S1049096520000438.

List of Figures

List if Tables

Acknowledgments

First of all, I would like to thank my family for their support over the years and for helping and backing me in the most important decisions.

I would like to thank Emma for always standing by me, for believing in my abilities and pushing me to do the same.

Despite the conditions that partly hindered this experience, I would like to thank all the people who have been close to me during the time away from home, starting with all the flatmates, passing through fellow students and those who helped me find moments of leisure.

Finally, I would like to thank Dr. Bartolomucci for his expertise and for allowing me to get passionate about this topic.