Establishing and implementing data collaborations for public good
A critical factor analysis to scale up the practice

Susha, Iryna

DOI
10.3233/IP-180117

Publication date
2020

Document Version
Final published version

Published in
Information Polity

Information Polity 25 (2020) 3–24
DOI 10.3233/IP-180117
IOS Press

Establishing and implementing data collaborations for public good: A critical factor analysis to scale up the practice

Iryna Susha[a,b]
[a]*Department of Informatics, School of Business, Örebro University, Sweden*
[b]*Section ICT, Faculty of Technology, Policy and Management, Delft University of Technology, The Netherlands*
E-mail: iryna.susha@oru.se

Abstract. Data analytics for public good has become a hot topic thanks to the inviting opportunities to utilize 'new' sources of data, such as social media insights, call detail records, satellite imagery etc. These data are sometimes shared by the private sector as part of corporate social responsibility, especially in situations of urgency, such as in case of a natural disaster. Such partnerships can be termed as 'data collaboratives'. While experimentation grows, little is known about how such collaborations are formed and implemented. In this paper, we investigate the factors which are influential and contribute to a successful data collaborative using the Critical Success Factor (CSF) approach. As a result, we propose (1) a framework of CSFs which provides a holistic view of elements coming into play when a data collaborative is formed and (2) a list of Top 15 factors which highlights the elements which typically have a greater influence over the success of the partnership. We validated our findings in two case studies and discussed three broad factors which were found to be critical for the formation of data collaboratives: value proposition, trust, and public pressure. Our results can be used to help organizations prioritize and distribute resources accordingly when engaging in a data collaborative.

Keywords: Critical success factors, data driven social partnership, data sharing, cross sector partnership, data innovation, inter-organizational collaboration

Key points for practitioners:

- Formation and implementation of data collaboratives requires consideration of organizational, technological, environmental, and legislation and policy factors.
- This study identifies Top 15 most critical factors for data collaboratives based on review of research and expert evaluation.
- Three broad factors emerge from case studies which emphasize the criticality of value proposition, trust, and public pressure for the formation of data collaboratives.

1. Introduction

The data revolution has catalyzed a number of innovations, including the way topical societal issues can be detected and addressed. Data analytics for public good has become a hot topic thanks to the inviting opportunities to utilize 'new' sources of data, such as social media insights, call detail records, satellite imagery etc. Increasingly, official data collected by governments is being complemented by and combined with data from the private sector, NGOs and individuals (ODI, 2013). Thus, governments are faced with a new task and role of capturing insights generated outside of public sector data ecosystem and acting upon them to their service to citizens. This also requires enhanced collaboration with actors outside of the public domain and across different levels of government. Many of the most complex public problems, such as climate change, overpopulation, disease outbreaks, are beyond the sphere of influence of any single government. Data exchange and collaboration between companies, governments, and other actors remain difficult because of legal barriers, silos, proprietary nature of data, fears and risks of misuse (Lisbon Council, 2017).

Recently the term 'data collaborative' (Verhulst & Sangokoya, 2015) emerged which stands for *a form of cross sector partnership in which participants – companies, government agencies, researchers, non-profits – collaborate to leverage data collection, sharing, or use for addressing societal problems* (Susha et al., 2017). Data collaboratives have proliferated recently in the world and can be found in different domains, such as environment, public health, disaster response, education, poverty etc. A repository of examples of data collaboratives[1] was created which showcases initiatives around the world. A notable example is the 2017 initiative of the UN Global Pulse called Data for Climate Action which invited a number of companies to share their data with teams of researchers in the framework of a challenge competition to help address topical research questions about climate change.

While experimentation grows, little is known about how such collaborations are formed and implemented. There are examples of data collaboratives which were not as successful as hoped at first. For instance, inBloom, a non-profit that offered a data warehouse solution designed to help public schools in the US embrace the promise of personalized learning by helping teachers integrate seamlessly the number of applications they use in their day-to-day teaching, collapsed and has ceased to exist, as privacy concerns from interested parties mounted over a period of many months (Forbes, 2014). Many initiatives do not get to this point because there are a number of barriers to the formation of data collaboratives. Private sector organizations are often unwilling to share data which otherwise can be used to their competitive advantage. Especially in situations when personal data is exchanged, security, privacy, and trust issues come into play. Also, it is often unclear which data is needed to address a certain societal issue which complicates the search for partners. Therefore, learning from practice and systematizing the evidence of what works is very important for advancing and scaling up data collaborative efforts.

In this paper, we investigate the factors which are influential and contribute to a successful data collaborative. The research question of our study thus reads: *What are the factors critical for the formation and implementation of data collaboratives?* Our study aims to make a practical contribution by developing a comprehensive overview of critical factors from research and practice that can be used as guidelines by the practitioner community. The research contribution of our study is that there is very limited research interrogating the novelty of this phenomenon and what it adds to our current knowledge on interorganizational collaboration and data sharing. Identifying and discussing critical factors will enable us to reflect on the nuances of implementing data collaboratives compared to more 'traditional' forms of partnerships.

The paper is structured as follows: in Section 2 we discuss the background literature regarding successful implementation of data collaboratives; in Section 3 we present the method we used to conduct our study; in Sections 4 and 5 we report on our results; we conclude our study with recommendations in Section 6.

2. Successful data collaboratives: Literature review

Overall, conceptualizing success is challenging because different stakeholders may have a different assessment of the project depending on their goals. Also, it matters at which point in time such an as-

[1] DataCollaboratives.org.

sessment is made. The level of success of a collaboration can be evaluated according to two dimensions: (1) the achievement of the expected outcomes as a result of the collaboration and (2) the level of satisfaction of the partners with the achievement of these outcomes (Barroso-Mendez et al., 2016). Thus, a successful partnership includes both dimensions (Ibid.).

Academic research focusing specifically on the phenomenon of data collaboratives is scarce but emergent. Data collaboratives can be viewed as both a novel form of partnership with distinct characteristics and a new promising concept for shedding new light on existing collaborations. Compared to other more 'traditional' collaborations around information exchange, data collaboratives are a product of the data revolution with all its affordances and dangers (Susha et al., 2019). They emerge in a completely novel environment, may involve different mix of actors (with private sector growing in importance when it comes to data), and may pose novel challenges, such as data ethics and privacy.

In terms of critical factors, traditional forms of partnering may offer a sound foundation for learning from best practice but are not able to fully explain and address these new realities. Previous research on cross-sector (public-private) partnerships and inter-organizational collaboration in general does not explicitly focus on the context of the data revolution. Comparable literature reviews on cross sector partnerships (e.g. Gray & Stites, 2013; Van Tulder et al., 2016) have so far not revealed any relevant studies on the phenomenon of data collaboratives. When it comes to interorganizational information sharing (IIS) research, there are distinct differences as well: in contrast to IIS which has focused on government-to-government sharing, data collaboratives attempt to use not only public but especially previously closed private data to address important social problems. They do not integrate all their data in a permanent system but take advantage of the availability of diverse and complementary public and private data to better understand a specific problem and propose a solution (Susha & Gil-Garcia, 2019).

Having said that, data collaboratives can also be viewed as a novel concept that can shed new light on past partnerships (Klievink et al., 2018). These authors proposed a conceptual model of value creation through data collaboratives and rooted the model in the work of Ansell and Gash (2008) on collaborative governance. Klievink et al.'s model suggests that the formation of data collaboratives is influenced by four types of antecedens – collaborative, information systems antecedents, prior experience and culture of data, and "rules of the game". Furthermore, according to this model, there is a positive feedback loop between collaboration and trust that leads to the institutionalization of trust which is required for successful data collaboratives (Ibid.). We find this model very helpful for our study and include its key elements in our integrative framework presented in Section 2.3.

Susha et al. (2019) reviewed literature on this emergent topic and provisionally elicited 13 success factors from previous studies. These factors concern motivations, organizational structure, stakeholder involvement, communication, roles and responsibilities, leadership, resources, political context, to name a few. All these factors were derived from cases in certain application domains and their generalizability remains to be tested. Other research (Van den Broek & Van Veenstra, 2018) looked more closely into the governance arrangements possible for data collaboratives and how the type of data shared influences the type of governance that is appropriate.

In line with this, we found that several contributions provide recommendations and guidelines for initiating and implementing public-private social partnerships around corporate data sharing. The important factors highlighted concerned: defining the problem, identifying data gaps and finding the right data, and assembling the right expertise (UNDP & UN Global Pulse, 2016). This shows that specific new professional roles need to be introduced in such partnerships, such as those of data scientist, data engineer, data visualizations expert, domain expert, and data privacy expert. In addition to this, the resource by The Gov Lab (2018) highlights the importance of several other steps, such as: defining value propositions and incentives for participants; developing a risk mitigation strategy; establishing a governance structure and agreeing on terms and conditions; defining an evaluation approach for impact assessment, to name a few.

There is limited academic literature on the topic therefore we turn to more established fields of research as well to draw relevant insights. By doing so, we follow the example of Gottschalk and Solli-Sæther (2005) who in their study of critical factors in IT outsourcing first reviewed relevant theories and then elicited critical factors from these theories. The phenomenon of data collaboratives spans several topics and fields of research. It is a socio-technical phenomenon combining 'soft' elements related to how actors collaborate and 'hard' elements such as data related activities which the collaboration focuses on. Therefore, in our view, the concept of data collaboratives builds on two main literature streams – *cross sector partnerships* and *cross boundary information sharing and integration*. These two streams conceptualize success in slightly different terms which we examine below. Both literatures offer a rich picture when it comes to factors contributing to successful collaborations or information sharing projects. In the following sections we review key points from the two literatures with the aim of combining them into an integrated framework that we can use to structure our analysis of critical factors for data collaboratives.

2.1. Cross sector partnerships perspective

The first relevant literature stream to review is *cross sector (social) partnerships (CCSP) research*. Cross sector social partnerships are "cross-sector projects formed explicitly to address social issues and causes that actively engage the partners on an ongoing basis" (Selsky & Parker, 2005, p.850). CCSPs occur in four 'arenas': business-nonprofit, business-government, government-nonprofit, and trisector (Ibid.).

There is a number of authoritative holistic frameworks which intend to capture the complexity of social partnerships. As summarized by Bryson et al. (2015), a common theme across these frameworks is their attention to the influence of starting conditions on the success of partnerships. We therefore include this element into our integrative framework. These authors further synthesized the knowledge about drivers and initial conditions of cross sector collaborations and highlighted the importance of the following factors: leadership by 'champions', (formal) agreement on the problem definition and the mission of the collaboration, prior relationship and existing networks, incentives to collaborate. These authors further summarized the issues important to collaboration processes, such as a trusting relationship, communication, legitimacy, collaborative planning of goals, governance, collaborative competences being the main ones. The authors however warn against creating 'recipes' for successful partnerships and see the role of research in this area as a design guidance.

Partnerships and collaboration research also offer a detailed view of the process and dynamics of collaboration between the partners. The framework of Emerson, Nabatchi and Balogh (2012) synthesized key ideas from previous work and conceptualized the 'inner workings' of collaboration through several constructs: principled engagement of actors, shared motivation, and capacity for joint action. This element shows how the process and the collaboration experience influence success of partnerships. We include this element in our integrative framework.

Several more specific frameworks exist which are applicable to certain kinds of cross sector collaborations. For instance, Barroso-Mendez et al. (2016) propose a model of success for social partnerships between businesses and non-profits which includes such constructs as shared values, trust, commitment, relationship learning, and cooperation. Hartman and Dhanda (2018) further propose three broad success factors of CCSPs: selecting compatible partners; defining and communicating clear and well-informed collaboration goals; and monitoring and measuring the impact of collaboration. In general, the yearly Corporate-NGO Partnerships Barometer finds that 90% of corporates in the UK think partnerships with NGOs with become more important in the near future (C&E, 2017).

Our review can also benefit from taking a broader perspective on partnerships and including relevant knowledge about public-private partnerships (PPP) and industry-academia collaborations. A systematic literature review by Osei-Kyei and Chan (2015) concluded that the top five factors in successful PPPs are: appropriate risk allocation and sharing, strong private consortium, political support, community/public support, and transparent procurement. In industry-academia collaborations diverse factors are mentioned as well and they can be grouped into organizational, contextual, and process factors (Thune, 2011).

None of the CCSP frameworks delves very deeply into the effects of the broader technical and institutional environments on collaboration (Bryson et al., 2015). Therefore, we will consider information sharing and integration literature to fill in this gap.

2.2. *Cross boundary information sharing and integration perspective*

The second relevant stream of research is the literature on interorganizational information sharing in the context of digital government domain. Several concepts are of particular interest here, such as cross boundary information sharing (CBIS) and interagency information sharing (IIS). Dawes (1996) proposed a theoretical model of interagency information sharing which outlined how expectations, information sharing experiences, and institutional/policy support influence each other. This study put forward two policy principles to stimulate information sharing between government organizations: information stewardship and information use (Ibid.). The work of Estevez et al. (2010) built on the Dawes' model and conceptualized information sharing as consisting of four 'packages': information sharing in government, sharing experience, infrastructure support, and information strategy.

Previous studies of interorganizational information sharing proposed several organizing frameworks to structure the different perspectives or areas. Dawes' (1996) research in interagency information sharing and Zhang et al.'s (2005) research in e-government knowledge sharing both define and view influential factors from three primary perspectives: technology, management, and policy. Yang and Wu (2014) illustrated the complexity of cross boundary information sharing via four groups of factors: organizational, technological, legislation and policy, and environment. Similarly, four perspectives were suggested in the work of Yang and Maxwell (2011) on interorganizational information sharing: technological, organizational and managerial, and political and policy perspectives. Five different 'layers' of interorganizational information sharing were proposed by Bigdeli et al. (2013): technological layer, organizational layer, business process layer, barrier/benefit/risk layer, and environmental layer. The aforesaid frameworks highlight the importance of multidimensional view going beyond just organizational and technological aspects. We therefore include this element in our integrative framework.

Several studies focused on specific factors, such as e.g. the importance of context in (transnational) knowledge and information exchange networks (Dawes et al., 2012). This study identified three layers of context – national, organizational, and informational – which create a number of 'distances' between participating organizations, such as cultural, political, resource, technical, intention etc. These distances may contribute differently towards the success of the exchange between the participants. Another study (Sayogo et al., 2018) showed that clarity of roles and responsibilities contributes to IIS project success and evaluated the determinants thereof. Research also found that perceived impediments could significantly detract from expected benefits in government information sharing projects, therefore it is recommended that project managers set clear goals, keep expectations real, and avoid control-oriented management style (Gil-Garcia, Chengalur-Smith & Duchessi, 2007). Studies also investigated the criticality of factors influencing the success of interorganizational information sharing and found compatibility of technical infrastructure and formally assigned project managers to be the most important factors (Gil-Garcia & Sayogo, 2016).

When it comes to CBIS literature, most research focused on the public sector, however, the discussion in this field moves towards putting in the spotlight the public-private relationship (Sutherland et al., 2018). Current private sector focused studies of CBIS focus mainly on the context of supply chain. These studies find that such factors as trust, interoperability, technical infrastructure, data standards and accuracy, timely communications are important for successful information sharing for supply chain (Ebrahim-Khanjari et al., 2012; Lee and Whang, 2000). More broadly, according to Yang and Maxwell's model (2011), information sharing among organizations is influenced directly by legislation and policies and indirectly by trust, lack of resource, concerns of information misuse, organizational boundaries of bureaucracy, different procedures, control mechanisms, and work flows between organizations. Sayogo and Pardo (2011) identified success factors for (scientific) data sharing in a collaborative network which include trust, common terminology, harmonization of external socio-factors, and common working principles, values, policies, and organizational commitment. More factors highlighted in this literature include a common need for shared data, strong collaborative leadership, clear agreements on data standards and data policy, good timing and clear need, strong partners, broad-based involvement, credibility and openness of process, support of established authorities, strong leadership, interim successes, and a shift to broader concerns (Hale et al., 2003).

2.3. Theoretical framework of critical factors for data collaboratives

We have seen that the two research streams we reviewed – CCSP and CBIS – and the emergent data collaboratives research propose a large number of factors which are seen as important for the formation and implementation of data collaboratives. To systematize and organize these factors, we needed a suitable framework which would capture the socio-technical nature and the various layers of complexity of this phenomenon. When taken in isolation, none of the frameworks we reviewed sufficiently explains the phenomenon of data collaboratives, therefore we combined key elements from both research streams and proposed an integrated framework below (Fig. 1). To reiterate, our approach rested on the view that data collaboratives combine features of information sharing and integration projects with the specifics of cross sector social partnerships. Therefore, we combined elements from CCSP frameworks that detail how partnerships are formed (starting conditions) and how collaboration between actors unfolds (collaboration dynamics) with elements from information sharing frameworks that detail the various layers of socio-technical complexity of such projects. Our integrative framework is depicted below.

We further identified 32 critical factors mentioned in these literature streams as reviewed in the previous sections. We grouped these factors in the four dimensions from the framework: Organizational, Technological, Legislation and Policy, and Environmental factors. We will discuss how these factors relate to the starting conditions or collaboration dynamics of data collaboratives in the following sections.

Having proposed a framework (Table 1) and list of factors from the literature, we now describe the empirical part of our study which was aimed at validating the insights from the literature.

3. Method

To conduct our study of success factors we used the Critical Success Factors approach (Borman & Janssen, 2012; Remus & Wiener, 2010; Rockart, 1981). The concept of success factors was first proposed

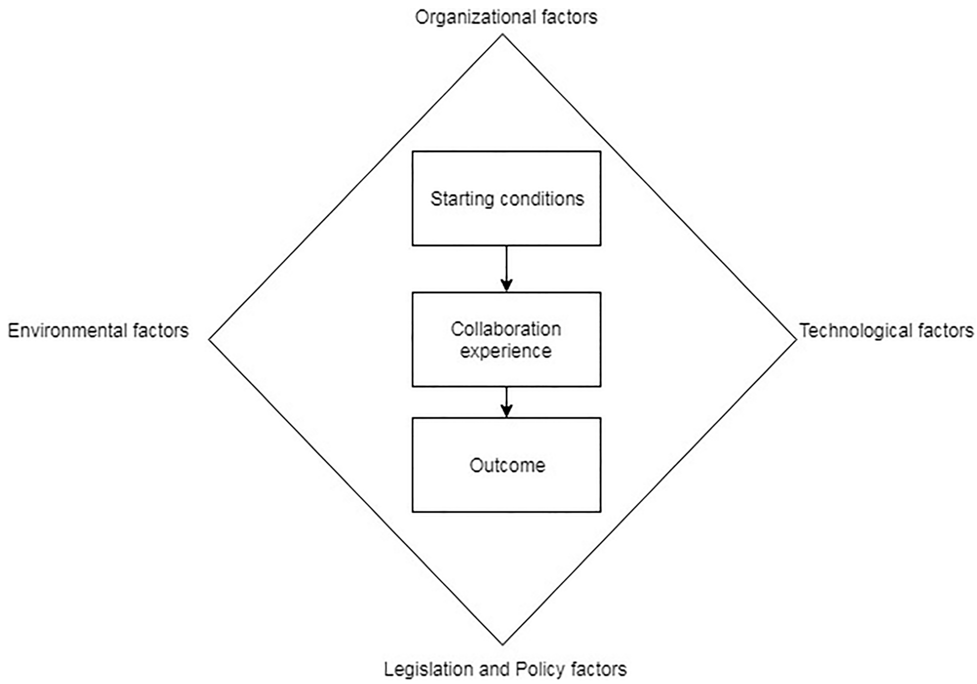

Fig. 1. Initial integrative framework of factors critical to data collaboratives.

in 1961 by Ronald Daniel of McKinsey & Company. The approach was formulated in 1979–1981 by John Rockart, co-founder of the MIT Centre of Information Systems research (Rockart, 1979, 1981). Rockart's approach was supposed to help executives who suffer from information overload to define their information needs. In its original interpretation, critical success factors (CSFs) are areas on which executives should focus their attention and seek data and reporting. Although this approach is decades old, it is still relevant for the present time of data explosion. Next to the original approach of Rockart, subsequent approaches were proposed which focus more on the universality of factors (rather than them being context specific), on the implementation process (rather than the outcome), and on the project as a level of analysis (rather than individual manager) (Borman & Janssen, 2012).

Overall, CSF is a widely used approach which proved useful in previous research on data sharing and digital government in general. There are many examples of research studies using this approach in this field, including on the topics of e-government adoption (Rana et al., 2013), national e-ID (Melin et al., 2013), emergency management (Chen, 2011), e-participation (Panopoulou et al., 2014), shared services in public sector (Borman & Janssen, 2012), adoption of electronic document management systems in government (Alshibly et al., 2016), open data publication and use (Susha et al., 2015), to name a few. The strengths of the CSF approach are that it can be helpful for improving project implementation and for making sense of complex problems where many factors occur, it offers an opportunity for learning and transferability of project lessons, and it is a widely tested method with a rigorous procedure. The weaknesses of the CSF approach, however, are that there is a lack of theoretical base (especially in the digital government domain), there is a discrepancy between the original and subsequent approaches, it is difficult to balance specificity with generality, and it may be challenging to define what is meant with 'success'.

There is a variety of research methods which can and have been used to identify success factors, such as case studies, literature reviews, focus groups, scenario analysis, structured interviewing, group

Table 1
List of critical factors for data collaboratives from the literature

#	CSFs	Select references
Organizational factors		
1	Appropriate and cost-effective business model	Robin et al., (2016)
2	Articulating a clear and compelling value proposition to stakeholders	The Chatham House (2015), Gao et al., (2015)
3	Availability of financial (and human) resources	Gil-Garcia and Sayogo (2016), Thune (2011)
4	Strong consortium with all required capacities and partner complementarity and fit	Osei-Kyei and Chan (2015), Thune (2011), Marlier et al. (2015)
5	Alignment of incentives of the participants	World Bank (2015), Perkmann and Schildt (2015)
6	Shared understanding of objectives, values, and expected outcomes	Ansell and Gash (2008), World Bank (2015), Smith and Dickson (2003), Sayogo and Pardo (2011), Thune (2011), Gao et al. (2015), Magee (2003)
7	Matching the problem with the data or data insights needed to address it	Gao et al. (2015), Magee (2003)
8	Long-term strategy (data sharing strategy)	Gao et al. (2015)
9	Clear measurable outcomes	Gao et al. (2015), Magee (2003)
10	Structured approach with clear (but flexible) agreements and regulatory mechanisms	Robin et al. (2016), Smith and Dickson (2003)
11	Broad participation of all affected stakeholders throughout the process	Robin et al. (2016), Ansell and Gash (2008), The Chatham House (2015), Magee (2003)
12	Top management support	Gao et al. (2015)
13	Building trust and investing in the relationship	Robin et al. (2016), Ansell and Gash (2008), Sayogo and Pardo (2011), Thune (2011), Wohlin et al. (2012), Marlier et al. (2015)
14	Facilitative leadership (via a formally assigned manager)	Ansell and Gash (2008), Gil-Garcia and Sayogo (2016), Thune (2011), Magee (2003)
15	Clear definition of responsibilities and process steps (iterative process)	Thune (2011), Gao et al. (2015), Marlier et al. (2015)
16	Continuous mutual adjustment of partners to each other and adaptation of their roles (interdependence)	Ber and Branzei (2010), Marlier et al. (2015)
17	Open and regular communications (personal contact)	Ansell and Gash (2008), Smith and Dickson (2003), Thune (2011), Marlier et al. (2015)
18	Commitment of stakeholders to the process	Ansell and Gash (2008), Smith and Dickson (2003), Sayogo and Pardo (2011), Thune (2011), Wohlin et al. (2012), Magee (2003)
19	Evaluations (to identify "small wins")	Ansell and Gash (2008), Thune (2011), Wohlin et al. (2012), Marlier et al. (2015)
20	Contingency planning	Smith and Dickson (2003)
21	Adequate technical/analytical skillsets and multidisciplinary teams	The Chatham House (2015), Gao et al. (2015)
22	Fast delivery of results	Gao et al. (2015)
Technological factors		
23	Compatibility of technical infrastructure and interoperable standards	Gil-Garcia and Sayogo (2016)
24	Using a systematic and transparent process to data sharing	Robin et al. (2016), Osei-Kyei and Chan (2015)
25	Using simple and familiar data sharing infrastructure	The Chatham House (2015)
26	Common concepts and terminology to enable data integration	Sayogo and Pardo (2011), Gao et al. (2015)
27	High quality of data	Gao et al. (2015), Magee (2003)
28	Innovative analysis tools	Gao et al. (2015)
Legislation and policy factors		
29	Adhering to standards and community norms, including privacy and security	The Chatham House (2015), Gao et al. (2015)

Table 1, continued

#	CSFs	Select references
30	Appropriate risk sharing and effective risk mitigation strategies	Robin et al. (2016), World Bank (2015), Osei-Kyei and Chan (2015)
Environmental factors		
31	Public pressure, community support, and/or political readiness	The Chatham House (2015), Osei-Kyei and Chan (2015), Magee (2003), Marlier et al. (2015)
32	Harmonizing geo-cultural differences among partners and/or acknowledging local context	Smith and Dickson (2003), Sayogo and Pardo (2011), Magee (2003)

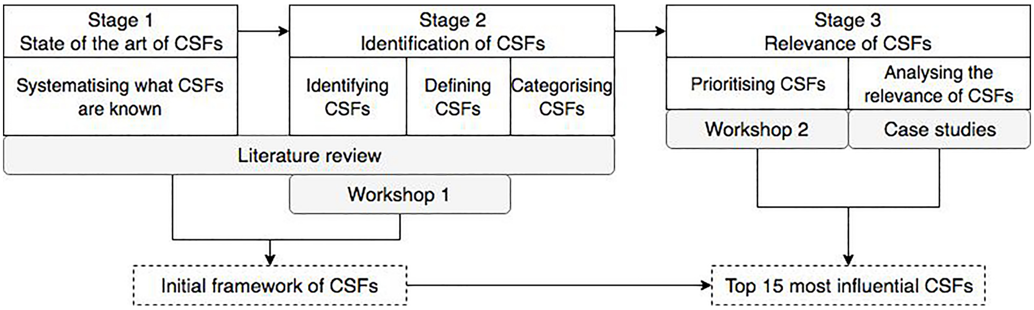

Fig. 2. Research design.

interviewing, Delphi technique, action research and others (De Sousa, 2004). Overall, the CSF approach includes four stages (De Sousa, 2004): state of the art, identification of CSFs, relevance of CSFs, and management of CSFs. We chose to follow this approach because it is informed by previous literature, it is straightforward and easy to apply, and it is comprehensive as it combines theory with practice in its different stages.

Figure 2 shows how we applied this approach in our research design. We have omitted the last phase concerning the management of CSFs as our study is exploratory.

Stage 1. State of the art. The first stage of the research – a literature review to systematize what factors are known – was described in the previous section.

Step 2. Identification of CSFs. To ensure that we did not miss any factors, we used the results of a practitioner workshop held at the International Data Responsibility Conference in The Hague (The Netherlands) in February 2016. The workshop was organized by The Gov Lab and was attended by representatives of companies, research institutions, non-profit sector, government (29 participants in total). The participants were asked to formulate factors important for data collaboratives from their experience. We did not provide them with our framework or categories beforehand, this ensured they were not biased towards any particular category or factor and could think about CSFs with an open mind. The factors proposed during the workshop were included in our framework developed from the literature.

Stage 3. Relevance. Once the framework was complete, we conducted an expert workshop to discuss the criticality of the identified CSFs and to prioritize them. The workshop took place during the IFIP WG 8.5 Electronic Government (EGOV) conference in September 2016. It was attended by 17 experts/researchers in the fields of open and big data, e-government, public private partnerships, and information sharing. The majority of the participants had over 5 years of experience. The experts were asked to select from our framework and rank the factors which (1) influence the process of data collaboratives the most, (2) influence the outcome of data collaboratives the most, and (3) without which a data collaborative would fail. This resulted in a list of top most critical CSFs.

Table 2
Case studies

	Case	Description
1	Nepal's telecom data and post-earthquake mobility	Shortly after 2015 earthquake in Nepal, call detail records of 12 million mobile phone users in Nepal was shared by the Nepali telecom operator NCell with a non-profit Flowminder in Sweden. Flowminder analyzed the data to map population flows after the disaster
2	Palm risk tool of Global Forest Watch	A tool launched in 2016 by the Global Forest Watch (GFW) which can be used by companies, as well as governments, conservation organizations, and the general public, to understand deforestation risks associated with palm oil mills. GFW is a partnership coordinated by the World Resources Institute

To validate the results of the ranking by the EGOV community experts, we conducted two case studies. To select suitable cases, we used the Data Collaboratives Explorer (see DataCollaboratives.org) which is a repository of cases compiled by The Gov Lab. We chose to limit our inquiry to a certain type of collaborative – the trusted intermediary model – to have a better ground for generalizing our insights. The trusted intermediary model, when a neutral actor mediates the transfer of corporate data into publicly useful knowledge, has seen an uptake recently. The repository of cases we used listed 29 of such partnerships at the time. The following were our selection criteria:

- Represents a trusted intermediary model of a data collaborative
- Has achieved positive outcomes (based on coverage in press)
- Available for interviews and document studies

We included cases from different domains, as long as they conformed to our selection criteria (selection by diversity). The following are the selected cases for our analysis:

To conduct the case studies (Table 2), we used document studies and interviews to collect the data. The interview protocol is presented in the Annex and was based on the guidelines of Rockart (1981) and Borman and Janssen (2012). The interviews were explorative and took place in November 2017. We were particularly interested in success factors from the perspective of the user of the data, therefore in both cases we interviewed organizations who were on the receiving end of the data chain, i.e. accessing and processing data to provide insights. In both cases we interviewed one top-level manager who played a leading role in conceiving and overseeing the collaboration with a private sector partner(s). We did not provide the interviewees with our framework beforehand to avoid steering the conversation towards any particular factor or category.

4. Findings

To identify more CSF (stage 2 in our CSF approach), in addition to the list we compiled from the literature, we used the results of an expert workshop as explained in the Method section. The participating experts did not suggest any factor that was not already present in our list, however they highlighted the importance of several. 10 factors in total were discussed during the workshop. The results of the workshop are available online.[2] In particular, the following factors were emphasized: matching supply and demand and data with problems (CSF7), identifying approaching business models (CSF1), long term strategy and sustainability plans (CSF8), building trust relationships (CSF13).

[2] https://bit.ly/2SLIjrr.

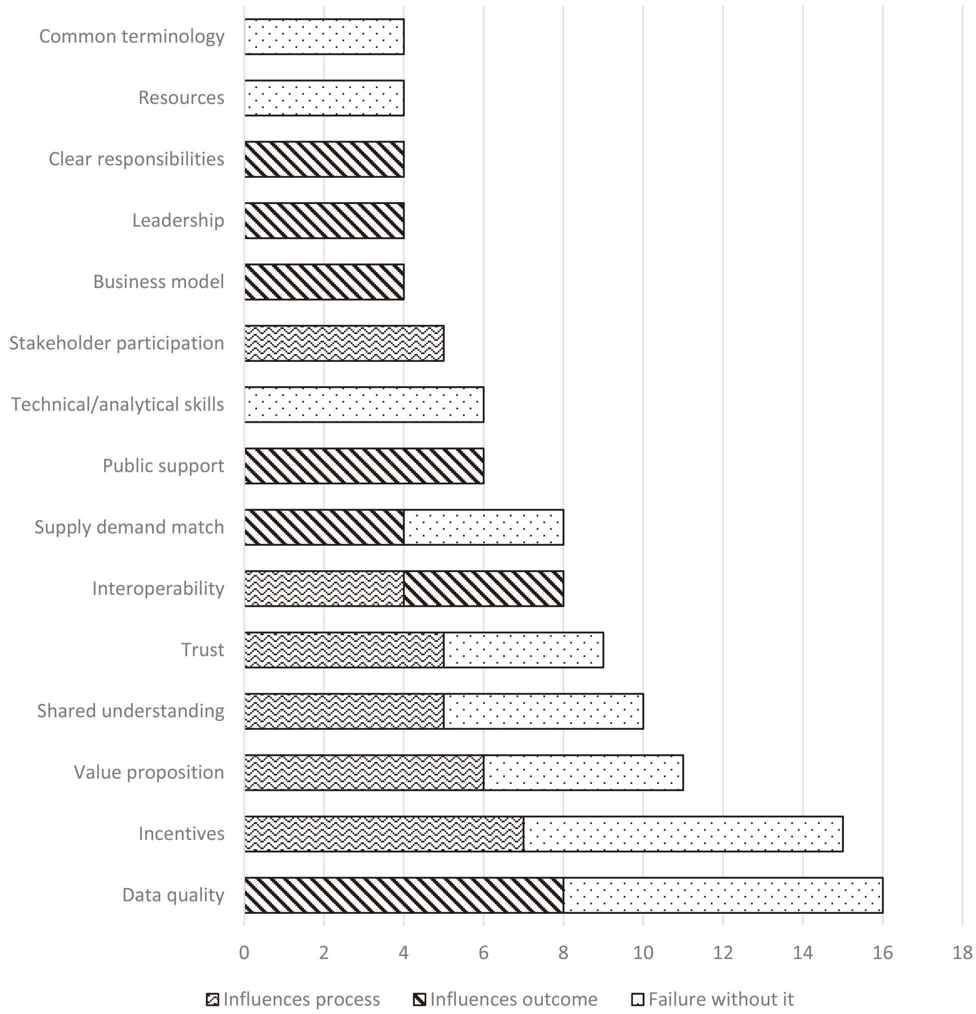

Fig. 3. Top 15 most critical factors for data collaboratives based on ranking by experts.

4.1. *Prioritization of factors*

In this section, we report on the results of the expert workshop (workshop 2) to assess the relevance of factors and prioritize their importance which is the third step of our CSF approach. The full results of the expert survey are available online.[3] We would like to highlight the following findings.

During the workshop, we asked the audience to select from our framework and rank the factors which (1) influence the process of data collaboratives the most, (2) influence the outcome of data collaboratives the most, and (3) without which a data collaborative would fail. Figure 3 shows the top picks and the distribution. The total number of experts who responded to these questions was between 15 and 17. The horizontal axis shows the number of votes each factor got from the experts.

Out of 32 factors in our theoretical framework this graph shows 15 which were found as critical by

[3] https://bit.ly/2uxP5pT.

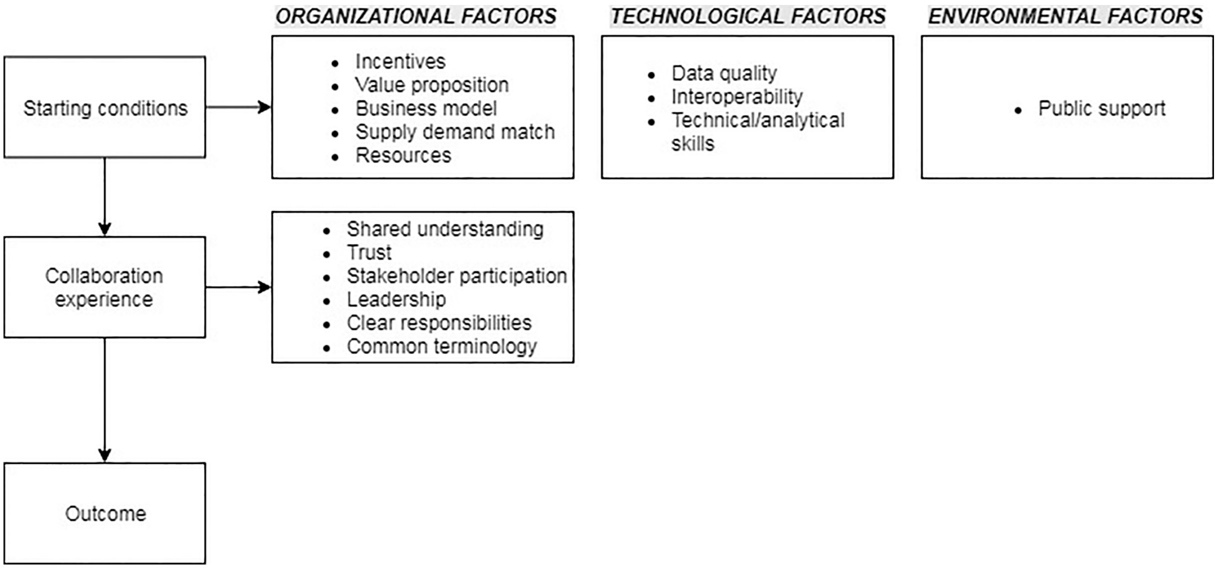

Fig. 4. Top most critical factors for the formation and implementation of data collaboratives.

the experts. The graph shows that *data quality* and *incentives* have the greatest combined influence over the success of a data collaborative. *Shared understanding, value proposition,* and *trust* were found to be critical by at least half of the panel. The graph also shows that different factors are seen as important for the process and for the outcome of collaboration. This means that while incentives are key for initiating a partnership, data quality matters the most when it comes to actually achieving the aims of the partnership in the end.

Figure 4 shows how the factors from the Top 15 list can be represented using the integrated framework we proposed in Section 2. It shows that the 15 factors can be divided between those which can be seen as starting conditions and those which are associated with the process of collaboration. We further grouped the factors by the layer to which it is relevant: Organizational, Technological, or Environmental. No factors which could be put into the Legislation and Policy category were included in the Top 15 ranking. We have thus organized the identified factors by whether they are seen as more critical for the formation (starting conditions) or for the implementation of data collaboratives (collaboration experience).

4.2. Analysis of case studies

4.2.1. Nepal's telecom data and post-earthquake mobility

On 25 April 2015, a 7.8 magnitude earthquake hit Nepal, including its capital city Kathmandu. 9 000 people were killed and 23 000 were injured. In the aftermath of the earthquake there were mass movements of the population as people were fleeing affected areas. The Nepali government faced a challenge of extending its relief activities because it did not have an overview of where the people were going to.

Shortly before the earthquake hit, a Swedish nonprofit organization Flowminder entered into a partnership with the largest mobile operator in Nepal NCell to access their anonymized call detail records and be able to map the population flows. Data of 12 million mobile phone users was shared by NCell.

In the result, the analysis of the data showed that an estimated 390,000 people above normal left the Kathmandu Valley soon after the earthquake. Many of these moved to the surrounding areas, and the highly-populated areas in the central southern area of Nepal. Reports containing these results, along with interpretation, were distributed to relevant humanitarian agencies working on the ground in Nepal. The success of the collaboration between Flowminder and NCell was recognized by the Global Mobile Award for Mobile in Emergency or Humanitarian Situations in February 2016.

Our inquiry focused on identifying what contributed to the success of this collaboration. As the interviewees explain, the partnership was initiated by Flowminder who had pitched their disaster response system to NCell one week before the earthquake. NCell were positive but reluctant to engage. When the earthquake hit, this was a decision point for NCell who agreed to share the data. The Telia Sonera group, who owned NCell at the time, offered support for the project and covered all the costs.

In conversation with Flowminder it became apparent that several factors played a role in the success of this collaboration. First, one should not look at this project in isolation as it was preceded by much work and engagements on the part of Flowminder. Flowminder pioneered the use of mobile data after the Haiti earthquake and have been trying to scale up since 2010. This process has been quite slow due to its technical complexity. Flowminder's mission was to get all telecom operators to support humanitarian response, and this was the goal towards which the organization had been working in the preceding 5–6 years. When approaching Ncell, Flowminder already had 5–6 years of track record and 3 years of negotiations with the Swedish operator group Telia Sonera (which at the time owned NCell). Flowminder also had endorsements from UN agencies and mobile industry association. All in all, it was a long discussion with many entry points in which the timing played an important role. Flowminder had a number of agreements with Telia Sonera before, although the local telecom operator was quite independent.

Second, one should look at this project from an ecosystem point of view. A lot of agencies have been advocating for this type of work (UN agencies, GSMA, UN Global Pulse, UN Data Innovation). According to the interviewee, it is very hard to isolate individual success factors. UN agencies have been trying to access mobile data for 6 years before that. Flowminder has been doing a lot of lobbying. "Most telecom operators Flowminder talked to receive about 1 call a week from UN agencies, NGO, university or donor who want to access their mobile data. It's been like that for the last few years", as the interviewee put it.

Third, the big picture is that "success cases are normally built on personal relationships, credibility, and some kind of value proposition", in the words of the interviewee. Using mobile data is very ad hoc in a way; those who use mobile data are research groups, such as at MIT. It was important that Flowminder had endorsements from UN agencies and GSMA who promote guidelines for that type of work. Discussions with Telia Sonera provided the connection to NCell. Flowminder met with key decision makers at NCell who received a full technical presentation of how they work. Without these factors the response would have been slower.

Therefore, we conclude that *prior collaboration, reputation, credibility, community support, personal relationships, and value proposition* have been crucial in this case to create a network effect. Thus, we find that the following factors from our Top 15 CSFs have been critical in this case:

- Articulating a clear and compelling value proposition to stakeholders
- Building trust and investing in the relationship
- Public pressure and/or community support and/or political readiness

The factors of prior collaboration, reputation, credibility, and personal relationships are not explicitly mentioned in our Top 15 CSFs, however, they can be seen as strategies for gaining trust of the partners and for investing in the collaboration.

4.2.2. Palm risk tool of Global Forest Watch

Global Forest Watch is an interactive online forest monitoring and alert system aimed to support better management and conservation of forests around the world. The platform is free to use for anyone. It is an initiative of the World Resources Institute, which is a global research nonprofit organization, who coordinates a large consortium of partners, including Google, Esri, USAID and many others. Global Forest Watch capitalized on three innovations, such as free governmental satellite data, and developments in cloud computing and machine learning.

In June 2016, in the framework of its Commodities program, Global Forest Watch introduced a new tool Palm Risk which can be used to identify which palm oil mills in the world are likely to lead to deforestation. Establishing supply chain traceability in this case is a challenging task because any given mill sources from a thousand of small farms. Even if companies have this data they are reluctant to share it because it is viewed as proprietary confidential information which should not be shared with competitors.

The tool was built using satellite data and data provided by the company FoodReg which traces supply chains of products around the world. Several multinationals making up about 80 percent of the global supply chain provided data to the project which was consolidated and anonymized by the GFW so that it was impossible to link which companies buy from which mills. As a result, a publicly available dataset was created of over a thousand mills. Palm Risk can also be used by governments, conservation organizations, and the general public who wish to monitor deforestation. GFW also created a tool specifically for companies to help them analyze deforestation risks associated with specific mills.

The Palm Risk tool was initially tested by Unilever who is one of the world's largest purchasers of palm oil. The pilot revealed that about 5 percent of the mills from the company's supply chain were at high risk of causing deforestation. In February 2018 Unilever was the first company to publish its full list of palm oil mills and is said to be taking steps to address this and other findings. At the time of the interviews, the project was perceived as successful thanks to the high level of commitment from companies with whom GFW collaborated on this issue. Another measure of success according to the interviewee was the number of companies with whom they signed MoUs and who used their tool.

In our exploratory interviews, we established that one of the main challenges GFW faced was finding out how to deliver information to companies that they can easily digest and use in their decision making. As the interviewee explained, tools built by NGOs are not typically used by big businesses. Companies want easy push-button tools to make decisions, thus understanding their needs was critical in the project. This involved a lot of scoping and prototyping to overcome these design challenges. As the interviewee phrased it, "we had the right solution at the right time when the companies had this need; demand was there, it was easy for us to take advantage of it". The value proposition and incentive for companies to share their supply chain data was that in return they would be able to use the Palm Risk tool for their decision making.

Another challenge was addressing companies' concerns about the sensitive nature of the data. Thus, GFW opted for a downloadable tool which can be used behind companies' firewalls so that proprietary data never has to leave their firewalls. This was seen as the first step before companies feel confident to share such data with each other and can see efficiencies from that.

To make a change, the uptake of the tool in the market is seen as critical. The tool becomes more useful once more actors begin to use it. At the same time, it is important to sustain the commitment of companies who are already using it. This will allow to obtain more granulated data to the level of farms and thus considerably improve supply chain traceability.

Another critical factor highlighted by the interviewee was public pressure and the combined work of other actors in the field. As the interviewee explained, success was due to "broad factors". For instance, Greenpeace had been campaigning to save the forests for years, and it is now that the impact of that is visible in companies making commitments such as in the work with GFW. “We can never completely attribute it to ourselves, so many organizations played a role”, according to the interviewee.

When asked what would have caused failure, the interviewee pointed out the issue of neutrality and the role of WRI as an evidence-based provider of information and analysis, as opposed to advocates and campaigners. It was reputation, credibility, and trust that played a major role in the success of the project. Furthermore, WRI was seen as a ‘trusted intermediary’, a convener of different parties, since they work with both public, private and nongovernment sectors. “We are a spider in the center of the web”, the interviewee commented. This was seen as the most important success factor. Following from that, GFW in general uses a very inclusive approach and avoided formal governance arrangements. Their governance structure is very flexible and can evolve. According to the interviewee, “multi-stakeholder governance approaches tend to slow things down”.

Therefore, we conclude that *responding to the need, offering value in return to data sharing, ensuring data confidentiality, obtaining commitment and trust from companies, benefitting from public and community pressure, and the right reputation* have been crucial in this case. Thus, we find that the following factors from our Top 15 CSFs were found to be critical in this case:

- Articulating a clear and compelling value proposition to stakeholders
- Alignment of incentives of the participants
- Building trust and investing in the relationship
- Stakeholder commitment and participation in the process
- Public pressure and/or community support and/or political readiness

The factor of responding to the need is related to the factor of public pressure to address a certain societal issue which is included in our initial framework. The factor of data confidentiality mentioned by the interviewee is not explicitly covered in the Top 15 CSFs but is present in the initial framework of factors from the literature.

5. Discussion

Based on the ranking by experts, we proposed a list of Top 15 critical factors for data collaboratives. This list includes factors from most categories of our initial framework of 32 factors: Organizational, Technological, and Environmental. To further validate these factors, we conducted two case studies.

Overall, we find an overlap between the prioritization of factors by experts and the findings of our case studies. Most factors mentioned by the interviewees were also top ranked by the experts, although the exact order of priority varies. The main discrepancy is that data quality was not explicitly mentioned as a critical factor in either of the cases. This can be explained by the fact that in both cases the receiving organization was confident about the quality of the data shared (call detail records, supply chain data). Furthermore, in both cases the research nonprofits were aware of the limitations of the data. In the Flowminder case, the data was only from one, although the largest, telecom operator. And in the case of GFW, the granularity and frequency of data could be improved. This however did not impede the course of the collaboration in either cases.

In both cases which we investigated a research non-profit organization approached a private sector partner(s) with an offer to collaborate in order to address a pressing societal problem (aftermath of a disaster and deforestation, respectively). The cases differ in many aspects, such as the purpose, the organizational arrangement, the type of data shared, duration etc. In both cases one of the main challenges mentioned was persuading the private sector to collaborate and addressing their concerns about potential risks to their data and implications thereof. Therefore, both interviewees tended to focus on the formation phase of the collaboration. Both Flowminder and GFW are organizations with resources and skills which are required to provide state of the art technical solutions in this respect. This may explain why fewer of the factors found as important in the case studies have to do with technology. Overall, we find that the factors which were considered to be most critical for the formation of these partnerships are similar. The common factors important in both cases are: *value proposition, gaining trust* (through reputation, personal relationships, or prior collaboration), *and effect of public pressure.* Relative to our integrative framework proposed in Section 2, the case studies emphasized the importance of starting conditions for the formation of data collaboratives. Value proposition, public pressure, and trusting relationship between partners are seen as key elements of a successful collaborative. We discuss them in more detail below. Although in our framework trust was included as part of collaboration experience, our case studies showed that it is equally (if not more) important at the outset of collaboration.

The first broad factor common between the case studies was the **value proposition**. In the recent CCSP literature a model of success of business – NGO partnerships was proposed (Barroso-Mendez et al., 2016) according to which shared values affect trust and commitment which in turn affect relationship learning and have a direct influence on the partnership success. According to this model, it is essential for businesses to share similar values and beliefs with their potential partners in the partnership formation phase. Shared values can be understood as "the degree to which the partners have beliefs in common about what behaviors, goals, and policies are important or unimportant, appropriate or inappropriate, and right or wrong' (Morgan & Hunt, 1994, p. 25). Our Top 15 list of critical factors includes several factors which are related to and add to this construct: shared understanding (of objectives, values, and expected outcomes), value proposition, incentives, and business model. Thus, our study conveys a view common to the data collaboratives literature which we reviewed that data collaboratives should not only generate social value but also value to the private sector engaging in such partnerships. In this respect, we concur with the underlying proposition of the 'shared value approach' of Kramer and Porter (2011) that engagements of the private sector with social issues should have an explicit connection to the company's profitability and competitive position instead of simply focusing on reputation (Ibid.). Besides positive reputation, incentives for companies to share data for a societal purpose include: new insights, recruitment opportunities, reciprocity, new revenue models, regulatory compliance, and "an ecosystem-supporting responsibility" (OECD & The Gov Lab, 2017). New studies (GSMA, 2018) described a variety of business models that are emerging around data partnerships for social good, although more systematic research is needed. Articulating these benefits in the formation stage of the data collaborative is of utmost criticality, as our case studies suggest. Current research on cross boundary information sharing does not explain the importance of value proposition sufficiently because until lately it has predominantly focused on data sharing between public sector entities. In the public sector, data is considered as a public good and there is often a moral imperative to open up information. The contrary applies to the private sector where data is treated as a resource and 'the new currency', thus, giving it away for free is counterintuitive to the business logics of companies. Therefore, we argue that this factor is specific to data collaboratives.

The second broad factor which was common in both cases is **trust**. Trust was also highly ranked by the experts in our Top 15 list of factors. Trust is a well described topic in the current CCSP and CBI literature. Therefore, our findings concur with these research streams. Trust can be defined as the expected collaboration in which the expectation is that partners will do what is best for the entire system, even when that may lead to a partner's disadvantage (Capaldo & Giannoccaro, 2015). The model of

Barroso-Mendez et al. (2016) puts trust at the core of partnership success. In fact, the better the relationship between partners in terms of trust and commitment, the greater is the extent to which they exchange information and knowledge (Sanzo et al., 2015). CBI literature concurs that trust is one of the most important factors in making cross boundary information sharing successful (Sutherland et al., 2018). Dawes et al. (2009) distinguish between three types of trust: information-based trust (rational decisions about who to trust), identity-based trust (familiarity between stakeholders), and institution-based trust (norms and social structures that define acceptable behavior). As evidenced by our case studies, all three types of trust are relevant for the formation of data collaboratives, as the interviewees mentioned the importance of personal relationships, credibility, reputation of their organization, and prior collaboration. Moreover, in the context of data collaboratives, there are a number of risks associated with the sharing of data, especially when it concerns personal (customer) data. Data can represent a "competitive advantage within the market" and companies need assurance that their data will not be leaked to unauthorized parties (Praditya & Janssen, 2015). Besides security and potential data leaks, there is a risk that the data may be misinterpreted or used inappropriately (The Gov Lab, 2018). Therefore, gaining trust of private sector partners in a data collaborative involves assuring them that the data will be handled responsibly. It is thus not coincidental that, according to the interviewee in the Palm Oil case, ensuring data confidentiality was one of the critical factors. This means that, besides reputation and credibility, the recipients of data are often expected to have data analytics capabilities which can be trusted. In fact, in an information exchange relationship competence-based trust is often found as more influential than affection-based trust (Sayogo & Pardo, 2011). Therefore, it is not incidental that technical and analytical skills are included by the expects in our Top 15 list of factors.

The third broad factor concerns the effect of **public pressure**. Unlike the two aforesaid ones, this is an external factor characterizing the environment in which the collaboration is formed. Some authors view such factors as 'antecedent conditions' (Bryson et al., 2015) or 'starting conditions' (Ansell & Gash, 2008) but few studies actually shed light on the relationship between public pressure and success of partnerships. In situations when public and private organizations collaborate, literature identifies political support, policy support, and/or community support as important for the progress of partnerships (Osei-Kyei & Chan, 2015; Marlier et al., 2015). However, we argue that public support is not equivalent to public pressure. Both of the cases we surveyed highlighted the 'network effect' of the public and community pressure to collaborate and share information. The specifics of data collaboratives are that these are partnerships often driven by the need to address a pressing and often urgent societal problem. Many of these problems, like climate change or poverty, are 'wicked' (Manning & Reinecke, 2016) and pose 'grand challenges' (George et al., 2017) because their magnitude is difficult to assess, and the cause-effect relationship is highly complex. At the same time, there is mounting pressure in the public sphere to address these problems. This explains why public pressure is seen as a particularly influential factor for the formation and implementation of data collaboratives. For instance, in the Flowminder case it was seen as critical that other actors in the ecosystem approached the telecom on a regular basis with requests to share data. And in the Palm Oil case responding to the need to facilitate supply chain transparency was seen as influential. We thus argue that public pressure is particularly influential factor for data collaboratives.

In our study we were interested to find out whether data collaboratives as a novel form of partnership should be implemented in any different way. What we found is that many of the factors known from more 'traditional' collaborations are still valid for data collaboratives, such as trust, public support, shared goals, motivations etc. Our Top 15 list of factors captures many factors described in previous research. However, we also found that some factors that received marginal attention or were not at all covered in established literatures (only in emergent data collaboratives research) were included in our Top 15 list. They are business models and value proposition and matching of data to problems and thereby of demand with supply. This is the niche where new research is needed to develop frameworks and tools for business model innovation and for problem formulation in the context of data-driven decision making.

6. Conclusion

In our study, we proposed a theoretical framework of critical factors relevant to data collaboratives based on the review of literature and discussions of an expert workshop (Table 1). We further prioritized the factors from the framework in an expert workshop at an academic conference and reduced the number of factors from 32 to 15. The Top 15 factors (Fig. 2) are expected to be of utmost criticality for data collaboratives. To test whether this is the case, we conducted two exploratory studies of data collaboratives in different domains. Our conclusion is that we found support for the identified factors from the interviews. Thus, the full framework is suggested to be used as a guideline for the implementation of data collaboratives, while the reduced list of 15 factors highlights the most important areas. In Section 5 we further discussed three broad factors which were found critical in both cases: value proposition, trust, and public pressure. We argue that in relation to each of these factors there are certain nuances when it comes to data collaboratives as an emerging form of partnership.

Our study makes a research contribution to the field as it integrates previously fragmented knowledge about what contributes to a successful data collaborative. This is an emerging topic which exists at the intersection of more established research fields. Our study is one of the first to situate this phenomenon in existing academic research and thereby contribute to its conceptual maturation. We offer our initial framework of factors, as well as our Top 15 list, to be tested in future empirical studies. We also contribute to the literatures on cross sector social partnerships and cross boundary information sharing by putting a spotlight on an innovative form of partnership catalyzed by the data revolution.

In terms of practice, our initial list of factors can be used as a general guideline by practitioners contemplating to engage in a data collaborative. It provides a holistic view of elements which come into play when a data collaborative is formed. The value of the Top 15 list is that it highlights elements which typically have a greater influence over the success of the partnership. This list thus can be used to help organizations prioritize and distribute resources accordingly when engaging in a data collaborative. This list also highlights factors which capture the specificity of data collaboratives compared to other types of partnerships. The value of the Top 15 list is not in the rankings of factors (which was exploratory and only provides an indication of relative importance) but in the fact that it offers a concise view of 15 select factors which deserve immediate attention.

To acknowledge the limitations of our study, the case studies primarily focused on the formation phase and did not capture long term impacts of these partnerships. Furthermore, both case studies are limited to the trusted intermediary model of data collaboratives; other forms of collaboratives may require a different mix of critical factors. Future research can test and elaborate our findings using more cases from other contexts (other types of data collaboratives, from other domains, involving other kinds of participants etc.).

The phenomenon of data collaborative is quite recent, so future research can include longitudinal studies of data collaboratives. This may generate a more comprehensive assessment of success of a data collaborative, also taking into account the level of satisfaction of the partners with the outcomes and the actual influence on the policy issue in question.

Acknowledgments

This research was funded by the Swedish Research Council under the grant agreement 2015-06563 “Data collaboratives as a new form of innovation for addressing societal challenges in the age of data”. The author wishes to thank Efthimios Tambouris, Marijn Janssen, and Åke Grönlund for co-organizing the workshop at the IFIP EGOV conference; to the interviewees who participated in the case studies; and to the anonymous reviewers of Information Polity for their feedback.

References

Alshibly, H., Chiong, R., & Bao, Y. (2016). Investigating the Critical Success Factors for Implementing Electronic Document Management Systems in Governments: Evidence From Jordan. *Information Systems Management*, *33*(4), 287-301.
Ansell, C., & Gash, A. (2008). Collaborative governance in theory and practice. *Journal of Public Administration Research and Theory*, *18*(4), 543-571.
Barroso-Mendez, M. J., Galera-Casquet, C., Seitanidi, M. M., & Valero-Amaro, V. (2016), Cross-sector social partnership success: A process perspective on the role of relational factors, *European Management Journal*, *34*, 674-685.
Ber, M. J., & Branzei, O. (2010), (Re)forming strategic cross-sector partnerships: Relational processes of social innovation. *Business and Society*, *49*(1), 140-172. doi: 10.1177/0007650309345457.
Bigdeli, A. Z., Kamal, M. M., & De Cesare, S. (2013), Electronic information sharing in local government authorities: Factors influencing the decision-making process. *International Journal of Information Management*, *33*(5), 816-830.
Borman, M., & Janssen, M. (2012), Reconciling two approaches to critical success factors: The case of shared services in the public sector. *International Journal of Information Management*. doi: 10.1016/j.ijinfomgt.2012.05.012.
Bryson, J. M., Crosby, B. C., & Stone, M. M. (2015), Designing and implementing cross-sector collaborations: Needed and challenging. *Public Administration Review*, *75*(5), 647-663.
Capaldo, A., & Giannoccaro, I. (2015), How does trust affect performance in the supply chain? The moderating role of interdependence. *International Journal of Production Economics*, 16636-49.
Chen, Y. (2011). Research on critical success foctors for Emergency Management Information System. Paper presented at the 2011 2nd International Conference on Artificial Intelligence, Management Science and Electronic Commerce, AIMSEC 2011 – Proceedings.
C&E Advisory (2017). Corporate-NGO Partnerships Barometer 2017. Headline findings. Accessed 15 October 2018 from https//www.candeadvisory.com/sites/candeadvisory.com/files/headline_findings_2017_7.pdf.
Dawes, S. S. (1996). Interagency information sharing: Expected benefits, manageable risks. *Journal of Policy Analysis and Management*, *15*(3), 377-394.
Dawes, S. S., Cresswell, A. M., & Pardo, T. A. (2009), From “need to know” to “need to share”: Tangled problems, information boundaries, and the building of public sector knowledge networks. *Public Administration Review*, *69*(3), 392-402.
Dawes, S. S., Gharawi, M. A., & Burke, G. B. (2012), Transnational public sector knowledge networks: Knowledge and information sharing in a multi-dimensional context. *Government Information Quarterly*, *29*, S112-S120.
Dawes, S. S., & Pardo, T. A. (2002), Building collaborative digital government systems. In *Advances in digital government (pp. 259–273)*. Springer, Boston, MA.
De Sousa, J. M. E. (2004), *Definition and analysis of critical success factors for ERP implementation projects.* Universitat Politècnica de Catalunya.
Ebrahim-Khanjari, N., Hopp, W., & Iravani, S. R. (2012), Trust and Information Sharing in Supply Chains. *Production & Operations Management*, *21*(3), 444-464.
Emerson, K., Nabatchi, T., & Balogh, S. (2012), An integrative framework for collaborative governance. *Journal of Public Administration Research and Theory*, *22*(1), 1-29.
Estevez, E., Fillottrani, P., & Janowski, T. (2010 June). Information sharing in government-conceptual model for policy formulation. In *Proceedings of The 10th European Conference on e-Government (ECEG)*. University of Limerick, Ireland, pp. 152-162.
Forbes (2014). inBloom’s Collapse Offers Lessons For Innovation In Education. Accessed 15 October 2018 from https//www.forbes.com/sites/michaelhorn/2014/12/04/inblooms-collapse-offers-lessons-for-innovation-in-education/#18579b4f525f.
Gao, J., Koronios, A., & Selle, S. (2015), *Towards a process view on critical success factors in Big Data analytics projects.* Paper presented at the 2015 Americas Conference on Information Systems, AMCIS 2015.
George, G., Howard-Grenville, J., Joshi, A. & Tihanyi, L. (2017). Understanding and Tackling Societal Grand Challenges through Management Research Academy of Management Journal, *59*(6).
Gil-Garcia, J. R., Chengalur-Smith, I., & Duchessi, P. (2007), Collaborative e-Government: impediments and benefits of information-sharing projects in the public sector. *European Journal of Information Systems*, *16*(2), 121-133.

Gil-Garcia, J. R., & Sayogo, D. S. (2016), Government inter-organizational information sharing initiatives: Understanding the main determinants of success. *Government Information Quarterly*. doi: 10.1016/j.giq.2016.01.006.

Gottschalk, P., & Solli-Sæther, H. (2005), Critical success factors from IT outsourcing theories: an empirical study. *PIndustrial Management & Data Systems*, *105*(6), 685-702.

Gray, B., & Stites, J. P. (2013), Sustainability through partnerships. Capitalizing on collaboration. Network for business sustainability, case study.

GSMA. (2018). Scaling Big Data for Social Good: The need for sustainable business models. Retrieved 24 July 2019 from https//bigdatatoolkit.gsma.com/resources/BD4SG_The_need_for_sustainable_business_models.pdf.

Hale, S. S., et al. (2003), Managing troubled data: Coastal data partnerships smooth data integration. *Environmental Monitoring and Assessment 81*(1-3), 133-148.

Hartman, L. P., & Dhanda, K. K. (2018). CrossâĂŘSector Partnerships: An Examination of Success Factors. *Business and Society Review*, *123*(1), 181-214.

Klievink, B., van der Voort, H., & Veeneman, W. (2018), Creating value through data collaboratives. *Information Polity*, (Preprint), 1-19.

Kramer, M. R., & Porter,| M. (2011), Creating shared value. *Harvard Business Review*, *89*(1/2), 62-77.

Lee, H. L., & Whang, S. (2000), Information sharing in a supply chain. *International Journal of Manufacturing Technology and Management*, *1*(1), 79-93.

Lisbon C. (2017). Making Europe a data economy: A new framework for free movement of data in the digital age. Retrieved *20* August 2017 from http//www.lisboncouncil.net/news-a-events/741-a-new-framework-for-free-movement-of-data.html.

Magee, M. (2003). Qualities of enduring cross-sector partnerships in public health. *The American Journal of Surgery*, *185*(1), 26-29. doi: 10.1016/S0002-9610(02)01143-1/.

Manning, S., & Reinecke, J. (2016). We're failing to solve the world's 'wicked' problems. Here's a better approach. The Conversation. Available at: http//theconversation.com/were-failing-to-solve-the-worlds-wicked-problemsheres-a-better-approach-64949, (accessed 31 January 2017).

Marlier, M., Lucidarme, S., Cardon, G., De Bourdeaudhuij, I., Babiak, K., & Willem, A. (2015), Capacity building through cross-sector partnerships: A multiple case study of a sport program in disadvantaged communities in Belgium Health policies, systems and management in high-income countries. *BMC Public Health*, *15*(1). doi: 10.1186/s12889-015-2605-5.

Melin, U., Axelsson, K., & Söderström, F. (2013), Managing the development of secure identification: investigating a national e-ID initiative within a public e-service context 21st European Conference on Information Systems (ECIS 2013), (p. 12). Utrecht, Netherlands.

Morgan, R. M., & Hunt, S. D. (1994), The commitment-trust theory of relationship marketing. *Journal of Marketing*, *58*(3), 20e38.

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

Overseas Development Institute (2013). The data revolution: Finding the missing millions. Retrieved on 20 August 2017 from https//www.odi.org/sites/odi.org.uk/files/odi-assets/publications-opinion-files/9604.pdf.

Osei-Kyei, R., & Chan, A. P. C. (2015), Review of studies on the Critical Success Factors for Public – Private Partnership (PPP) projects from 1990 to 2013. *International Journal of Project Management*, *33*(6), 1335-1346. doi: 10.1016/j.ijproman.2015.02.008.

Panopoulou, E., Tambouris, E., & Tarabanis, K. (2014), Success factors in designing eParticipation initiatives. *Information and Organization*, *24*(4), 195-213. doi: 10.1016/j.infoandorg.2014.08.001.

Perkmann, M., & Schildt, H. (2015), Open data partnerships between firms and universities: The role of boundary organizations. *Research Policy*, *44*(5), 1133-1143. doi: 10.1016/j.respol.2014.12.006.

Praditya, D. D., & Janssen, M. M. (2015), Benefits and Challenges in Information Sharing Between the Public and Private Sectors. *Proceedings of the European Conference on E-Learning*, pp. 246-253.

Rana, N. P., Dwivedi, Y. K., & Williams, M. D. (2013), Analysing challenges, barriers and CSF of egov adoption. *Transforming Government: People, Process and Policy*, *7*(2), 177-198.

Remus, U., & Wiener, M. (2010), A multi-method, holistic strategy for researching critical success factors in IT projects. *Information Systems Journal*, *20*(1), 25-52.

Robin, N., Klein, T., & Jütting, J. (2016), Public-Private Partnerships for Statistics: Lessons Learned, Future Steps.

Rockart, J. F. (1979). Chief executives define their own data needs. *Harvard Business Review*, *57*(2), 81-93.

Rockart, J. F. (1981). A Primer on Critical Success Factors. In C. V. Bullen (Ed.), *The Rise of Managerial Computing: The Best of the Center for Information Systems Research*. Homewood, IL: Dow Jones-Irwin.

Sanzo, M. J., Alvarez, L. I., Rey, M., & García, N. (2015), Business-nonprofit partnerships: Do their effects extend beyond the charitable donor-recipient model? *Nonprofit and Voluntary Sector Quarterly*, *44*(2), 379e400.

Sayogo, D. S., Gil-Garcia, J. R., & Cronemberger, F. (2018), Clarity of roles and responsibilities in interagency information sharing (IIS) projects: determinants and impact on success. *International Journal of Electronic Governance*, *10*(3), 296-316.

Sayogo, D. S., & Pardo, T. A. (2011), *Understanding the capabilities and critical success factors in collaborative data sharing network: The case of dataONE.* Paper presented at the ACM International Conference Proceeding Series.
Selsky, J. W., & Parker, B. (2005), Cross-sector partnerships to address social issues: Challenges to theory and practice. *Journal of Management 31*(6), 849-873.
Smith, H. L., & Dickson, K. (2003), Geo-cultural influences and critical factors in inter-firm collaboration. *International Journal of Technology Management*, *25*(1-2), 34-50. doi: 10.1504/IJTM.2003.003088.
Susha, I., & Gil-Garcia, J. R. (2019 January). A Collaborative Governance Approach to Partnerships Addressing Public Problems with Private Data. In Proceedings of the 52nd Hawaii International Conference on System Sciences.
Susha, I., Gronlund, A., & Van Tulder, R. (2019), Data driven social partnerships: Exploring an emergent trend in search of research challenges and questions. *Government Information Quarterly*, *36*(1), 112-128.
Susha, I., Janssen, M., & Verhulst, S. (2017). Data collaboratives as a new frontier of cross-sector partnerships in the age of open data: taxonomy development, Proceedings of the 50th Hawaii International Conference on System Sciences in Waikoloa, HI.
Susha, I., Zuiderwijk, A., Charalabidis, Y., Parycek, P., & Janssen, M. (2015), Critical factors for open data publication and use: A comparison of city-level, regional, and transnational cases. *JeDEM-eJournal of eDemocracy and Open Government*, *7*(2), 94-115.
Sutherland, M.K., Pardo, T. A., Park, S., Ramon Gil-Garcia, J., Roepe, A. (2018). Cross-boundary information sharing and the nuances of financial market regulation: Towards a research Agenda ACM International Conference Proceeding Series, pp. 133-142.
The Chatham House. (2015). *Data Sharing for Public Health: Key Lessons from Other Sectors*. Retrieved from http//bit.ly/1DHFGVl.
The Gov Lab (2018). Design of data collaboratives. DataCollaboratives.org.
Thune, T. (2011). Success Factors in Higher Education-Industry Collaboration: A case study Of collaboration in the engineering field. *Tertiary Education and Management*, *17*(1), 31-50. doi: 10.1080/13583883.2011.552627.
UNDP & UN Global Pulse (2016). A Guide to Data Innovation for Development – From idea to proof-of-concept. Retrieved on 21 August 2019 from https://www.undp.org/content/dam/undp/library/innovation/UNGP_BigDataGuide2016_%20Web.pdf.
van den Broek, T., & van Veenstra, A. F. (2018), Governance of big data collaborations: How to balance regulatory compliance and disruptive innovation. Technological Forecasting and Social Change, *129*, 330-338.
Van Tulder, R., Seitanidi, M., & Crane, A. S. Brammer. (2016). Enhancing the impact of cross-sector partnerships: Four impact loops for channelling partnership studies. *Journal of Business Ethics*, *135*, 1-17.
Verhulst, S., & Sangokoya, D. (2015), Data Collaboratives: Exchanging Data to Improve People's Lives. Retrieved from https//medium.com/@sverhulst/data-collaboratives-exchanging-data-to-improve-people-s-lives-d0fcfc1bdd9a.
Wohlin, C., Aurum, A., Angelis, L., Phillips, L., Dittrich, Y., Gorschek, T., et al., (2012), The success factors powering industry-academia collaboration. *IEEE Software*, *29*(2), 67-73. doi: 10.1109/MS.2011.92.
World Bank. (2015). *Public-Private Partnerships for Data: Issues Paper for Data Revolution Consultation*. Retrieved from http//bit.ly/1ENvmRJ.
Yang, T. M., & Wu, Y. J. (2014), Exploring the determinants of cross-boundary information sharing in the public sector: An e-Government case study in Taiwan. *Journal of Information Science*, *40*(5), 649-668.
Yang, T. M., & Maxwell, T. A. (2011), Information-sharing in public organizations: A literature review of interpersonal, intra-organizational and inter-organizational success factors. Government Information Quarterly, *28*(2): 164-175.
Zhang, J., Dawes, S. S., & Sarkis, J. (2005), Exploring stakeholders' expectations of the benefits and barriers of e-government knowledge sharing. *Journal of Enterprise Information Management*, *18*(5), 548-567.

Annex

Interview questions

1. Introduction explaining objectives and interview procedure
2. Questions about background of data collaborative
    a. How it was initiated
    b. What was the motivation
    c. What were objectives and expected value
    d. What role was assumed (data provider or user or else)

   e. How was rolled out
   f. Who was driving it
   g. What was the interviewee’s role
3. Questions about success
   a. How successful it was in your opinion
   b. What barriers were encountered (overcome)
   c. Did it meet your expectations
4. Questions about CSFs and prioritization
   a. What were CSFs in your opinion
   b. Which were absolute priority and why
   c. What could have potentially caused failure
5. Summary and conclusions