



Contents lists available at ScienceDirect

## Government Information Quarterly

journal homepage: [www.elsevier.com/locate/govinf](http://www.elsevier.com/locate/govinf)

## Data driven social partnerships: Exploring an emergent trend in search of research challenges and questions

Iryna Susha<sup>a,c,\*</sup>, Åke Grönlund<sup>a</sup>, Rob Van Tulder<sup>b</sup><sup>a</sup> Örebro University, School of Business, Department of Informatics, Örebro 701 82, Sweden<sup>b</sup> RSM Erasmus University Rotterdam, Department of Business-Society Management, Partnerships Resource Centre, Burgemeester Oudlaan 50, 3062, PA, Rotterdam, The Netherlands<sup>c</sup> Section Information and Communication Technology, Faculty of Technology, Policy and Management, Delft University of Technology, Jaffalaan 5, 2628 BX Delft, The Netherlands

## ARTICLE INFO

## Keywords:

Data partnership  
Data collaborative  
Data philanthropy  
Data donation  
Big data  
Collaboration

## ABSTRACT

The volume of data collected by multiple devices, such as mobile phones, sensors, satellites, is growing at an exponential rate. Accessing and aggregating different sources of data, including data outside the public domain, has the potential to provide insights for many societal challenges. This catalyzes new forms of partnerships between public, private, and nongovernmental actors aimed at leveraging different sources of data for positive societal impact and the public good. In practice there are different terms in use to label these partnerships but research has been lagging behind in systematically examining this trend. In this paper, we deconstruct the conceptualization and examine the characteristics of this emerging phenomenon by systematically reviewing academic and practitioner literature. To do so, we use the grounded theory literature review method. We identify several concepts which are used to describe this phenomenon and propose an integrative definition of “data driven social partnerships” based on them. We also identify a list of challenges which data driven social partnerships face and explore the most urgent and most cited ones, thereby proposing a research agenda. Finally, we discuss the main contributions of this emerging research field, in relation to the challenges, and systematize the knowledge base about this phenomenon for the research community.

## 1. Introduction

Opening public data for reuse has been associated with many benefits, including positive impact on societal issues. Governments around the world have made available important datasets which are key for addressing many societal challenges. Poverty reduction, climate change, access to education, and protection against violence are just a few of such challenges. The UN Sustainability Goals<sup>2</sup> provide a prominent example of a joint effort – after 3 years of multi-stakeholder engagement – to set the global agenda for focused efforts to address these challenges. Vital in these recent efforts is the acknowledgement that the solution to many of the world’s ‘grand challenges’ (George, Howard-Grenville, Joshi, & Tihanyi, 2017) require collaborative and coordinated efforts in which the action of non-governmental actors, such as companies and civil society organizations, is equally vital. Strategic datasets crucial for addressing these complex problems are not only held by governments, but also rest in private hands. Companies

around the world, as part of their corporate social responsibility, have also begun to explore opportunities to contribute to addressing societal problems by sharing some of their data. For instance, in the aftermath of the 2015 earthquake NCell, a telecom operator in Nepal, shared mobile call records with data scientists from the non-profit Flowminder in Sweden to help direct disaster response efforts in the area.

This form of collaboration between different actors can be referred to as “data collaboratives” (Verhulst & Sangokoya, 2015). The term itself is new, although the concepts underlying it – data sharing and collaboration – are well known in the digital government research and practice. A “collaborative” is an organized group of people or entities who collaborate towards a particular goal (Wiktionary, 2016). Although a well-founded conceptualization of a data collaborative is lacking, the following definition can give a preliminary idea of the term: data collaboratives are “a new form of collaboration, beyond the classic public-private partnership model, in which participants from different sectors – in particular companies – exchange their data to

\* Corresponding author.

E-mail addresses: [iryana.susha@oru.se](mailto:iryana.susha@oru.se) (I. Susha), [ake.gronlund@oru.se](mailto:ake.gronlund@oru.se) (Å. Grönlund), [rtulder@rsm.nl](mailto:rtulder@rsm.nl) (R. Van Tulder).<sup>1</sup> Present/permanent address: Örebro University, Örebro 701 82, Sweden.<sup>2</sup> <https://www.un.org/sustainabledevelopment/sustainable-development-goals/><https://doi.org/10.1016/j.giq.2018.11.002>

Received 24 August 2017; Received in revised form 14 September 2018; Accepted 11 November 2018

0740-624X/ © 2018 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

create public value” (Verhulst & Sangokoya, 2015).

In 2017 the first repository of data collaboratives to document the growing number of initiatives around the world was launched (see Data Collaboratives Explorer<sup>3</sup>). There is much interest from different global actors to learn and experiment with this form of partnership. For instance, the UN Global Pulse has as its mission to advance the use of big data, including corporate data, for humanitarian and development action. Since 2015 there is an international practitioner conference dedicated to discussions of how to use data responsibly for addressing different societal problems (see International Data Responsibility Conference<sup>4</sup>).

Whereas there is an increasing number of initiatives, academic research lacks systematic insight into this form of partnership. This gap also makes it difficult to assess whether data partnerships can actually become a force for social good or can be abused for different – less social – purposes. Data collaboratives is an ill-defined concept, the novelty of which is not obvious. Stakeholders advocating data sharing for public good often use other terms to label similar initiatives. “Data for good” focuses on the purposes of data sharing and use (Howard, 2012). “Data donations” and “data philanthropy” (Kirkpatrick, 2013) emphasize the act of disclosing data free of charge for a societal cause. In this paper, we propose that an integrative concept can be found to encompass the different elements emphasized by these terms and serve the purpose of distinguishing the specific phenomenon of actors of various kinds working together for the purpose of societal impact from collaboration for individual purposes only. To explore this, we conduct a literature review to map the similarities and differences between the terms. A systematic literature review is needed for several reasons: (1) various overlapping terms are used for different aspects of this phenomenon; (2) this topic is extremely interdisciplinary, involving contributions from business administration, information systems, data analytics, computer science among others; (3) most publications on the topic have appeared in a short period of time following the hype; and (4) there is no previous literature review conducted on this topic. From this point onwards, we will use the term *data driven social partnerships* to refer to the phenomenon of our interest. By choosing this label, we imply a link to an existing concept of cross sector social partnerships (CSSP) on which the literature abounds. CCSPs are understood as voluntary collaborative efforts between organizations from two or more sectors which combine complementary resources to purposefully address complex societal problems (such as environmental protection, economic development, poverty alleviation, health care or education) (Vurro, Dacin, & Perrini, 2010). Cross-sector social partnerships thus put an emphasis on collaborating for social impact, however they do not explicitly consider data as a new driver and resource for such collaboration.

Furthermore, most of the extant studies on cross-sector partnerships have difficulty in defining societal impact in broader and systemic terms (Glasbergen, 2011; Van Tulder & Keen, 2018). From a more practical point of view, enhancing the impact of partnerships is also acknowledged to be dependent on the type of configuration of the partnership and the way progress can be measured (Branzei & Le Ber, 2014; Gray & Stites, 2013; Van Tulder, Seitanidi, Crane, & Brammer, 2015). Cross sector partnerships are often aimed at ‘transformational social’ change but have difficulty in assessing the extent to which that change can actually be achieved. The latter also refers to the search for institutional antecedents of effective partnerships (Vurro et al., 2010) and the impact of the choice of particular configurations of partnerships (cf. Wettenhall, 2003) in terms of multi-stakeholder platforms (Selsky & Parker, 2010). But in particular the degree to which the ‘partnering space’ is created through specific forms of collaboration between public, private, for-profit and non-profit organizations (Van Tulder &

Pfisterer, 2014) defines their effectiveness. The classical PPP literature that has been developed in the public management domain in general does not take the private side of this discourse into account as a separate entity of research. Recent insights have nevertheless been formulated in support of the importance of cross-sector partnering (in particular with Bryson, Crosby, & Bloomberg, 2015; Bryson, Ackermann, & Eden, 2016). The public management discourse is moving to a ‘collaborative governance’ approach in which common ‘goal systems’ can be defined and in which partnerships share collaborative advantages by pooling resources that have a positive bearing on the whole of society (Bryson et al., 2015). This discourse, however, has not really integrated the rapidly developing literature and insights on the private side of cross-sector partnerships: the so-called social partnerships between profit and non-profit actors. These partnerships can be considered from the for-profit side of the partnership (cf. Seitanidi & Crane, 2014, for an overview of this discussion that is shaped by business scholars) or from the nonprofit or social/citizen’s side of the partnerships (cf. for instance Gray & Stites, 2013 for an overview of the discussion that is largely shaped by social movement theory and sociologists).

There is a growing understanding that social partnerships are needed for particularly complex or ‘wicked’ problems (Waddock, Meszoely, Waddell, & Dentoni, 2015) for which individual actors lack the competencies or willingness (Kolk, Van Tulder, & Kostwinder, 2008) to address the complexity of the problem (Pattberg & Widerberg, 2016). Such partnerships therefore can create ‘collaborative advantage’ (Huxham & Vangen, 2004). A vital part of the effectiveness challenge of partnering is formed by the immense lack of data and data sharing with researchers that engage in complexity-sensitive research and monitoring activities (Patton, 2011). But the problem is also affected by low levels of data sharing with practitioners from public and private domains. Classic challenges in particular of public-private partnerships originate in governance problems – largely trust and accountability (Brinkerhoff & Brinkerhoff, 2011) – and a better understanding of the systemic goals for which the partnership is created (Bryson et al., 2016), including the validity of the proposed interventions and the necessary data sharing that is at stake (Babiak, 2009; Liket, Rey-Garcia, & Maas, 2014; Maani, 2017; Patton, McKegg, & Wehipeihana, 2016). Our label – data driven social partnerships – accounts for the additional dimension of collaborating for societal impact while building on the legacy of more traditional partnerships for societal benefit.

The goal of this article is to review the state of the art of research on *data driven social partnerships* by answering the following research questions:

1. What are the core elements of data driven social partnerships?
2. What concepts are used in research to describe this phenomenon, and can an integrative definition be proposed?
3. What are the challenges such partnerships face?
4. What are the main research contributions in the field?

Our literature study systematizes what is already known and what needs further exploration in this emerging field. The expected deliverable is an overview of main results and knowledge gaps and a research agenda for future work. The contribution to research is that this review contextualizes the phenomenon of data driven social partnerships in existing academic research, proposes a well-founded definition, and discusses future research directions. This is a theoretical contribution which can serve to systematize the field. This is needed as the literature stems from several research disciplines and fields and, partly as a consequence of that, uses many partly overlapping concepts. Clearer definitions will help integrate research from different fields. This review is also of value to practitioners, such as parties interested in or advocating for initiating a data driven social partnership, as it extracts and integrates various findings from the disparate body of relevant academic and practitioner literature and thus can be used as a roadmap for

<sup>3</sup> <http://datacollaboratives.org/explorer.html>

<sup>4</sup> <http://www.responsible-data.org>

efforts to advance practices in the field.

## 2. Datafication as a catalyst for new forms of partnerships

The volume of data collected by multiple devices, such as mobile phones, sensors, satellites, is growing at an exponential rate. The term “data revolution” has become a household name used to refer to this development. Data revolution is an explosion in the volume of data, the speed with which data are produced, the number of producers of data, the dissemination of data, and the range of things on which there are data, coming from new technologies such as mobile phones and the internet of things, and from other sources, such as qualitative data, citizen-generated data and perceptions data (IEAGDRSD, 2014). These data may be held by citizens, or by public or private organizations.

To benefit from the explosion of these data, it has to be made available and accessible to allow for data analytics through the processes of data access, use, and reuse. However, for instance in the EU, 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). Accessing and aggregating different sources of data, including data outside the public domain, has the potential to provide insights for problems not envisaged at the point of data collection. Increasingly, official data collected by governments is being complemented by and combined with traditional and big data from the private sector, NGOs and individuals (ODI, 2013). For instance, private sector is increasingly more engaged in ‘smart disclosure’, whereby data about consumer products, companies, services, and consumers themselves is opened up by businesses to foster innovation and enable better purchasing decisions by consumers (Sayogo et al., 2014; Sayogo & Pardo, 2013). Gasco-Hernandez, Feng, and Gil-Garcia (2018) discuss smart disclosure in the context of food traceability and how small farms and institutional buyers can be incentivized to share their data in a way that contributes to food safety, public health, and other societal goals. Besides private sector data, leveraging data about individuals also creates unprecedented opportunities for data science and evidence based policy making. For instance, in 2017 the largest study of human mobility was made possible using the data of 717,527 anonymous users of a smartphone app tracking physical activity (National Institutes of Health, 2017). The study found that more than 5 million people die each year from causes associated with inactivity (Ibid).

Facilitating easier data flows however also requires new forms of organizing. As the data becomes ‘big’, an entirely new ecosystem is emerging comprising new actors moved by their own incentives (Data-Pop Alliance, 2014). There is undoubtedly much research available on information sharing (De Tuya, Cook, Sutherland, & Luna-Reyes, 2017; Gil-Garcia & Sayogo, 2016; Welch, Feeney, & Park, 2016) and cross sector collaboration (Bryson, Crosby, & Stone, 2006; Picazo-Vela, Gutiérrez-Martínez, Duhamel, Luna, & Luna-Reyes, 2017; Vurro et al., 2010) in the digital government domain and beyond; however, the datafication trend adds an *extra layer of complexity* to these partnerships. The evolution of data into big, open, and linked data changes the way governments operate and can transform their functioning and organization (Janssen & van den Hoven, 2015). There are several ongoing shifts in terms of what skills are required to handle data, who should be involved and in what roles, on which conditions data can be shared, and what conclusions can be made and enacted in policies. Because data collection is no longer a prerogative of the government and is very decentralized, data access becomes a negotiation; it creates new hierarchies and inequalities between those who are invited to collaborate and who are not (Boyd & Crawford, 2012).

For public sector organizations data exchange involves a complex social process and critical organizational and managerial capacity (Welch et al., 2016). In fact, governments may be more likely to engage in data sharing collaborations if they have appropriate technical infrastructure and human capital for that (Ibid.). This points to the need

for new or improved capabilities, skills, and resources for engaging in partnerships to leverage data for societal impact. The shortcomings of data and algorithms – such as issues of objectivity, representativeness, privacy – impose an increased demand for transparency and openness on governments too (Janssen & Kuk, 2016). Moreover, the outcomes of algorithmic decision making may not always be positive (Newell & Marabelli, 2015), which may require novel frameworks for risk assessment and mitigation when entering in partnerships around (big) data use.

The nature of societal problems we face nowadays also leaves a mark on how organizations work together. Many of today’s problems are very complex ‘wicked’ problems which often cannot be solved by any single authority in the public sector, such as climate change or refugee crises. Nor can they be solved by other societal actors on their own (Selsky and Parker, 2005; Kolk et al., 2008). The magnitude of such problems is often hard to estimate and the cause-effect relations are complex (Manning & Reinecke, 2016; Van Tulder & Keen, 2018). This means that partnerships aiming to leverage data to address such problems often face a new challenge of ‘breaking down’ the problem in question into feasible and actionable tasks and obtaining relevant information to address a shared goal (Utting & Zammit, 2009). This comes in addition to keeping track of the various phases of the partnership process, that define the degree of trust partners can have in creating an equal and mutual relationship (cf. Glasbergen, 2011; Tennyson, 2010), and to developing shared monitoring and impact measurement (Van Tulder, Seitanidi, Crane, & Brammer, 2016). Partnering processes are also used as a means to navigate relations around societal issues that are often ‘contested’ (Mert & Chan, 2012) and involve unequal relationships (Richter, 2004) and power relations (Ellersiek, 2011). This also has implications on who should be involved in such collaborations and to what effect.

Previous research on cross-sector (public-private) partnerships and inter-organizational collaboration in general does not explicitly focus on the aforesaid challenges in the context of the data revolution. Comparable content-driven systematic literature reviews on cross sector partnerships (cf. Branzei & Le Ber, 2014; Gray & Stites, 2013; Van Tulder et al., 2016) have so far not revealed any relevant studies on the phenomenon of data-driven social partnerships. However, we recognize that there is a solid foundation to build upon when researching how organizations collaborate, including around data exchange for social good. The institutional context thereby dictates the conditions of effective social partnerships (Vurro et al., 2010). More specifically, the institutional context can be identified as consisting of three separate spheres of actors that represent complementary logics, interests, and value propositions (Bryson et al., 2015; Van Tulder & Pfisterer, 2014). Actors from each of these societal spheres need to collaborate and exchange information in order to develop the ‘collective intelligence’ (Patton, 2011; Van Tulder, 2018) that is needed to create a basis of meaningful data creation and exchange. Generally accepted classifications of these societal and institutional spheres are: state, market (firms,) and civil society (social and representative of citizens). Consequently, four types of cross sector partnerships appear: public private (classic infrastructure PPPs) between state and firms, public-nonprofit partnerships (between state and civil society organizations and NGOs), profit-nonprofit partnerships (between companies and NGOs) (cf. Austin & Seitanidi, 2012 for overviews of this particular interaction), and tripartite partnerships that involve all parties. The latter category is generally acknowledged to be necessary to deal with ‘super-wicked’ problems (Levin, Cashore, Bernstein, & Auld, 2012; Warner & Sullivan, 2004), such as climate change for which all relevant societal actors need to engage and share relevant information (Cf. Pinkse & Kolk, 2011, for concrete examples). With each sphere come additional roles and aims of the collaboration. While civil society might want to collaborate for advocacy (Kourula & Laasonen, 2010), corporate-NGO collaboration is often aimed at creating new business propositions for instance to reach unserved markets and needs at the ‘bottom of the

pyramid’ (Cf. [Rufin and Rivera Santos, 2012](#)). Public-private partnerships specifically face a so-called ‘governance paradox’ ([Vangen, 2016](#)) which in short implies that the desire to control and the need to hold each other accountable, in particular triggered by the need for public authorities to be transparent, creates considerable barriers to effectively collaborate ([Brinkerhoff & Brinkerhoff, 2011](#); [Huxham, 2010](#)).

An example of the datasets needed for addressing the type of complex problems that require tripartite partnerships, can perhaps best be illustrated by the experience of the Sustainable Development Goals (SDGs). As already explained in [Section 1](#), the SDGs provide one of the most advanced efforts to create relevant datasets to address complex societal challenges. This effort requires an immense amount of data sharing and data development. For instance, the 17 goals were further elaborated in 169 sub-targets for which more than 230 official indicators were agreed upon of which 150 have more or less well established definitions ([UN, 2015](#)). Most of these indicators have been developed by national statistics bureaus and thus have a considerable macro-oriented bias. Furthermore, when countries started to measure for these indicators, they encountered at least two problems for almost half of the indicators: some of the indicators could not be measured because they were difficult to quantify (which prompted countries to search for different indicators), other indicators were not available in countries (which made it difficult to compare). Interestingly, Dutch policy research shows that the challenge of non-available or measurable indicators is particularly relevant for the more complex or wicked SDG16 (Peace and institutions) and SDG17 (Partnering for the goals) ([Statistics Netherlands, 2018](#)). In these areas a number of data driven partnerships have been initiated, such as between the Bertelsmann Foundation and Sustainable Development Network that developed an SDG Index and Dashboard, which concentrates on international spill-overs, but also identified major indicator and data gaps (around 40) that require further elaboration.

All the above points to the fact that data driven social partnerships is a certain kind of collaboration which faces extreme socio-technical as well as organizational complexity. So far both public and private organizations have been quite cautious about engaging in partnerships to exchange data for societal impact. This notwithstanding, many flagship initiatives exist which pioneer this practice and address diverse societal problems, e.g. disaster response, environment, urbanisation, health-care, education, mobility etc. In this paper, we present a view on these partnerships as a distinct emerging phenomenon and systematize relevant literature to on this issue. Our expected contribution is to map the knowledge landscape and provide a unified view on this form of partnership to help guide further research efforts.

### 3. Research method

To conduct our literature review, we followed the *grounded theory literature review method* formulated by [Wolfswinkel, Furtmueller, and Wilderom \(2013\)](#). The method is particularly suited for reviews aiming to develop a conceptualization of an emerging term. The phenomenon of data driven social partnerships is an emerging one and does not belong to any clear-cut academic niche. It lacks comprehensive concepts and basic theoretical constructs. Therefore, by emphasizing theory development we aim to contribute to scientifically grounding this topic. The method also provides more comprehensive guidance for achieving a better legitimized and thus replicable literature review than the more conventional guidelines by [Webster and Watson \(2002\)](#).

#### 3.1. Data collection

This method consists of five stages, depicted in [Fig. 1](#). Stage 1 is defining the criteria for inclusion, fields of research, outlets and data-bases, and the search terms. Stage 2 is searching. Stage 3 is refining the sample by selecting relevant articles.

The phenomenon of data driven social partnerships is not

**Table 1** Five-stage grounded-theory method for reviewing the literature in an area: to be used in an iterative fashion

Number	Task
<b>1. DEFINE</b>	
1.1	Define the criteria for inclusion/exclusion
1.2	Identify the fields of research
1.3	Determine the appropriate sources
1.4	Decide on the specific search terms
<b>2. SEARCH</b>	
2.1	Search
<b>3. SELECT</b>	
3.1	Refine the sample
<b>4. ANALYZE</b>	
4.1	Open coding
4.2	Axial coding
4.3	Selective coding
<b>5. PRESENT</b>	
5.1	Represent and structure the content
5.2	Structure the article

**Fig. 1.** Grounded theory literature review method of [Wolfswinkel et al. \(2013\)](#).

encompassed by any single research field and is very interdisciplinary. The phenomenon is addressed in different ways using different labels in different fields and projects. Finding commonalities and arriving at unifying definitions would be useful in order to more clearly define the phenomenon and hence be able to more stringently research it. For that purpose, we review articles from several fields based on a number of criteria defining the phenomenon.

To find relevant academic literature, we searched in Scopus and in Google Scholar using the system of keywords displayed in [Table 1](#). These keywords were identified in iterations based on the screening of key literature and snowballing. To locate key literature, we used the repository [DataCollaboratives.org](#) which is a knowledge resource dedicated to the phenomenon. We further used snowballing to identify what other literature is referenced in these papers and which keywords are used there.

Our keyword selection certainly has some limitations. We chose not to use Boolean operators (e.g. data AND philanthropy, data AND collaborative) for two reasons. First, we are interested in whether there is an emerging particular form of partnerships which is labelled and conceptualized in some way. Articles found with Boolean operators,

**Table 1**  
Results of academic literature search.<sup>a</sup>

Keywords	Scopus		Google Scholar	
	Articles found	Articles selected	Articles found	Articles selected
1 “data collaborative”	65	6	89*	3
2 “data philanthropy”	3	1	132	7
3 “data partnership”	14	4	12*	2
4 “data donation”	13	6	7*	2
5 “big data” and “partnership”	9*	1	7*	0
6 “big data” and “collaboration”	18*	1	43*	0
Total	122	19	290	14

<sup>a</sup> As of 6 July 2017.



and not quotation marks, do not contain any specific label/term/concept and are often too broad for us to infer any conceptualization of an emerging phenomenon. Second, using Boolean operators is simply not practical and returned too many results. Also, several other keywords, such as e.g. ‘data sharing’ or ‘cross sector collaboration’, were excluded because they are too broad (and thus generate too many random results) given our interest in a particular type of data sharing and collaboration (aimed at social good).

In Scopus, the search was performed in the title, abstract, and keywords in principle; for the combinations of keywords 5 and 6 the search was performed in the article title only. In Google Scholar, the search was performed anywhere in the article in principle; in cases when it returned over 200 results, the search was done in the article title only (marked with \* in Table 1). In Google Scholar only the first ten pages of results were surveyed. The relevance of the found articles was determined by reading the title and abstract. Our selection criteria were that the article describes, at some level of detail, a partnership or collaboration, as captured by our keywords, which is fueled by data and has a social orientation. In other words, we aimed to select the articles which can help us conceptualize the phenomenon we are interested in. Only articles available in full text were selected. A large portion of the found articles included a comma between the terms (e.g. “data, collaborative”) and thus were excluded on this basis. Also, articles which only mentioned any of the terms, without further discussing them, were excluded as well. In total, we included 33 articles found using this method for our in-depth review.

In addition to the research literature we identified five *practitioner resources* which have the purpose of guiding interested parties to implement a data driven social partnership (labelled in different ways). By practitioner literature we mean non-academic literature, such as reports, guides and working papers, found outside of academic databases. Through desk research, we selected the resources listed in Table 2. These are mainly reports and how-to resources aimed at a wide audience of practitioners, such as policy makers, data advocates, information managers, data analysts in public and private organizations. Our search for these resources was guided by identifying, based on our prior insight into the issue, key institutions which are actively involved in or advocate for datafication for societal benefit (such as UN Global Pulse, The Gov Lab, OECD, among others). The main selection criterion was that the resource provides a conceptualization of the terms used and/or discusses lessons learnt or challenges facing this phenomenon. Additional inclusion criteria were their visibility (number of occurrences when searched for), availability online, and authoritativeness (how often they are referred to). We excluded a potentially large number of reports discussing open data initiatives and the partnerships emerging from that from our review. This is because open data initiatives rest on a different premise: they imply universal openness of data to all and reuse of data for any purpose untargeted to any social issues.

### 3.2. Data analysis

Stage 4 is the analysis which was conducted using Excel by the first

author. The first step was reading all articles, in random order, and highlighting any findings and insights in the text that seem relevant to our research questions. Selecting articles for reading randomly allows for theoretical sampling, i.e. an unbiased approach with an open mind for identifying further concepts and properties. Then, by re-reading the highlighted excerpts, we formulated a set of concepts/categories and meta-insights (*open coding*) which capture a bird's eye view of the findings of the articles. In parallel, we established the interrelations between categories and their sub-categories when this was relevant (*axial coding*). Our final step was to integrate and refine the categories and develop the relations between the main concepts (*selective coding*). All three steps however (open, axial, and selective coding) were performed in an intertwined fashion, going back and forth between papers, excerpts, concepts, categories and sub-categories. This process was performed until the theoretical saturation was achieved, i.e. no more new concepts or interesting links could be identified.

### 4. Findings

In our sample the earliest research using any of the terms in Table 1 is Hale et al. (2003) who discuss “data partnerships” between government agencies from multiple jurisdictions in the context of environmental monitoring. The first academic article which uses the term “data collaborative” is the work of Jonson (2005) which describes the MetroGIS project – a collaboration between geospatial data producers and user communities to assemble, document, and distribute geospatial data in the state of Minnesota. With regards to the term “data donation”, the earliest article in our sample is Weitzman et al. (2011) which describes the case of the TuAnalyze app used for collecting biomedical data from the users for research on diabetes. Finally, the term “data philanthropy” is the most recent and can be attributed to the activities of the UN Global Pulse (Kirkpatrick, 2013).

Our article sample ( $n = 38$ , including the practitioner resources) represents a very eclectic collection of resources. The topic of data driven social partnerships attracts research from a variety of *research subjects* (see Fig. 2). While many research subjects are represented by just one article, several more populated clusters emerge, such as medicine, multidisciplinary studies, and practitioner literature. Research in medical sciences shows a more established tradition of data collaboratives compared to other fields. As a clarification, to identify the disciplines, we looked at the publication outlet of the articles and inferred the research subject based on the outlet title. The category ‘practitioner’ refers to the papers which were not published in an academic outlet (e.g. Hemerly, 2012). Besides the five practitioner resources, we also found other practitioner papers when we searched in Google Scholar.

Our sample also spans various *application domains*: humanitarian, healthcare, international development, education, agriculture, spatial data, statistics. The largest category of articles discuss partnerships in the healthcare domain (12 articles) or in general terms discussing multiple domains as examples (11 articles).

In terms of *methods*, the overwhelming majority of the articles are

**Table 2**  
Practitioner resources selected for review.

Resource	Author	Description
1 Data Collaboratives Guide (2017)	The Gov Lab	A guide on the stages and techniques for establishing an effective data collaborative
2 A Guide to Data Innovation for Development (2016)	UN Global Pulse UNDP	A how-to resource covering data innovation project design phases from idea to proof of concept
3 Data-Driven Development: Pathways for Progress (2015)	World Economic Forum	A report outlining the current landscape, challenges and pathways for progress on big data for development
4 The Data Revolution: Finding the Missing Millions (2013)	Overseas Development Institute	A report setting out a vision for a fully-fledged data revolution with an examination of outstanding challenges
5 Access to new data sources for statistics (2017)	OECD The Gov Lab	A working paper discussing legal requirements and business incentives to obtain agreement on private data access

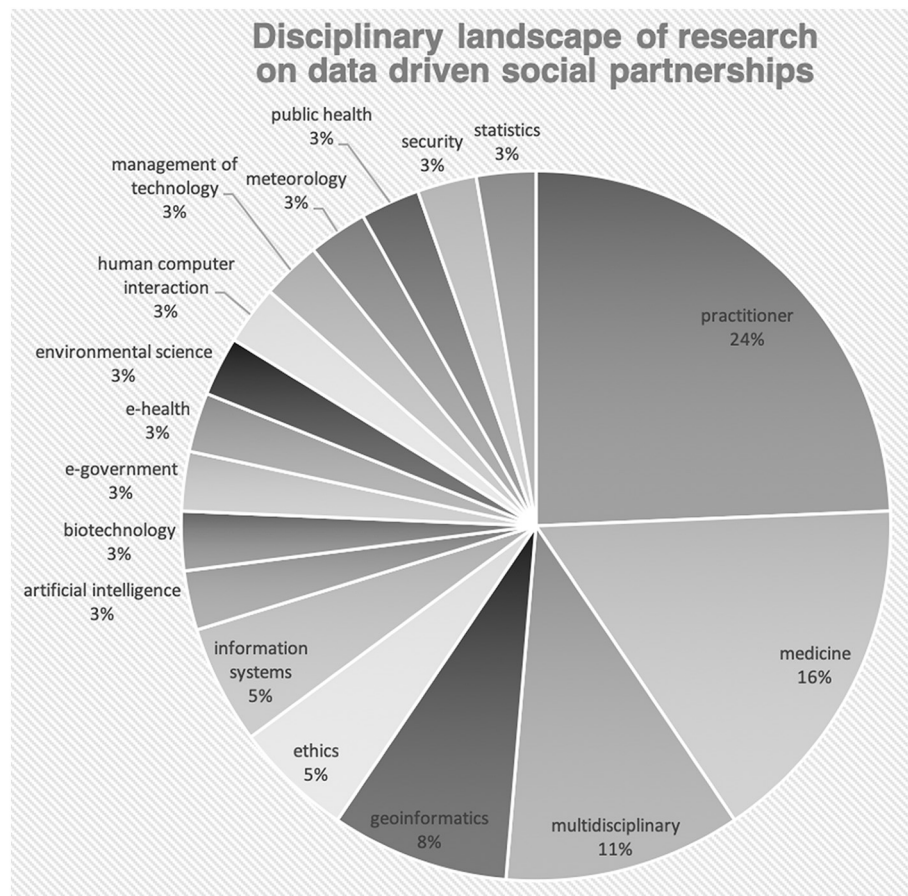


Fig. 2. The disciplinary landscape of research on data driven social partnerships.

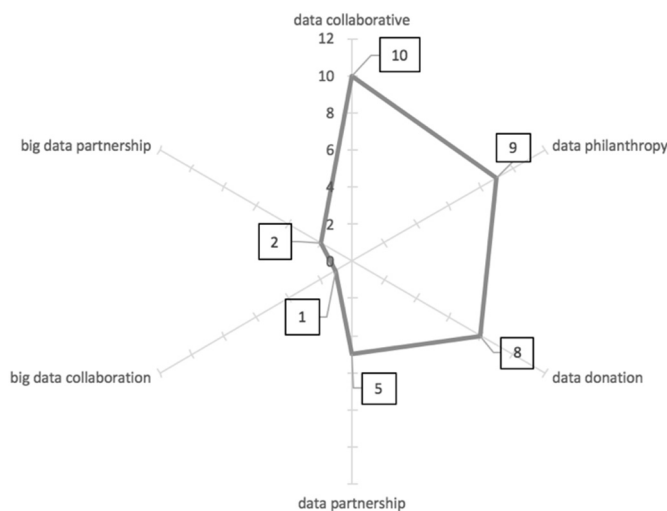


Fig. 3. Concepts used to term data driven social partnerships.

conceptual papers; there are only 8 studies in our sample using empirical methods, such as surveys (Liu et al., 2017; Petersen et al., 2014; Skatova, Ng, & Goulding, 2014), interviews (Taylor & Broeders, 2015; Buda, A (2015)), case studies (Perkmann & Schildt, 2015; Sussha et al., 2017a; Weitzman et al., 2011). This points to a wide gap in evidence-based knowledge on this topic.

Many articles do not explicitly use or elaborate on any of the terms explored here but still discuss the opportunities of using big data from the private sector for advancing science or public good, e.g. in health-care (Hansen, Miron-Shatz, Lau, & Paton, 2014; Schmidt, 2012; Vayena,

Salathé, Madoff, & Brownstein, 2015), peacekeeping (Karlsrud, 2014), agriculture (Kshetri, 2014), human rights (Latonero & Gold, 2015), international development (Lokanathan & Gunaratne, 2015; Taylor & Schroeder, 2015; UN Global Pulse, 2013; World Economic Forum, 2015), transnational politics (Madsen et al., 2016), disaster response (Meier, 2013; Qadir et al., 2016) and others. These articles are not included in the article analyses below, but they do form part of the general understanding of the nature of the field to the extent that they explicitly discuss the collaboration dynamics or data sharing mechanisms involved in such initiatives.

#### 4.1. Integrative conceptualisation of data driven social partnerships

The first research question of our study was concerned with defining the phenomenon: *What are the core elements of data driven social partnerships? What concepts are used in research to describe this phenomenon? Can an integrative definition be proposed?*

To answer these questions, we coded our article sample based on which term was used and how it was conceptualized in the articles. These conceptualizations led us to the formulation of core elements for each term. By comparing these core elements and assessing whether there is common ground, we were able to answer the question about an integrative definition.

Fig. 3 shows the distribution of concepts in use in our sample of articles. The most used term is “data collaborative”, followed by “data philanthropy” and “data donation”.

When coding the articles, we also noted the terms which are used as synonyms or any related terms mentioned in the articles. Our goal was to map the ‘labels’ which are used synonymously to our concepts of interest. In most articles, the authors used one or several synonyms to the main concept (data collaborative, data partnership etc.). Mapping

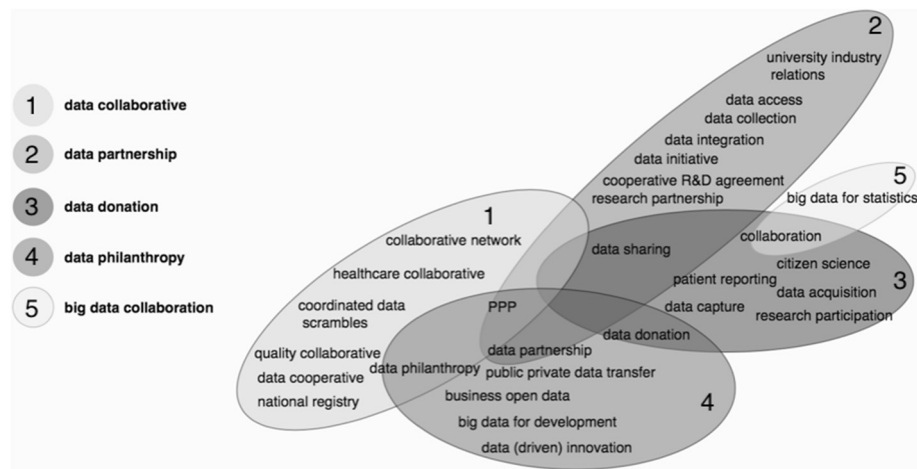


Fig. 4. Vocabulary used in research on data driven social partnerships.

these synonyms forms a vocabulary of terms used in the sample of literature we reviewed. Fig. 4 has five clusters and does not include “big data” and “partnership” from Table 1 because we identified no synonyms in these papers to be included in the vocabulary in Fig. 4.

Fig. 4 visualizes this diverse vocabulary and highlights shared terms and overlaps. It shows that for the most part researchers publishing on the topic of data driven collaboration for social good ‘speak in different languages’ and have also a few shared concepts: public private partnerships, data sharing, collaboration, data partnership, data philanthropy, and data donation. For instance, it shows that data collaboratives can sometimes be referred to as data philanthropy projects, and data philanthropy projects can be referred to as data donations or as data partnerships. However, in most cases (except for articles on data donation and big data collaboration) the shared concept of public private partnerships is emerging.

To take this analysis further, as a next step, we coded the articles by highlighting the definitions or conceptualizations given to any of these terms with the aim of distilling the core elements. In most of the articles in our sample there is no dedicated definition or conceptualization of the term used, other than a description based on the case on which the paper focuses (e.g. a data collaborative in education and its constituting elements).

There is more clarity when it comes to *data donations*: there is consensus among the articles that it is about people donating their data directly for science or other social good free of charge and on a voluntary (consent dependent) basis. There are however different interpretations as to what data is in focus of data donations – primarily personal data or also contextual data, such as for instance data from smartphone apps (Liu et al., 2017) or online transactions (Skatova et al., 2014) which are collected as a by-product. In either case, data donations presume a direct transaction between researchers and users, without the use of commercial apps and involvement of companies as intermediaries collecting these data. Also, although researchers (especially in medicine) are the main recipient of data donations as described in our sample, other actors can receive data donations too, such as public health institutions or disease communities (Weitzman et al., 2011) or even app developers and other innovators (Taylor & Mandl, 2015). All articles on data donations were academic with no relevant practitioner resources identified.

Another relatively well defined and consistent term is *data philanthropy*. Ajana (2017) provides the most comprehensive definition available. The articles focusing on data philanthropy agree that it is about companies donating data (about their customers) for research or social good. Some authors highlight one purpose over the other and phrase it differently, such as to achieve “positive societal impact” (Data-Pop Alliance, 2015) or “enhancement of policy action” (Ajana, 2017)

but the overall meaning is the same. On the other hand, it is not clearly delineated who is the recipient of these donated private sector data: research projects in general (Kirkpatrick, 2013), public sector organizations (Ajana, 2017), or a wider ecosystem of domain-specific practitioners (Buda, A (2015); Taylor & Broeders, 2015). Most authors scope this phenomenon as a data sharing practice by focusing on how companies make data available; with the exception of Ajana (2017) who defines data philanthropy as a form of partnership thus also highlighting the two-way collaboration dynamics. The majority of papers focusing on data philanthropy are in the domain of international development or discuss the term in relation to multiple domains; domain-specific contributions are only two, i.e. in statistics (OECD and The Gov Lab, 2017) and medicine (Ajana, 2017). A large portion of papers on data philanthropy are practitioner literature, with only a few academic contributions (Ajana, 2017; Buda, A (2015); Mir, 2015; Taddeo, 2017; Taylor & Broeders, 2015).

The term *data partnerships* shows very little consistency and was represented by only 5 articles. Except Perkmann and Schildt (2015), all articles focus on intra-sectoral partnerships between public sector organizations at various levels, with a particular focus on initiatives between federal and state agencies. The purpose of these partnerships is typically to integrate disparate data into a centralized data infrastructure to eliminate duplication and fill in gaps. Thus, efficiency is a strong driver of such data partnerships according to our sample, as well as policy improvement (Love et al., 2008; Prescott, Michelau, & Lane, 2016) and research (Love et al., 2008). Exchanging resources, next to data, is mentioned as another activity for data partnerships (Mueller et al., 2009). Of all articles on data partnerships only one was a practitioner paper (Prescott et al., 2016). There is also a variety of application domains described, such as education (Prescott et al., 2016), healthcare (Love et al., 2008), agriculture (Mueller et al., 2009), and environment (Hale et al., 2003). The article of Perkmann and Schildt (2015) using the term “open data partnership” is a special case, since it focuses on university-industry collaboration around access to private sector data by researchers and on opening these data together with research results to the public as well. As explained in the Method section, we did not explicitly include open data initiatives in the scope of our review. However, we find that the term ‘data partnerships’ is sometimes used to refer to collaborations between public organizations at various levels, including those centered on open data.

The term *data collaborative* shows some interesting patterns. It is the most represented category in our sample (10 papers). Most papers describe initiatives either in healthcare, geoinformatics, or across multiple domains. There is just one practitioner resource available using this term (The Gov Lab, 2017). Among these articles there is no consensus about what to term a data collaborative. A working definition of data



collaborative is only provided in most recent literature (Sussha et al., 2017a; The Gov Lab, 2017); the remaining earlier articles only conceptualize this term in relation to the case they describe. Furthermore, data collaboratives can refer to both cross sector (public private) initiatives (Sussha et al., 2017a; The Gov Lab, 2017) and to initiatives mainly between public sector agencies (Byrd, 2011; Priest et al., 2014; Scheich & Bingham, 2015). This can arguably be explained by the evolution of thinking around data collaboratives and the diffusion of this concept beyond the original boundaries of the public sector. However, at the same time, there is a disconnect between prior scientific literature using the term and more contemporary contributions. Similarly, there is a divide in the literature as to which activities a collaborative can focus on: data collection (Scheich & Bingham, 2015), data integration (Astley et al., 2011; Byrd, 2011), data curation and distribution (Johnson, 2005; Masser & Johnson, 2006; Priest et al., 2014), data exchange (The Gov Lab, 2017), or all of the above (Sussha et al., 2017a; Sussha et al., 2017b). However, mostly there is agreement that a data collaborative has a socio-technical nature and requires establishing a data infrastructure on the one hand and a process and an organizational system for collaboration on the other. Van den Homberg (2017) even proposes to consider a more formal institutionalization of data collaborative practices and a long-term timeframe.

It is also interesting to explore if there are any differences in the conceptualization of the terms proposed in academic vs practitioner literature. We find that all five practitioner resources that were included in the analysis focus on the sharing of private sector data but use different terms for that, such as data collaboratives (The Gov Lab, 2017); data philanthropy (Data-Pop Alliance, 2015; Kirkpatrick, 2013; UNDP & UN Global Pulse, 2016); public private partnerships focused on sharing proprietary data (OECD and The Gov Lab, 2017; World Economic Forum, 2015); private data access or data exchange (OECD and The Gov Lab, 2017). Besides, UNDP and UN Global Pulse (2016) use yet another term, “data innovation”, defined as “the use of new or non-traditional data sources and methods to gain a more nuanced understanding of development challenges”. This has to do with the fact that most of these organizations are in the business of advancing development goals in an ecosystem of stakeholders from different sectors.

The articles using a combination of terms “big data” and “collaboration” or “partnership” made up a miniscule portion of our sample (3 articles), therefore we will omit detailed analysis of them. It is only worth noting that, next to the focus on accessing (big) data from new sources (Crump, Sundquist, & Winkleby, 2015; Vale, 2015), big data partnerships can also stand for initiatives to modernize access to (government) data by transferring it to cloud infrastructures of third parties (Ansari et al., 2017).

Having discussed the specifics of each of the term above, we further explore whether there are any common elements used to define more than one term which can contribute towards an integrative definition. Our goal with this is to find out whether these terms refer to different phenomena or whether they can be merged.

Table 3 below gives further insight into the various conceptualizations of the terms found in the sample. These elements were formulated using open coding and grouped into categories by means of selective coding. The articles in the sample were assigned numbers (last column) found in the list of references.

As described above, we find that each of the six terms have a distinct meaning, however, there are several prominent points of contact among them. This allows us to propose an integrative definition of data driven social partnerships as follows. We construct the definition by identifying commonalities and generalizing where appropriate across the terms concerning each of the core elements: actors, activities, object of exchange, purpose, infrastructure, and conditions. These aforesaid elements are the building blocks of our definition. For instance, in the category of actors we propose to generalize towards ‘collaboration between actors in one or more sectors’ to include all mentioned alternatives across the terms (public-public, public-private, involving data

subjects).

Furthermore, Table 3 shows how different content elements were cited across the sample, with the most cited cross-category (occurring in the highest number of papers and in more than one category) highlighted in italics in the first column. Thus, we also made sure that the italicized elements feature prominently in our definition, where appropriate. For instance, we combined ‘data sharing and access’ and ‘exchanging data or resources’ into ‘leveraging data’ (see the definition below). We however excluded the elements of centralized data infrastructure and free-of-charge sharing from the definition, because they are not generic enough to be used to distinguish data driven social partnerships from other types of partnerships (e.g. there may or may not be a data infrastructure for data sharing in a data driven social partnership). Thus, the following is the definition which these steps resulted in:

*Data driven social partnership is a collaboration between actors in one or more sectors to leverage data from different parties, at any stage of its lifecycle, for public benefit in policy or science.*

The benefits of having one defining concept are obvious in a field which spans multiple research disciplines without having a natural home discipline and can be expected to grow and hence requires not only research in general but research that can be inspired, cross-fertilized and compared across disciplines. Because many pressing problems today cannot be solved by government, business and civil society organizations individually, because the increased availability of big data is one key ingredient in solving or managing such problems, and because challenges in the field are many and diverse, as we have shown here, the field we propose to name data driven social partnership is worthy of shared definitions.

#### 4.2. Key challenges for data driven social partnerships

Having proposed an integrative definition, our next step is to answer the second research question: *What are the challenges facing data driven social partnerships?* To answer this question, we used open and axial coding to systematize the challenges mentioned and to create a categorization (Table 4).

We identified 35 challenges in four categories: regulatory, organizational, data-related, and societal. We kept coding the articles, and finding new ones by snowballing and incidental discovery, until saturation was achieved and no new challenges were identified. The categories proposed are for convenience; many challenges span several categories and can be addressed by a combination of legal, technical, or organizational measures.

Table 4 above shows that data-driven social partnerships face a significant number of problems which require further research and action to address. Overall, we observe that the identified challenges to data driven social partnerships concern the supply, as well as the demand sides. On the one hand, there is a lack of incentives, unclear value proposition, and resource constraints for data providers to share data; on the other hand, there is difficult data discovery, lack of communities of practice, and challenging matching of data to problems on the user side, to name a few.

The most cited challenges mentioned by the highest number of authors are (highlighted in italics in the first column): privacy issues; conflicting or lack of appropriate legal provisions; difficult data discovery or costly access; lack of insight into incentives; soliciting participation of data providers; and resource constraints. Two of these challenges can be considered *meta-challenges*, as they were mentioned by all streams of the literature included in our analysis which shows that they are relevant for data collaboratives, data partnerships, data donations, data philanthropy alike. These challenges are difficult data discovery or costly access and conflicting or lack of appropriate legislative provisions. On the other hand, some challenges were mentioned by just one or a few authors from one of the literature streams but these



**Table 3**  
Matrix of core elements used to conceptualize data driven social partnerships.

Core elements	D Coll	D Phil	D Don	D Part	BD Part	BD Coll	References
<b>Actors</b>							
Intra-sectoral: Between government organizations	x	x		x			(Love et al., 2008; Mueller et al., 2009; Scheich & Bingham, 2015; Byrd, 2011; Priest et al., 2014; Prescott et al., 2016)
Cross sector (or public private)	x	x			x		(Sussha et al., 2017a; Ansari et al., 2017; Ajana, 2017; Taylor & Broeders, 2015; The Gov Lab, 2017)
Active engagement of data subjects			x				(Weitzman et al., 2011)
<b>Activity</b>							
Standardized data collection or acquisition	x						(Astley et al., 2011; Love et al., 2008; Scheich & Bingham, 2015; Sussha et al., 2017a; Sussha et al., 2017b; Priest et al., 2014)
Data sharing and access	x	x	x	x	x	x	(Crump et al., 2015; Editorial, 2015; Hale et al., 2003; Liu et al., 2017; Petersen et al., 2014; Shaw et al., 2016; Sussha et al., 2017a; Sussha et al., 2017b; Taddeo, 2017; Vale, 2015; van den Homberg, 2017; Weitzman et al., 2011; Wicks & Heywood, 2014; Skatova et al., 2014; Taylor & Mandl, 2015; Ajana, 2017; Buda et al., 2015; Data-Pop Alliance, 2015; Kirkpatrick, 2013; Hemerly, 2012; Taylor & Broeders, 2015; UNDP & UN Global Pulse, 2016)
Data integration	x			x			(Hale et al., 2003; Love et al., 2008; Byrd, 2011; Priest et al., 2014)
Exchanging data and/or resources	x			x			(Mueller et al., 2009; Vale, 2015; Masser & Johnson, 2006; Ansari et al., 2017; Prescott et al., 2016; The Gov Lab, 2017)
Collaborative data processing	x			x		x	(Sussha et al., 2017a; Sussha et al., 2017b; Vale, 2015)
Discussion and collaboration	x					x	(Scheich & Bingham, 2015; Sussha et al., 2017a; Sussha et al., 2017b; Vale, 2015; The Gov Lab, 2017)
<b>Object of exchange</b>							
(Sharing of) data held by companies	x	x					(Taddeo, 2017; Vale, 2015; Buda et al., 2015; Data-Pop Alliance, 2015; Kirkpatrick, 2013; Hemerly, 2012; Taylor & Broeders, 2015; The Gov Lab, 2017; UNDP & UN Global Pulse, 2016)
(Sharing of) user generated (personal) data	x	x	x				(Editorial, 2015; Liu et al., 2017; Petersen et al., 2014; Shaw et al., 2016; Weitzman et al., 2011; Wicks & Heywood, 2014; Skatova et al., 2014; Ajana, 2017; Hemerly, 2012)
(Sharing of) mined data about persons		x					(Ajana, 2017)
Universal access to data and insights	x	x		x			(Perkmann & Schildt, 2015; Ansari et al., 2017; Buda et al., 2015)
<b>Purpose</b>							
<i>Research purpose</i>							
To address societal problem or for public good	x	x	x				(Editorial, 2015; Liu et al., 2017; Perkmann & Schildt, 2015; Petersen et al., 2014; Shaw et al., 2016; Weitzman et al., 2011; Wicks & Heywood, 2014; Kirkpatrick, 2013)
For innovation or economic purposes					x		(Sussha et al., 2017a; Sussha et al., 2017b; Skatova et al., 2014; Prescott et al., 2016; Ajana, 2017; Data-Pop Alliance, 2015; The Gov Lab, 2017; UNDP & UN Global Pulse, 2016)
To modernize data access and dissemination					x		(Ansari et al., 2017)
To eliminate duplication and satisfy common information needs	x		x	x			(Prescott et al., 2016; Johnson, 2005; Liu et al., 2017; Love et al., 2008; Mueller et al., 2009; Perkmann & Schildt, 2015; Petersen et al., 2014; Scheich & Bingham, 2015; Shaw et al., 2016; Sussha et al., 2017a; Sussha et al., 2017b; Taddeo, 2017; Vale, 2015; van den Homberg, 2017; Weitzman et al., 2011; Wicks & Heywood, 2014; Byrd, 2011; Masser & Johnson, 2006)
<b>Infrastructure</b>							
(Centralized) data infrastructure	x			x		x	(Astley et al., 2011; Hale et al., 2003; Johnson, 2005; Liu et al., 2017; Love et al., 2008; Mueller et al., 2009; Perkmann & Schildt, 2015; Petersen et al., 2014; Scheich & Bingham, 2015; Shaw et al., 2016; Sussha et al., 2017a; Sussha et al., 2017b; Taddeo, 2017; Vale, 2015; van den Homberg, 2017; Weitzman et al., 2011; Wicks & Heywood, 2014; Byrd, 2011; Masser & Johnson, 2006; Love et al., 2008; Vale, 2015; Byrd, 2011; Priest et al., 2014; Ansari et al., 2017)
Organizational structure, capacities, and governance system	x						(Astley et al., 2011; Johnson, 2005; Liu et al., 2017; Love et al., 2008; Mueller et al., 2009; Perkmann & Schildt, 2015; Petersen et al., 2014; Scheich & Bingham, 2015; Shaw et al., 2016; Sussha et al., 2017a; Sussha et al., 2017b; Taddeo, 2017; Vale, 2015; van den Homberg, 2017; Weitzman et al., 2011; Wicks & Heywood, 2014; Byrd, 2011; Masser & Johnson, 2006; van den Homberg, 2017; Priest et al., 2014)
<b>Conditions</b>							
Free of charge (as part of CSR or charity)		x	x				(Editorial, 2015; Perkmann & Schildt, 2015; Taddeo, 2017; Skatova et al., 2014; Buda et al., 2015; Data-Pop Alliance, 2015; Kirkpatrick, 2013; Taylor & Broeders, 2015)
Voluntary participation	x		x				(Johnson, 2005; Liu et al., 2017; Skatova et al., 2014)
Rewards for sharing		x					(Data-Pop Alliance, 2015; Hemerly, 2012)
Long term	x						(van den Homberg, 2017)

**Table 4**  
Categorization of challenges to data driven social partnerships from the literature.

Challenge	Description	References
<b>Regulatory</b>		
1 <i>Lack of consistent and comprehensive legal provisions</i>	Relevant legislation is not up-to-date or specific enough or differs across jurisdictions	The Gov Lab (2017); Taylor and Mandl (2015); Shaw et al. (2016); Hale et al. (2003); Data-Pop Alliance (2015); World Economic Forum (2015); OECD and The Gov Lab (2017); Bellagio Big Data Workshop Participants (2014); Vale (2015)
2 Ambiguous data sharing policies of organizations	Policies lack transparency in terms of how personal data is handled and how it flows to and from third parties	Ajana (2017); World Economic Forum (2015)
3 Lack of clear and accepted ethical guidelines	Existing ethical guidelines are not clear enough and come from multiple sources	Taddeo (2017); Petersen et al. (2014); Data-Pop Alliance (2015)
4 Problem of informed consent of data subjects	Data subjects often give implicit or generic consent to data sharing without knowing how, when, and by whom the data is used exactly	Petersen et al. (2014); Taylor and Mandl (2015); Shaw et al. (2016)
<b>Organizational</b>		
5 <i>Lack of or misalignment of incentives</i>	Data sharing may be counterintuitive	Editorial (2015); Sussha et al. (2017b); Liu et al. (2017); Taylor and Mandl (2015); Skatova et al. (2014); World Economic Forum (2015); OECD and The Gov Lab (2017)
6 Unclear value proposition for data providers	Risk of losing competitive advantage, lack of insight into how value can be created for the data provider	Perkmann and Schildt (2015); Buda et al. (2015); Data-Pop Alliance (2015); World Economic Forum (2015); OECD and The Gov Lab (2017)
7 Lack of coordination of roles, resources, and activities	Coordination may be costly or difficult	Johnson (2005); Sussha et al. (2017b); Vale (2015)
8 Difficulties in collaboration	Achieving an effective collaboration among diverse parties may be challenging	Sussha et al. (2017b); Van den Homberg (2017); The Gov Lab (2017)
9 <i>Low uptake of data providers</i>	Attracting participation of data providers may be difficult	Scheich and Bingham (2015); Editorial (2015); Sussha et al. (2017b); Liu et al. (2017); Weitzman et al. (2011); Taylor and Mandl (2015); Skatova et al. (2014)
10 <i>Resource constraints</i>	Lack of financing	Johnson (2005); Hale et al. (2003); Data-Pop Alliance (2015); World Economic Forum (2015); OECD and The Gov Lab (2017); Ansari et al. (2017); Vale (2015)
11 <i>Difficult data discovery or costly access</i>	Lack of insight into what data is available and how it can be accessed	Van den Homberg (2017); Taylor and Mandl (2015); Shaw et al. (2016); Hale et al. (2003); Love et al. (2008); Kirkpatrick (2013); World Economic Forum (2015); Bellagio Big Data Workshop (2014)
12 Differences in organizational norms, cultures, practices	Participants from different organizations have different practices	The Gov Lab (2017); Hale et al. (2003)
13 Differences in terminologies and frames of reference	Interdisciplinary teams speak different 'languages' which may impede collaboration	Hale et al. (2003);
14 Lack of data stewardship	Data lifecycle in an organization is not monitored by dedicated personnel and through formal procedures	The Gov Lab (2017); Ansari et al. (2017)
15 Lack of communities of practice	Existing communities are fragmented and emerging which impedes learning	World Economic Forum (2015)
16 Fear of losing control and lack of trust	Data may be 'overprotected' due to lack of trust in the data recipient and in the process of sharing	World Economic Forum (2015)
<b>Data-related</b>		
17 <i>Privacy issues</i>	Risk of re-identification of persons exists regardless of anonymization and aggregation	Taddeo (2017); The Gov Lab (2017); Ajana (2017); Mir (2015); Data-Pop Alliance (2015); Kirkpatrick (2013); World Economic Forum (2015); OECD and The Gov Lab (2017); Vale (2015)
18 Data bias	Data may be biased because it is not representative	Scheich and Bingham (2015); The Gov Lab (2017); Data-Pop Alliance (2015); OECD and The Gov Lab (2017); Bellagio Big Data Workshop (2014)
19 Low or uncertain data quality	Data may be of insufficient granularity, have different scope of coverage, range in timeliness which may lead to difficulties to combine data	The Gov Lab (2017); Sussha et al. (2017b); Hale et al. (2003)
20 Data security	Risks of unauthorized access or data leaks	The Gov Lab (2017); Editorial (2015); Ajana (2017); Data-Pop Alliance (2015); World Economic Forum (2015)
21 Risk of flawed data analysis	Risk of data being incorrectly interpreted leading to inadequate conclusions	The Gov Lab (2017)
22 Risk of incongruous data use or misuse	Data insights may be misused on purpose or by mistake which may harm individuals described by the data	The Gov Lab (2017)
23 Matching data with problem	Data may be of limited analytic utility for a certain problem, problem formulation of complex societal issues is difficult	The Gov Lab (2017); Sussha et al. (2017b); Love et al. (2008); Kirkpatrick (2013); World Economic Forum (2015)
24 Lack of appropriate tools, methodologies, expertise	Risk of inaccurate modelling and bias in algorithms, lack of skills and domain expertise	The Gov Lab (2017); Kirkpatrick (2013); World Economic Forum (2015); Vale (2015)
25 Questionable legitimacy of new sources of data	Data outside the public domain may be viewed as not reliable enough by policy makers	Wicks and Heywood (2014); Bellagio Big Data Workshop (2014)
26 Lack of consistency of data and resources	Data may be heterogeneous and resources of parties may be disparate	Hale et al. (2003); Love et al. (2008)
27 Data archival	Maintaining data beyond the lifetime of a project may be difficult due to limited resources and evolving technologies	Hale et al. (2003)
28 Lack of control over data	Once shared, it is difficult to control how data is used by data recipients	Sussha et al. (2017b)
<b>Societal</b>		
29 Measuring impact and value	Measuring direct and indirect benefits and value is complex	Johnson (2005)
30 Data ownership	Companies have full or partial rights to customer data and its commercial use	Editorial (2015); Sussha et al. (2017b); Weitzman et al. (2011); Shaw et al. (2016); Ajana (2017); OECD and The Gov Lab (2017)

(continued on next page)

Table 4 (continued)

	Challenge	Description	References
31	Data divide and exclusion of digitally invisible	'New' data sources do not capture people who have no access or do not use these media which leads to bias	<a href="#">Ajana (2017)</a> ; <a href="#">World Economic Forum (2015)</a>
32	Institutional and political power shift	Role of private sector as data collector about persons increases over that of the state	<a href="#">Taylor and Broeders (2015)</a>
33	Uneven distribution of data analytics capacities	Data analytics expertise is concentrated in developed world	<a href="#">Data-Pop Alliance (2015)</a>
34	Public perception	Public attitudes towards surveillance and privacy have an impact on data sharing initiatives	<a href="#">Buda et al. (2015)</a> ; <a href="#">Data-Pop Alliance (2015)</a> ; <a href="#">World Economic Forum (2015)</a>
35	Implementing interventions based on data insights	Implementing actions based on data insights often encounters a "response gap"	<a href="#">Data-Pop Alliance (2015)</a> ; <a href="#">Kirkpatrick (2013)</a>

challenges are clearly relevant for data-driven social partnerships in general. For instance, such challenges as measuring impact and value of partnerships ([Johnson, 2005](#)), lack of communities of practice ([World Economic Forum, 2015](#)), differences in terminologies of parties from different organizations/domains ([Hale et al., 2003](#)) can be relevant for partnerships of different types – either involving public-public or public-private participants and either involving data integration or data donation activities. This means this overview of challenges can be used for learning across these literature and practice domains and for identifying points of contact for collaboration and knowledge exchange.

To take this overview to the next level, we continued selective coding of the articles to identify how the challenges relate to one another. Fig. A.1 illustrates the relationships which we identified. The cells marked with a star show the challenges which are influenced by several other factors and thereby form *clusters*. We will discuss them in more depth below.

The category of *data-related* challenges is the most populated showing that data driven social partnerships face complex technical challenges. However, articles which mentioned data-related challenges mostly originated from the literature streams of data collaboratives, data philanthropy, and practitioner resources. Many challenges in this category point to one issue of concern – ensuring that the data is analysed in a correct and appropriate manner. The opposite can occur for several reasons: because private sector data most often contains bias (e.g. represents a market share of a certain service provider), the methods or algorithms used for data analytics may be flawed or biased, the data may be of low quality, or simply the data obtained may not be exactly relevant for the problem in question. These challenges, however, are not only relevant to public-private initiatives but also for public-public ones. Data bias may become an issue in data collection or integration initiatives when organizations choose not to contribute their data on grounds of cost or effort required ([Scheich and Bingham \(2015\)](#)). Similarly, the issue of data quality is relevant in a public-public collaboration aiming to integrate data from different sources ([Hale et al., 2003](#)).

Accurate and comprehensive data analysis is related to the other big issue of concern in this category – ensuring that the data analysis is used towards a legitimate and justified purpose. The opposite can happen for several reasons: compromised data security may lead to unauthorized access and misuse of data, data privacy may be compromised leading to re-identification of individuals, flawed data analysis may lead to wrong conclusions and unjustified decisions. Moreover, the question of legitimacy of data is current. Partnerships involving data outside the public domain are more susceptible to this problem, since public data is typically seen as more trustworthy. The legitimacy of 'alternative' sources of data is linked to several problems: to what extent the data can be trusted, how rigorous the data collection process was, how it is possible to verify its representativeness. For public-public partnerships this issue is solved by means of standardized protocols and hierarchical structures thereby ensuring confidence in the data obtained from other public sector parties. In situations where parties from different sectors collaborate – either to access customer transactions or user generated

health data – there are few prior structures for trust building and for creating guarantees of how the data will be used.

The issue of legitimacy of data is linked to a cluster of challenges in the *societal* category. [Wicks and Heywood \(2014\)](#) give an example of clashes in legitimacy between 'old' and 'new' data by describing a case in which patients used the PatientsLikeMe platform to submit their healthcare data which was used by researchers to disprove the effects of a certain medication in opposition to traditional trials and experiments. While it is important to validate that the analysis is accurate, there are also societal implications of making interventions informed by these data analytics. [Taylor and Broeders \(2015\)](#) discuss this in the framework of institutional and political shift of power from state to private sector actors. Besides, there is asymmetry in the geographical distribution of data analytics capabilities ([Data-Pop Alliance, 2015](#)) – often data driven social partnerships involve data scientists from developed countries working on problems in the developing world. In other words, how can actions informed by data describing a limited segment of population be justified, whereas the mandate of governments is to provide services to all in equal manner? Furthermore, designing interventions based on data insights is also complex from organizational, logistical, strategic points of view which often leads to "response gaps" ([Data-Pop Alliance, 2015](#)). This also makes measuring the impact of data driven social partnerships difficult. The lack of dedicated communities of practice only exacerbates this.

Handling personal data also involves challenges from the *regulatory* point of view; the most prominent one is informed consent of data subjects. This challenge is particularly highlighted in the literature on data donations which discusses different forms of consent and the practicalities of obtaining agreement of individuals for the use of their data in research ([Petersen et al., 2014](#); [Shaw et al., 2016](#); [Taylor & Mandl, 2015](#)). This issue however is equally critical in cases of corporate data sharing, because typically individuals as service users give their implicit consent to data sharing when subscribing to the service (e.g. Facebook, Google, Uber etc.). This type of consent cannot be called 'informed', not least in the sense conveyed to this term in research ethics. This problem is complicated by the lack of clear regulatory provisions that are specific to data sharing social partnerships involving personal data. At least as of 2015, the respective EU legislation on data privacy is considered to have many loopholes ([Data-Pop Alliance, 2015](#)).

This lack of clarity of what is and is not allowed when it comes to data sharing affects the dynamics of collaboration and the ease with which organizations are willing to provide access to their data. In the category of *organizational* challenges, these form a cluster of problems. Organizations tend to overprotect their data when they have no incentives to share and see no clear value proposition for 'giving away' their data. The cost or other required resources for sharing data may also be a contributing factor. Next to the aforesaid pragmatic factors comes fear of losing control and potentially compromising one's reputation if the data is of low quality, is leaked or misused. Besides, collaboration between organizations may be complicated by different parties having different rules, practices, cultures, and terminologies

**Table 5**  
Main contributions of research on data driven social partnerships mapped to challenges.

Challenge	Main contributions
Lack of coordination of roles, resources, and activities	• Theoretical coordination mechanisms to address main coordination problems (Sussha et al., 2017b)
Matching data with problem	• Conceptual approach for matching data with problems (Van den Homberg, 2017)
Lack of or misalignment of incentives	• Conceptual factors which are important for incentivizing individuals to donate their data (Editorial, 2015)
	• Survey-based motivations to donate data for scientific research (Liu et al., 2017)
	• Survey-based motivations to donate data to charity (Skatova et al., 2014)
	• Conceptual incentives structuring the sharing of call detail records by companies (Data-Pop Alliance, 2015)
	• Conceptual incentives for corporations to share data within data collaboratives (OECD & The Gov Lab, 2017)
Problem of informed consent of data subjects	• Survey-based contextual factors which influence individuals' consent to donating data for research and how they differ across countries (Petersen et al., 2014)
	• Discussion of different forms of consent and their advantages and disadvantages (Petersen et al., 2014)
	• Proposal to introduce a "data donor" card for donating data after death for research (Shaw et al., 2016)
Lack of consistent and comprehensive legal provisions	• Recommendations to overcome the regulatory vacuum and harness data donations (in the context of health system innovations) (Taylor & Mandl, 2015)
Unclear value proposition for data providers	• Organizational arrangements which are particularly adept at generating productive outcomes while mitigating firms' challenges (Perkmann & Schildt, 2015)
Privacy issues	• Conceptualization of data philanthropy through the prism of infraethics to alleviate the tension with individual rights (Taddeo, 2017)
	• Proposal for a "forward extensible" data sharing approach to counteract privacy risks (Mir, 2015)
Difficult data discovery or costly access	• Decision support framework for companies to decide on opening up their data (Buda, A (2015))
	• Data sharing mechanisms for corporations to share data within data collaboratives (OECD & The Gov Lab, 2017)
	• Conceptualization of data-driven development as informational capitalism (Taylor & Broeders, 2015)
Institutional and political power shift	• Proposal for ethical principles for data philanthropy projects (Data-Pop Alliance, 2015)
Lack of clear and accepted ethical guidelines	• Case-based success factors for public-public partnerships focused on data exchange and integration (Johnson, 2005; Masser & Johnson, 2006)
Failure of projects	• Case-based lessons learnt for the establishment of public-public partnerships focused on data integration (Priest et al., 2014)
	• Case-based success factors for data donation initiatives (Weitzman et al., 2011)
	• Drivers of successful data partnerships in the context of environmental monitoring (Hale et al., 2003)
	• Essential elements of effective federal-state data partnerships in the context of education policy (Prescott et al., 2016)
	• Potential solutions to barriers characterizing public-public partnerships focused on data integration (Love et al., 2008)
	• Taxonomy characterizing different elements and forms of data collaboratives to guide design (Sussha et al., 2017a)
	• Step-by-step recommendations how to design a data collaborative (The Gov Lab, 2017)
	• Guide and tools for designing a data innovation project (UNDP & UN Global Pulse, 2016)
Inception and design of partnerships	

with which they are accustomed. It may be challenging to create a shared understanding in such multi-party, multi-disciplinary teams.

#### 4.3. Main contributions of research on data driven social partnerships

Having outlined the challenges facing data driven social partnerships, our next step is to answer the forth research question: *What are the main contributions in this research field?* To answer this question, we again used open and axial coding to systematize the results achieved in our sample of articles (Table 5). Table 5 below maps the contributions to the challenges from Table 4.

As Table 5 shows, the body of research we collected for our analysis explicitly contributes to 10 out of 35 challenges identified in the previous chapter. Two new challenges were identified (highlighted in italics) not explicitly mentioned as such but instead inferred based on the results of the articles: 1) project failure and 2) inception and design of partnerships, which has a strong link with the formation of a particular configuration of partnerships (PrC, 2012; Tennyson, 2005) In fact, the latter is the *most populated* category offering diverse contributions and perspectives on success factors, drivers, essential elements, and lessons learnt from different kinds of data driven social partnerships. The category of incentives is also well populated with research contributions thereby offering a good initial base of knowledge for addressing this challenge. We discuss these contributions in more detail below.

With regards to incentives, our sample of articles offers insights about motivations of two kinds of actors who may be involved in data driven social partnerships – companies as providers of customer data or individuals as providers of personal data to researchers. Thus, the streams of data donations and data philanthropy literature contribute to this research question the most. In the case of data donations, there is a range of factors which can motivate people to donate their data for research or charity purposes: reputation of the recipient organization

(Liu et al., 2017; Skatova et al., 2014), social signal of knowing how many people already donated (Liu et al., 2017); concern for others (Skatova et al., 2014); direct personal benefit (Skatova et al., 2014); convenience and utility of the data donation tool (Editorial, 2015); and perceptions of data security (Editorial, 2015). There appears to be a difference between the motivations to donate data for science versus for charity, as in the case of the latter the motivation of personal benefit is found to be of influence. This assumption, however, is yet to be tested; as well as the question of whether there are differences in motivations based on which scientific discipline the donation is for (e.g. medicine or digital media research).

In the case of partnerships in which companies share customer data for public good, there are also some advances in the understanding of incentives to do so for companies. The motivations of social signal (knowing that other companies shared data) and self-benefit (pursuing business interests from sharing) are found to be important in this context as well (OECD & The Gov Lab, 2017). Besides, several other incentives come into play in this context: expectations of reciprocal benefits, tapping into external expertise outside the company, enhancing the reputation and visibility of the company, expectations of generating revenue from sharing, improving transparency by sharing data generated for regulatory compliance, philanthropic and socially responsible drivers (Ibid.). However, these incentives have not been empirically tested in scientific studies yet and remain at conceptual level. It is also not yet known how the consent and attitudes of customers determine the motivations of companies as service providers to share data. It is important to note that, based on our sample, research on data driven social partnerships have not paid due attention to the incentives of public sector organizations to share or exchange data. We therefore view this as a research gap waiting to be bridged. In the literature overview (Section 2), this phenomenon was referred to as the 'governance paradox' of cross sector partnerships (Vangen, 2016) which refers to the dual challenge of combining control/accountability



and cooperation at the same time. The partnership literature in general considers this a challenge of the proper sequencing of the partnership in which goal-alignment and adjusted ‘theories of change’ can lead to increasing trust and exchange of more detailed and meaningful sets of data (Van Tulder & Keen, 2018; Patton et al., 2016).

The other relatively ‘rich’ category contains contributions about preventing project failure of data driven social partnerships. Here we find articles which propose lessons learnt or success factors, whichever the term preferred, based on descriptions or analyses of successful cases. In this category, all contributions except one (Weitzman et al., 2011) concern partnerships between public organizations, horizontal or vertical. Weitzman et al. (2011), on the other hand, focus on donations of personal data by individuals and argue that establishing a “research relationship” with participants and providing flexible options for data sharing controls to the users (e.g. sharing with researchers only versus sharing with the user community) are important for the success of these partnerships. We did not find any explicit contribution discussing success factors of other kinds of partnership, such as corporate data sharing initiatives, in our sample of literature. In the case of public-public data driven partnerships for social good the following success factors are found to be important:

- building inter-organizational relationships (Masser & Johnson, 2006);
- stakeholder involvement (Masser & Johnson, 2006), particularly of data stewards (Love et al., 2008);
- open communication (Hale et al., 2003; Priest et al., 2014);
- horizontal organizational structure to build trust (Masser & Johnson, 2006);
- support of local politicians and sense of public purpose (Masser & Johnson, 2006);
- self-interest and motivation of all stakeholders (Masser & Johnson, 2006; Priest et al., 2014) creating a win-win situation (Love et al., 2008);
- structured team with clear responsibilities (Priest et al., 2014);
- formal quality assurance process for data (Priest et al., 2014);
- creating a shared need for and dependence on data (Hale et al., 2003);
- collaborative leadership (Hale et al., 2003);
- committing resources towards a long-term partnership (Hale et al., 2003);
- adopting uniform data standards to make data exchange easier (Hale et al., 2003);
- embracing relevant new technology (Hale et al., 2003).

Many of these factors are also covered in a broader sense in the cross-sector partnering literature with special reference to the importance of formation (PrC, 2012) and governance (Branzei & Le Ber, 2014; Gray & Stites, 2013; Seitanidi & Crane, 2014; Vangen, 2016).

The characteristics of data and data access also influence the degree to which a partnership can be successful. These issues include how secure and privacy-preserving data sharing is, how timely and complete the data is, how easy and flexible data collection, sharing, and use are, how reliable the data is, and to what extent it is combinable with other sources (Prescott et al., 2016). This overview shows that much can be learnt from these initiatives, however all these factors were derived from cases in certain application domains without testing their generalizability. Many of the factors may be relevant for public-private data driven partnerships involving corporate data sharing, such as building relationships, ensuring a win-win situation, committing resources, and having a quality assurance process in place. However, this remains to be tested empirically.

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. These contributions highlight several other factors which were not explicitly

mentioned in the aforesaid articles on success factors for public-public partnerships. This shows that this type of partnerships requires a nuanced approach and faces case-specific complexities. 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 (Ibid.). This is in line with extant partnering research which reiterates the importance of so-called partnership brokers. They can perform three types of information brokering roles: (1) at the formation phase of the partnership, (2) at the continuation and final stages of the partnership and (3) as internal brokers to facilitate data gathering and data sharing within the participating organizations (Manning & Roessler, 2014; Stadler & Probst, 2012; Tennyson, 2005). In addition to this, the resource by The Gov Lab (2017) 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. These recommendations align well with many of the identified challenges to this phenomenon demonstrating high awareness about these challenges in the practitioner community; however, they remain unverified by scientific means.

It is important to note that most of the listed research contributions are either conceptual or case descriptions. Few articles arrive at their conclusions empirically by conducting rigorous case studies or surveys, with a few notable exceptions (Liu et al., 2017; Perkmann & Schildt, 2015; Petersen et al., 2014; Skatova et al., 2014). A portion of articles only describes the results of a certain project (Ansari et al., 2017; Astley et al., 2011; Mueller et al., 2009; Scheich & Bingham, 2015; Vale, 2015) without explicating the lessons learnt beyond this given situation. One may conclude that the field, nascent as it is, abounds in ideas which so far lack firm scientific evidence. More work should be directed towards using empirical research methods to prove or disprove the many propositions circulating in the literature.

## 5. Discussion and conclusions

In this paper, our goal was to provide systematic insight into the phenomenon of data driven social partnerships by conducting a comprehensive knowledge review. This study focused on four questions: (1) what are the core elements of data driven social partnerships; (2) what concepts are used in research to describe this phenomenon, and can an integrative definition be proposed; (3) what are the challenges such partnerships face; (4) what are the main research contributions in the field.

As a result, we identified several concepts which are used in research to describe this phenomenon and proposed an integrative definition of data driven social partnerships based on them. We propose that data driven social partnerships are *collaborations between actors in one or more sectors to leverage data from different parties, at any stage of its lifecycle, for public benefit in policy or science*. The utility of introducing this integrative definition is that it can contribute to better cross fertilization between the multiple research streams dealing with this phenomenon but using different labels for it.

We also identified a list of challenges which data driven social partnerships face and explored the most urgent and most cited ones. Table 4 can be used as an initial research agenda guiding future research in this field. In total, we identified 35 challenges in four categories: regulatory, organizational, data-related, and societal. The *most cited challenges* mentioned by the highest number of authors are: privacy issues; conflicting or lack of appropriate legal provisions; difficult data discovery or costly access; lack of insight into incentives; soliciting participation of data providers; and resource constraints. Two of these challenges can be considered *meta-challenges*, as they were mentioned

by all streams of the literature included in our analysis: difficult data discovery or costly access and conflicting or lack of appropriate legislative provisions. In our analysis, we explored the interplay between the challenges and discussed several emerging *clusters* of problems in each of the categories, such as the risk of flawed data analysis, inappropriate data use, problem of informed consent, difficulties in collaboration. This shows that many problems having to do with data driven social partnerships cannot be tackled in isolation.

Finally, we discussed the main contributions of this emerging research field, in relation to the challenges, and systematized the knowledge base of what is known about this phenomenon. The body of research we collected for our analysis explicitly contributes to 10 out of 35 challenges. Two new challenges were identified: 1) project failure and 2) inception and design of partnerships. In fact, the latter is the *most populated* category offering diverse contributions and perspectives on success factors, drivers, essential elements, and lessons learnt from different kinds of data driven social partnerships. The category of incentives is also well populated with research contributions thereby offering a good initial base of knowledge for addressing this challenge. However, many challenges facing data driven social partnerships remain unaddressed in the literature. As rightly pointed out by Taylor and Broeders (2015), “the shift towards a combination of datafication and privatisation is still in its early stages and the evidence is not yet available to draw conclusions about its medium or longer-term impacts”. Based on our analysis, apart from incentives and design of partnerships, little is known about how the effects and impact of these partnerships can be measured, how risks can be mitigated, how trust can be forged, how citizens as data subjects can be better involved in the process.

The datafication challenge points to a general problem of enhancing the effectiveness and impact of partnerships for delivering social value, public good and/or addressing complex problems that require transformational change such as the SDGs (Van Tulder & Keen, 2018). Data driven social partnerships therefore have to deal with very specific dimensions of the governance paradox, as well as defining what is at stake (Van Tulder et al., 2016). This includes defining the relevant type of data gathering and data exchange that is needed for maximum societal impact. In particular, the governance paradox plays out on at least two levels in data driven social partnerships: related to the configuration (input of the partnership) and the nature of the topic (aim of the partnership). Firstly, the kind of data that can be exchanged depends on the specific interface of the organizations that pool resources in a partnership. It makes a difference whether the actors are public, private, for profit, or non-profit. For further research and policies, it will be relevant to define the configurations of the data driven social partnerships in terms of the societal interface that the participating partners try to combine: state-market (public private partnerships), state-civil society (public non-profit partnerships), market-civil society (profit non-profit partnerships), tripartite partnerships. The challenge for relevant data gathering is to create the right organizational ‘fit’ between the participants and the problem addressed (Van Tulder & Pfisterer, 2014). One of the biggest challenges of partnerships is that the partners are ‘misaligned’ (PrC, 2015). This implies that one of the parties sees

the partnership as a philanthropic activity, whereas the other party sees the partnership as strategic. We have seen that a considerable number of the data driven social partnerships are philanthropic. This creates a considerable risk for the continuity of the partnerships, due to a limited organizational and expectation fit between the participating organizations. Secondly, the topic that the partnership wants to address frames the evaluation of its effectiveness. The more wicked or complex an issue is, the more the partnership should not only be aimed at measuring indicators but also at developing joint indicators for impact in the longer run. Longer run or societal impact questions are in any case difficult to measure, but it has also been found that impact measurement can be done in a meaningful manner at four levels of analysis: individual, organizational, partnering, and societal (Van Tulder et al., 2016). Data driven social partnerships have the potential to exchange data at all these levels, provided the right conditions can be created for a trusted and goal-aligned relationship. The same applies to the challenge of jointly developing more data points. In case of a joint societal goal (such as the SDGs) and a formation process that builds up trust relations (PrC, 2012), organizations can not only become more willing to share data, but also to invest in each other's data collection abilities, thus creating the collective intelligence needed to make the partnership effective also in the longer run. In the monitoring and evaluation literature this is referred to as ‘developmental evaluation’ (Patton, 2015).

The theoretical contribution of this study is primarily to delineate an emerging research field of data driven social partnerships spanning many disciplines for the purpose of being better able to research it. Clearly many major societal problems require cooperation among various kinds of actors and it is valuable to study how and under which circumstances collaborations beyond purely business and government domains occur and what they lead to. For this, sharing of research from many disciplines is needed, and our definition and outline of the field provides a start for such sharing.

The study also offers value to practitioners by recognizing and systematizing a large set of issues pertinent to initiatives in the field. While each solution needs to be situated, any project will benefit from having an overview of issues involved.

For instance, Herala, Vanhala, Porras, and Krri (2016) in their review of how companies share their data as open data found that, although many positive impacts are described in the literature, practice does not follow and companies still refrain from opening their data. They also point out that the majority of the articles they surveyed made assumptions rather than observed these impacts. Our findings are in line with these.

This study draws on a relatively small number of papers and many non-scientific sources. This was a restriction given by the nascence of the field. Further research should be able to expand the evidence base as the field grows.

## Acknowledgements

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”.

## Appendix A

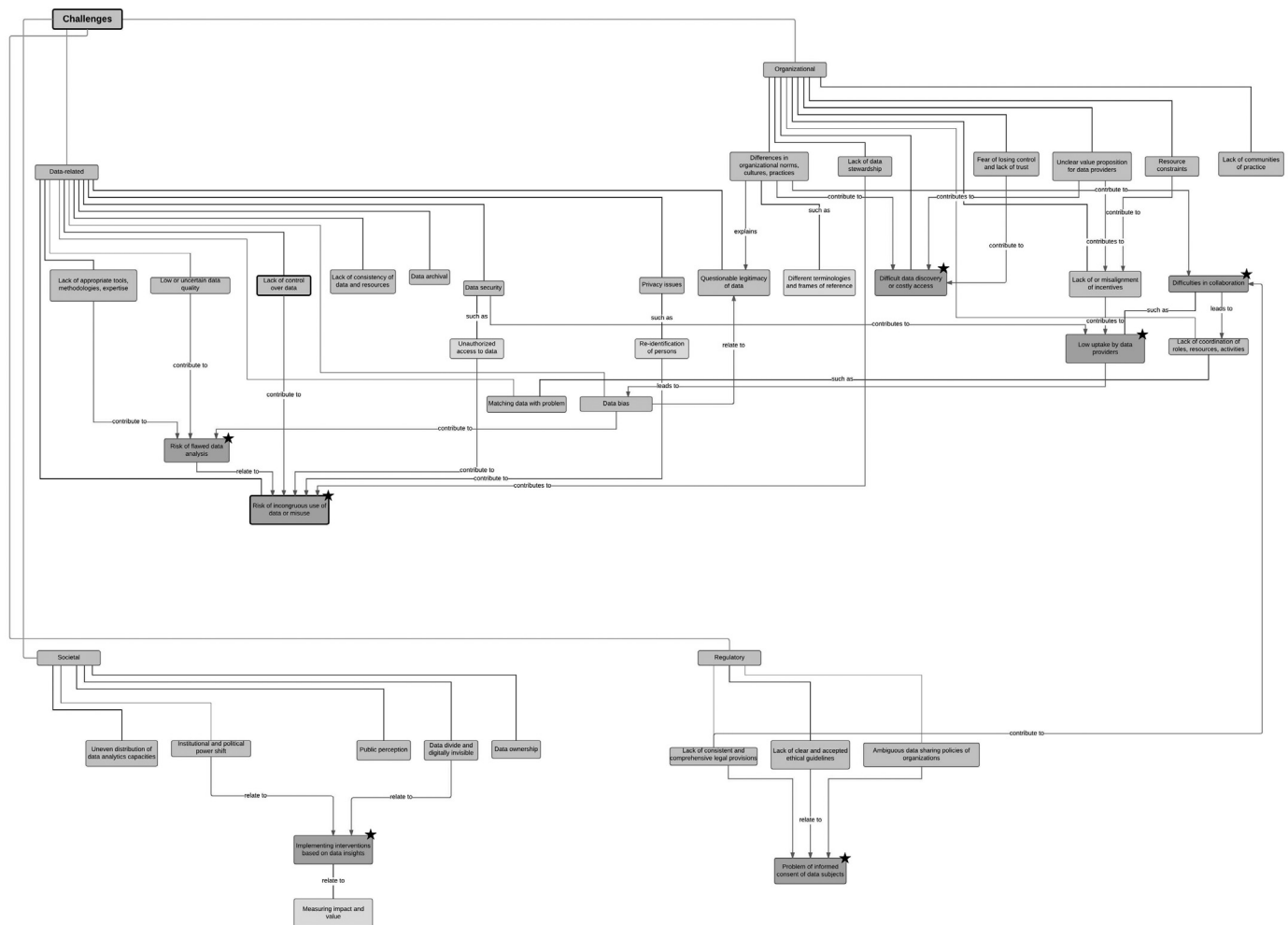


Fig. A.1. Mind map of challenges to data driven social partnerships.

## References

- Austin, J. E., & Seitanidi, M. M. (2012). Collaborative value creation: A review of partnering between nonprofits and businesses: Part I. Value creation spectrum and collaboration stages. *Nonprofit and Voluntary Sector Quarterly*, 41(5), 726–758.
- Babiak, K. M. (2009). Criteria of effectiveness in multiple cross-sectoral interorganizational relationships. *Evaluation and Program Planning*, 32, 1–12.
- Bellagio Big Data Workshop Participants (2014). *Big data and positive social change in the developing world: A white paper for practitioners and researchers*. Oxford: Oxford Internet Institute. Retrieved 21 August 2017 from <http://www.rockefellerfoundation.org/uploads/files/c220f1f3-2e9a-4fc6-be6c-45d42849b897-big-data-and.pdf>.
- Boyd, D., & Crawford, K. (2012). Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon. *Information, Communication and Society*, 15(5), 662–679.
- Branzei, & Le Ber (2014). In M. M. Seitanidi, & A. Crane (Eds.). *Social partnerships and responsible business. A research handbook*. Routledge.
- Brinkerhoff, D., & Brinkerhoff, J. (2011). Public-private partnerships: Perspectives on purposes, publicness, and good governance. *Public Administration and Development*, 31(1), 2–14.
- Bryson, J., Ackermann, F., & Eden, C. (2016). Discovering collaborative advantage: The contributions of goal categories and visual strategy mapping. *Public Administration Review*, 76(6), 912–925.
- Bryson, J., Crosby, B., & Bloomberg, L. (Eds.). (2015). *Creating public value in practice. Advancing the common good in a multi-sector, shared-power, no-one-wholly-in-charge world*. CRC Press: Taylor and Francis Group.
- Bryson, J. M., Crosby, B. C., & Stone, M. M. (2006). The design and implementation of cross sector collaborations: Propositions from the literature. *Public Administration Review*, 66, 44–55. <https://doi.org/10.1111/j.1540-6210.2006.00665.x>.
- collaborative. (2016, March 13). Wiktionary, The Free Dictionary. Retrieved 10:28, April 11, 2016 from <https://en.wiktionary.org/w/index.php?title=collaborative&oldid=37591086>.
- Data-Pop Alliance (2014). Official statistics, big data and human development. Retrieved on 20 August 2017 from <http://datapopalliance.org/item/white-paper-official-statistics-big-data-and-human-development/>.
- De Tuya, M., Cook, M., Sutherland, M. K., & Luna-Reyes, L. F. (2017). Information requirements to create public value: Sharing and opening data to address urban blight. *Transforming Government: People, Process and Policy*, 11(1), 79–98.
- Ellersiek, A. (2011). *Same same but different. Power in partnerships; an analysis of Origins, Effects and Governance*. published doctoral dissertation Tilburg: Tilburg University.
- Gasco-Hernandez, M., Feng, W., & Gil-Garcia, J. R. (2018, January). Providing public value through data sharing: Understanding critical factors of food traceability for local farms and institutional buyers. *Proceedings of the 51st Hawaii International Conference on System Sciences*.
- 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., & Sayogo, D. S. (2016). Government inter-organizational information sharing initiatives: Understanding the main determinants of success. *Government Information Quarterly*, 33(3), 572–582.
- Glasbergen, P. (2011). Understanding partnerships for sustainable development analytically: The ladder of partnership activity as a methodological tool. *Environmental Policy and Governance*, 21(1), 1–13.
- Gray, B., & Stites, J. (2013). Sustainability through partnerships. *Capitalizing on Collaboration, Network for Business Sustainability* case study.
- Hansen, M., Miron-Shatz, T., Lau, A. Y. S., & Paton, C. (2014). *Big data in science and healthcare: A review of recent literature and perspectives*. IMIA Yearbook of Medical Informatics.
- Heralal, A., Vanhala, E., Porras, J., & Krri, T. (2016). Experiences about opening data in private sector: A systematic literature review. *Proceedings of 2016 SAI Computing Conference, SAI 2016, art. No. 7556060* (pp. 715–724).
- Howard, A. (2012). Data for the public good: From healthcare to finance to emergency

- response, data holds immense potential to help citizens and government. *Radar O'Reilly*. Accessed 15 August 2017 from <http://radar.oreilly.com/2012/02/data-public-good.html>.
- Huxham, C. (2010). The challenge of collaborative governance. *Public Management*, 2, 337–357.
- Huxham, C., & Vangen, S. (2004). Doing things collaboratively: realizing the advantage or succumbing to inertia? *Organizational Dynamics*, 33(2), 190–201.
- Independent Experts Advisory Group on the Data Revolution for Sustainable Development (IEAGDRSD) (2014). A world that counts: Mobilising the data revolution for sustainable development. Retrieved 4 August 2017 from <http://www.undatarevolution.org/wp-content/uploads/2014/11/A-World-That-Counts.pdf>
- Janssen, M., & Kuk, G. (2016). The challenges and limits of big data algorithms in technocratic governance. *Government Information Quarterly*, 33(3), 371–377.
- Janssen, M., & van den Hoven, J. (2015). Big and Open Linked Data (BOLD) in government: A challenge to transparency and privacy? *Government Information Quarterly*, 32(4), 363–368.
- Karlsrud, J. (2014). Peacekeeping 4.0: Harnessing the potential of big data, social media, and cyber technologies. *Cyberspace and international relations* (pp. 141–160). Berlin Heidelberg: Springer.
- Kolk, A., Van Tulder, R., & Kostwinder, E. (2008). Business and partnerships for development. *European Management Journal*, 26(4), 262–273.
- Kourula, A., & Laasonen, S. (2010). Nongovernmental Organizations in business and society, management and international business research. Review and implications from 1998 to 2007. *Business & Society*, 49(1), 35–67.
- Kshetri, N. (2014). The emerging role of Big Data in key development issues: Opportunities, challenges, and concerns. *Big Data & Society*, 1(2) (2053951714564227).
- Latonero, M., & Gold, Z. (2015). *Data, human rights & human security*.
- Levin, K., Cashore, B., Bernstein, S., & Auld, G. (2012). Overcoming the tragedy of super wicked problems: Constraining our future selves to ameliorate global climate change. *Policy Sciences*, 45, 123–152.
- Lisbon Council (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>.
- Lokanathan, S., & Gunaratne, R. S. (2015). Mobile Network Big Data for development: Demystifying the uses and challenges. *Communications and Strategies*, 97(1), 75–94.
- Maani, K. (2017). *Multi-stakeholder decision making for complex problems*. Singapore: World Scientific Publishing.
- Madsen, A. K., Flyverbom, M., Hilbert, M., Ruppert, E., et al. (2016). Big Data: Issues for an international political sociology of data practices. *International Political Sociology*. <https://doi.org/10.1093/ips/olw010>.
- Manning, S., & Reinecke, J. (2016). We're failing to solve the world's 'wicked' problems. Here's a better approach. *The Conversation*. Retrieved on 31 January 2017 from <http://theconversation.com/were-failing-to-solve-the-worlds-wicked-problems-heres-a-better-approach-64949>.
- Manning, S., & Roessler, D. (2014). The formation of cross-sector development partnerships: How bridging agents shape project agendas and longer-term alliances. *Journal of Business Ethics*, 123(3), 527–547.
- Meier, P. (2013). Human computation for disaster response. *Handbook of human computation* (pp. 95–104). New York: Springer.
- Mert, A., & Chan, S. (2012). The politics of partnerships for sustainable development. In P. Pattberg, F. Biermann, S. Chan, & A. Mert (Eds.). *Public-private partnerships for sustainable development: Emergence, influence and legitimacy* (pp. 21–44). Cheltenham, Northampton: Edward Elgar Publishing Limited.
- National Institutes of Health (2017). NIH-funded team uses smartphone data in global study of physical activity. Retrieved on 15 July 2017 from <https://www.nih.gov/news-events/news-releases/nih-funded-team-uses-smartphone-data-global-study-physical-activity>.
- Newell, S., & Marabelli, M. (2015). Strategic opportunities (and challenges) of algorithmic decision-making: A call for action on the long-term societal effects of 'datification'. *Journal of Strategic Information Systems*, 24(1), 3–14.
- 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>.
- Pattberg, P., & Widerberg, O. (2016). Transnational multistakeholder partnerships for sustainable development: Conditions for success. *Ambio*, 45, 42–51.
- Patton, M. (2011). *Developmental evaluation. Applying complexity concepts to enhance innovation and use*. New York: The Guilford Press.
- Patton, M., McKegg, K., & Wehipeihana, N. (Eds.). (2016). *Developmental evaluations exemplars. Principles in practice*. New York: The Guilford Press.
- Picazo-Vela, S., Gutiérrez-Martínez, I., Duhamel, F., Luna, D. E., & Luna-Reyes, L. F. (2017). Value of inter-organizational collaboration in digital government projects. *Public Management Review*, 1–18.
- Pinkse, J., & Kolk, A. (2011). Addressing the climate change-sustainable development nexus: The role of multistakeholder partnerships. *Business & Society*, 51, 176–210.
- PrC- Partnerships Resource Centre (2012). *The partnering formation process*. RSM. Rotterdam: Erasmus University.
- PrC-Partnerships Resource Centre (2015). *The state of the partnership report*. RSM. Rotterdam: Erasmus University.
- Qadir, J., Ali, A., Rasool, R., Zwitter, A., Sathiaselan, A., & Crowcroft, J. (2016). *Journal of International Humanitarian Action*, 1(12).
- Richter, J. (2004). Public-private partnerships for health: A trend with no alternatives? *Development*, 47(2), 43–48.
- Rufin, C., & Rivera-Santos, M. (2012). Between commonweal and competition understanding the governance of public-private partnerships. *Journal of Management*, 38, 1634–1654.
- Sayogo, D., Zhang, J., Pardo, T. A., Tayi, G. K., Hrdinova, J., Andersen, D. F., & Luna-Reyes, L. F. (2014). Going beyond open data: Challenges and motivations for smart disclosure in ethical consumption. *Journal of theoretical and applied electronic commerce research*, 9(2), 1–16.
- Sayogo, D. S., & Pardo, T. A. (2013, June). Understanding smart data disclosure policy success: The case of Green Button. *Proceedings of the 14th annual international conference on digital government research* (pp. 72–81). ACM.
- Schmidt, C. W. (2012). Using social media to predict and track disease outbreaks. *Environmental Health Perspectives*, 120(1), 30–33.
- Seitanidi, M., & Crane, A. (Eds.). (2014). *Social Partnerships and Responsible Business. A research handbook*. London: Routledge.
- 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. <https://doi.org/10.1177/0149206305279601>.
- Selsky, J. W., & Parker, B. (2010). Platforms for cross-sector social partnerships: Prospective sensemaking devices for social benefit. *Journal of Business Ethics*, 94(1), 21–37.
- Stadtler, L., & Probst, G. (2012). How broker organizations can facilitate public-private partnerships for development. *European Management Journal*, 30(1), 32–46.
- Statistics Netherlands (2018). *Duurzame ontwikkelingsdoelen; de stand voor Nederland*. The Hague.
- Taylor, L., & Schroeder, R. (2015). Is bigger better? The emergence of big data as a tool for international development policy. *GeoJournal*, 80, 503–518.
- Tennyson, R. (2005). *The brokering guidebook. Navigating effective sustainable development partnerships*. London: The Partnering Initiative.
- Tennyson, R. (2010). *The Partnering Toolkit*. London: The Partnering Initiative. The Partner Broker Association.
- UN Global Pulse (2013). Big data for development: A primer. Retrieved 21 August 2017 from [http://www.unglobalpulse.org/sites/default/files/Primer%202013\\_FINAL%20FOR%20PRINT.pdf](http://www.unglobalpulse.org/sites/default/files/Primer%202013_FINAL%20FOR%20PRINT.pdf)
- United Nations (2015). *Transforming our World: The 2030 agenda for sustainable development*. NY: UN.
- Utting, P., & Zammit, A. (2009). United nations-business partnerships: Good intentions and contradictory agendas. *Journal of Business Ethics*, 90(1), 39–56.
- Van Tulder, R. (2018). *Skill Sheets. An integrated approach to research, management and study*. Amsterdam: Pearson International.
- Van Tulder, R., & Keen, N. (2018). Capturing collaborative complexities – Designing complexity sensitive theories of change for transformational partnerships. *Journal of Business Ethics*, 150(2), 315–332.
- Van Tulder, R., & Pfisterer, S. (2014). Creating partnering space: Exploring the right fit for sustainable development partnerships. *Seitanidi and crane*. 2014. *Seitanidi and crane* (pp. 105–125).
- Van Tulder, R., Seitanidi, M., Crane, A., & Brammer, S. (2015). Enhancing the impact of cross-sector partnerships. Four impact loops for channelling partnership studies. *Journal of Business Ethics*, 105(5), 111–130.
- Van Tulder, R., Seitanidi, M., Crane, A., & Brammer, S. (2016). Enhancing the impact of cross-sector partnerships. Four impact loops for channelling partnership studies. *Journal of Business Ethics*, 105(5), 111–130.
- Vangen, S. (2016). Developing practice-oriented theory on collaboration: A paradox lens. *Public Administration Review*, 77(2), 263–272.
- Vayena, E., Salathé, M., Madoff, L. C., & Brownstein, J. S. (2015). Ethical challenges of Big Data in public health. *PLoS Computational Biology*, 11(2).
- 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>.
- Vurro, C., Dacin, M. T., & Perrini, F. (2010). Institutional antecedents of partnering for social change: How institutional logics shape cross-sector social partnerships. *Journal of Business Ethics*, 94(1), 39–53. <https://doi.org/10.1007/s10551-011-0778-0>.
- Waddock, S., Meszoly, G. M., Waddell, S., & Dentoni, D. (2015). The complexity of wicked problems in large scale change. *Journal of Organizational Change Management*, 28(6), 993–1012.
- Warner, M., & Sullivan, R. (2004). Putting partnerships to work. *Strategic alliances for development between government, the private sector and civil society*. Sheffield: Greenleaf Publishing.
- Webster, J., & Watson, R. T. (2002). Analyzing the past to prepare for the future: Writing a literature review. *Management Information Systems Quarterly*, 26(2), 3.
- Welch, E. W., Feeney, M. K., & Park, C. H. (2016). Determinants of data sharing in US city governments. *Government Information Quarterly*, 33(3), 393–403.
- Wettenhall, R. (2003). The rhetoric and reality of public-private partnerships. *Public Organization Review*, 3(1), 77–107.
- Wolfswinkel, J. F., Furtmueller, E., & Wilderom, C. P. M. (2013). Using grounded theory as a method for rigorously reviewing literature. *European Journal of Information Systems*, 22(1), 45–55. <https://doi.org/10.1057/ejis.2011.51>.

#### List of articles included in the sample: No reference assigned

- Astley, C. M., Tonkin, A. M., Mahar, L. J., Davidson, P. M., Boyden, A. N., Brieger, D. B., Chew, D. P., et al. (2011). Clinical effectiveness in everyday practice: Improving outcomes for all patients through a national acute coronary syndrome data collaborative. *Internal Medicine Journal*, 41(2), 206–210.
- Crump, C., Sundquist, K., & Winkleby, M. A. (2015). Transnational research partnerships: Leveraging big data to enhance US health. *Journal of Epidemiology and Community Health*, 0, 1–2.



- Editorial (2015). Incentivizing data donation. *Nature Biotechnology*, 33(9), 885.
- Hale, S. S., Hale, M. A., Bradley, M. P., Belton, T. J., Cooper, L. D., Frame, M. T., Peterjohn, B. G., et al. (2003). Managing troubled data: Coastal data partnerships smooth data integration. *Environmental Monitoring and Assessment*, 81(1–3), 133–148.
- Johnson, R. (2005). Minnesota MetroGIS geospatial data collaborative Minneapolis – St. Paul metropolitan area (2002 – Enterprise system). *URISA Journal*, 17(2), 41–45.
- Liu, Y., Ferreira, D., Gonçalves, J., Hosio, S., Pandab, P., Kostakos, V., et al. (2017). Donating context data to science: The effects of social signals and perceptions on action-taking. *Interacting with Computers*, 29(2), 132–146.
- Love, D., Rudolph, B., Shah, G. H., et al. (2008). Lessons learned in using hospital discharge data for state and national public health surveillance: Implications for centers for disease control and prevention tracking program. *Journal of Public Health Management and Practice*, 14(6), 533–542.
- Mueller, R., Boryan, C., Seffrin, R., et al. (2009). Data partnership synergy: The Cropland Data Layer. 2009 17th International Conference on Geoinformatics, *Geoinformatics 2009*.
- Perkmann, M., & Schildt, H. (2015). Open data partnerships between firms and universities: The role of boundary organizations. *Research Policy*, 44(5), 1133–1143.
- Petersen, I., Desmedt, C., Harris, A., Buffa, F., Kollek, R., et al. (2014). Informed consent, biobank research, and locality: Perceptions of breast cancer patients in three European countries. *Journal of Empirical Research on Human Research Ethics*, 9(3), 48–55.
- Scheich, B., & Bingham, D. (2015). Key Findings from the AWHONN Perinatal Staffing Data Collaborative. *JOGNN – Journal of Obstetric, Gynecologic, and Neonatal Nursing*, 44(2), 317–328.
- Shaw, D. M., Gross, J. V., Erren, T. C., et al. (2016). Data donation after death: A proposal to prevent the waste of medical research data. *EMBO Reports*, 17(1), 14–17.
- Sussha, I., Janssen, M., Verhulst, S., et al. (2017a). 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 Science (HICSS-50). Waikoloa Village, Hawaii, USA* (pp. 2691–2700).
- Sussha, I., Janssen, M., Verhulst, S., et al. (2017b). Data collaboratives as “bazaars”: A review of coordination problems and mechanisms to match demand for data with supply. *Transforming Government: People, Process and Policy*, 11(1), 157–172.
- Taddeo, M. (2017). Data philanthropy and individual rights. *Minds and Machines*, 27(1).
- Vale, S. (2015). International collaboration to understand the relevance of big Data for official statistics. *Statistical Journal of the IAOS*, 31(2), 159–163.
- Van den Homberg, M. (2017). Toward a Balkans' Data for disaster management collaborative? *Implications of climate Change and disasters on Military Activities* (pp. 11–18). Dordrecht: Springer.
- Weitzman, E. R., Adida, B., Kelemen, S., Mandl, K. D., et al. (2011). Sharing data for public health research by members of an international online diabetes social network. *PLoS One*, 6(4).
- Wicks, P., & Heywood, J. A. (2014). Data donation could power the learning health care system, including special access programs. *American Journal of Bioethics*, 14(11), 27–29.
- Byrd, J. (2011). Education data collaborative. Retrieved 23 August 2017 from [http://www.spnlive.org/spn/media/files/articles/research/EDC%20Research%20Report\\_Final%20Electronic\\_031811.pdf](http://www.spnlive.org/spn/media/files/articles/research/EDC%20Research%20Report_Final%20Electronic_031811.pdf)
- Masser, I., & Johnson, R. (2006). Implementing SDIs through effective networking: The MetroGIS geospatial data collaborative. *GeoInformatics*, 9(6), 50–53.
- Priest, E. L., Klekar, C., Cantu, G., Berryman, C., Garinger, G., Hall, L., ... Masica, A. (2014). Developing electronic data methods infrastructure to participate in collaborative research networks. *EGEMS*, 2(1).
- Skatova, A., Ng, E., & Goulding, J. (2014). Data donation: Sharing personal data for public good? *Proceedings of Digital Economy all Hands meeting*.
- Taylor, P. L., & Mandl, K. D. (2015). Leaping the data chasm: Structuring donation of clinical data for healthcare innovation and modeling. *Harvard health policy review: A student publication of the Harvard Interfaculty Initiative in Health Policy*, 14(2), 18.
- Ansari, S., Del Greco, S., Kearns, E., Brown, O., Wilkins, S., Ramamurthy, M., ... Lakshmanan, V. (2017). Unlocking the potential of NEXRAD data through NOAA's big Data Partnership. *Bull. Amer. Meteor. Soc.*. <https://doi.org/10.1175/BAMS-D-16-0021.1> (in press).
- Prescott, B., Michelau, D., & Lane, P. (2016). *Creating an ideal state-federal data partnership to improve policymaking related to college affordability*. State-Federal Partnerships in Postsecondary Education. Education Commission of the States.
- Ajana, B. (2017). Digital health and the biopolitics of the quantified self. *Digital Health*, 3 (2055207616689509).
- Buda, A., et al. (2015). *Decision support Framework for opening Business Data*. Master thesis, TU Delft. Retrieved on 23 August 2017 from <https://repository.tudelft.nl/islandora/object/uuid:9033ea68-20ee-4acd-8bd4-91f595d82b5f/datastream/OBJ>.
- Data-Pop Alliance (2015). The law, politics, and ethics of cell phone data analytics. Retrieved 21 August 2017 from [http://datapopalliance.org/wp-content/uploads/2015/04/WPS\\_LawPoliticsEthicsCellPhoneDataAnalytics.pdf](http://datapopalliance.org/wp-content/uploads/2015/04/WPS_LawPoliticsEthicsCellPhoneDataAnalytics.pdf).
- Kirkpatrick, R. (2013). Big data for development. *Big Data*, 1(1), 3–4.
- Hemerly, J. (2012). *Policy Colloquium notes: Empowering data-driven innovation-co-convened by Google and O'Reilly Media at Google's Mountain View Campus*.
- Taylor, L., & Broeders, D. (2015). In the name of development: Power, profit and the datafication of the global South. *Geoforum*, 64, 229–237.
- The Gov Lab (2017). Data collaboratives guide. Retrieved 21 August 2017 from <http://datacollaboratives.org/canvas.html>.
- UNDP & UN Global Pulse (2016). A guide to data innovation for development: From idea to proof-of-concept. Retrieved 21 August 2017 from <http://www.undp.org/content/undp/en/home/librarypage/development-impact/a-guide-to-data-innovation-for-development—from-idea-to-proof.html>.
- World Economic Forum (2015). Data-driven development: Pathways for progress. Retrieved 21 August 2017 from [http://www3.weforum.org/docs/WEFUSA\\_DataDrivenDevelopment\\_Report2015.pdf](http://www3.weforum.org/docs/WEFUSA_DataDrivenDevelopment_Report2015.pdf).
- OECD & The Gov Lab (2017). *Access to new data sources for statistics: Business models and incentives for the corporate sector*. OECD Statistics Working Papers, 2017/06Paris: OECD Publishing.
- Mir, D. (2015). Frameworks for companies to Share Data with researchers [Point of View]. *Proceedings of the IEEE*, 103(9), 1439–1444.

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

**Åke Grönlund** is a (full) Professor in informatics at the School of Business of Örebro University in Sweden. His research concerns the use of ICT in various human activities. Key words include information systems, informatics, electronic government, ICT for Development (ICT4D), e-learning, education, mobile technologies and others. The common denominator involved in all projects is to understand how people arrange their work, their organizations and other activities pertaining to private life, such as socializing on the Web, and how ICT can be used to make improvements.

**Rob Van Tulder** is the academic director and founder of the Partnerships Resource Centre. He is also co-founder of the department of Business-Society at RSM (1997), which now is a world leader in issues of sustainability and corporate social responsibility. He coordinates the SCOPE databank project which keeps track of the strategies of the largest multinational companies in the world. His particular interests are in international business, sustainability, issues management and cross sector partnerships. He has published extensively on these topics in academic as well as more popular settings. His numerous books are highly acclaimed in both academic and entrepreneurial circles. His publications have received many awards, while his books have twice been nominated as 'best management' book in the Netherlands. He ranked among the 100 most sustainable Dutchmen. Van Tulder teaches executive courses on International Strategic Management and Sustainability aimed at (executive) managers and academics. He acts as a consultant for international organizations, ministries, international companies and NGOs. He has been visiting professor at several renowned international institutes. Rob van Tulder holds a PhD degree (cum laude) in social sciences from the University of Amsterdam.