The Legitimacy of Data Partnerships for Sustainable Development

Abstract:

This paper examines the legitimacy attached to different types of multi-stakeholder data partnerships occurring in the context of sustainable development. We develop a framework to assess the democratic legitimacy of two types of data partnerships: open data partnerships (where data and insights are mainly freely available) and closed data partnerships (where data and insights are mainly shared within a network of organizations). Our framework specifies criteria for assessing the legitimacy of relevant partnerships with regard to their input legitimacy as well as their output legitimacy. We demonstrate which particular characteristics of open and closed partnerships can be expected to influence an analysis of their input and output legitimacy.

INTRODUCTION

The Global Mobile Industry Association (GSMA), an industry association representing approximately 800 mobile operators, estimates that close to a billion people, mostly in developing countries in Africa and Asia, will come online in the next three years (GSMA, 2016). The evolution in volume and resolution of new private data sources is starting to have a transformational effect on development policies. Development analyses increasingly supplement small-sample survey data with an examination of continuous data sources with large or sometimes close to universal population coverage (Einav & Levin, 2014). These new data sources have received considerable attention in the international development space due to their potential for better monitoring and assessing the Sustainable Development Goals (SDGs) (see e.g., Madsen et al., 2016; Melamed, 2014; Steele et al., 2017).

One of the most recent organizational forms to collect, analyse, and distribute development-related data that rests on these new data sources are *Big Data Partnerships for Sustainable Development* (hereafter "data partnerships"). Such partnerships occur among multiple actors, including, but not limited to: private companies, international organizations, national governments, non-governmental organizations (NGOs), as well as academic institutions (Flyverbom et al., 2017). With the rise of the internet and communication technology (ICT) industries and the concurrent evolution of the SDGs, the hyperconnectivity of the data revolution makes such partnerships a powerful asset that has awakened utopian dreams of it being a new "public commons" (Lohr, 2013). The number of data partnerships with an ambition to address sustainable development has increased markedly. Much hope has been expressed by governments, civil society actors, and businesses for such partnerships to positively impact sustainable development due to their ability to handle data that is characterized by high volume, high velocity, and high variety (Etter et al., 2017; United Nations, 2013; Wesolowski et al., 2015).

Consider the following example of the UN collaborating with Twitter. The UN Global Pulse initiative analyzed large amounts of tweets commenting on the price of rice in Indonesia. The analysis showed that the quantity of tweets on the topic followed the official inflation for the food basket in the country, indicating that social media data can be used as a predictor of price trends on local markets (UN Global Pulse, 2014; UN Global Pulse and Crimson Hexagon, 2011). Such predictors, in turn, can help to ensure food safety. Because Twitter data is sourced and analyzed in real-time, it can provide insights for regions where it is difficult, costly, and time consuming to collect data. Considering the potentially evasive impact of such partnerships, and their subtle institutionalization as a key mechanism to address social development, our main research question is: *What are the conditions that lead to the legitimation and delegitimation of data partnerships for sustainable development?*

Our analysis is motivated by the observation that, because of their increasing relevance for policy makers and international institutions, data partnerships depend on being recognized as *legitimate* tools to collect, analyze, and distribute data in support of sustainable development. Yet, despite their growing relevance for advancing sustainable development, little is known about what legitimizes such collaborative agreements. While the work of international organizations and national governments is legitimized by reference to direct or representative democracy (Scharpf, 1997), data partnerships cannot refer to such democratic mechanisms (e.g., as they include non-elected private actors). Although the literature on multi-stakeholder partnerships has discussed legitimacy (see e.g., Bäckstrand, 2010; Bitzer & Glasbergen, 2015; Mena & Palazzo, 2012), existing results cannot be easily applied in the context of data partnerships. The main difficulty is that data partnerships exist on a continuum between "open" initiatives (in which the data that is used for the analysis and the results are, more or less, freely available) and "closed" initiatives (in which the sourcing of data and the dissemination of

results are restricted). As we show below, the open or closed nature of a partnership influences the way in which initiatives can legitimize themselves.

Our paper makes two key contributions to the emerging discussion of data partnerships in the context of sustainable development. First, we introduce the distinction between open and closed data partnerships and show that both types differ with regard to three criteria (i.e., how membership is regulated, how data is sourced, and how insights are shared). Although this distinction outlines two *ideal* types of data partnerships and therefore obscures some nuances, it helps scholars working in this field to acquire a better understanding of their unit of analysis. As Susha et al. (2017) argue, we need an improved understanding of different types of data partnerships to describe and analyse relevant initiatives in a systematic manner. Second, we develop a conceptual framework to show which criteria are likely to influence the input and output legitimacy of both types of partnerships. Our discussion also demonstrates which particular characteristics of open and closed partnerships *can be expected* to influence an analysis of their input and output legitimacy. Although we show the relevance of our theoretical claims by discussing selected examples, we caution that our analysis is not based on a large-scale empirical assessment of a whole population of partnerships. Nevertheless, our theoretical study can be used as a springboard for future research.

While we develop a framework to analyse the democratic legitimacy of data partnerships, our discussion rests on an acknowledgment that such legitimacy is contextually dependent (Suddaby et al., 2017). Our analysis is focused on data partnerships in the context of sustainable development, and it does therefore not cover data partnerships as a general phenomenon (e.g., when such partnerships are used by companies to gain better consumer insights; McAfee & Brynjolfsson, 2012). We therefore view data partnerships for sustainable development as a distinct unit of analysis, which also differs in various ways from the general notion of cross-sector partnerships for sustainable development (see our discussion below).

Our analysis proceeds as follows. The next section defines big data, as it underlies the discussion of relevant collaborations, and then conceptualizes two ideal types of data partnerships: open and closed agreements among multiple stakeholders. Section three develops a conceptual framework for assessing the input and output legitimacy of open and closed data partnerships. We first develop criteria to judge the input and output legitimacy of relevant partnerships and then show which particular characteristics of open and closed partnerships can be expected to influence the analysis. Section four discusses the implications of our analysis: (a) by showing in how far our framework differs from legitimacy assessments of partnerships for sustainable development that are *not* tied towards big data analyses and (b) by outlining future research challenges that arise from our study, both in terms of possible research topics and appropriate methods.

BIG DATA PARTNERSHIPS FOR SUSTAINABLE DEVELOPMENT

Big Data

Big data is commonly defined as digital data that ranks high on a number of criteria; specifically characterised by large volume, high velocity of collection and processing, and increased variety (McAfee & Brynjolfsson, 2012). Big data is not a homogenous entity but can take a number of manifestations or levels of aggregation; for instance with regard to scale (various sample sizes vs. full dataset), scope (e.g. geospatial, time-series or cross-sectional), and level of analysis (e.g. individual-level data, aggregated data). Established big data sources are satellite data, mobile network data, social media data and data on internet behaviour, the majority of which are increasingly collected and accessed through private firms (Einav & Levin, 2014). The big data paradigm is defined by a high amount of available information flow, information storage, and information processing (Hilbert, 2013). Human-generated telecommunication flows, surveillance cameras, health sensors and the so-called "internet-of-things" are all central parts

of this increasing data stream that is today more easily exchanged in networks (Manyika, et al., 2011). The resulting masses of data, coupled with the emergence of advanced data mining technologies and visualization techniques, constitute the foundation of big data.

Big data makes it possible to render data in many areas where we have not been able to quantify before. Geographical location, words, friendships and sentiments are today datafied, and we are provided with large amounts of data rather than small samples on personal and collective preferences. With big data we are able to collect and analyse massive quantities of real-time information about events that are perhaps less precise than statistics but yet sufficiently precise to inform here-and-now decision-making to improve local conditions. Hence, big data has marked a change in how society processes information and how we think of the world. As a consequence, big data, as pointed out by Meyer-Schönberger & Cukier (2013, p.3), has made us understand that "many aspects of life are more probabilistic, rather than certain."

Data Partnerships for Sustainable Development

Open and Closed Data Partnerships. Based on Waddock's (1991) early definition of cross-sector partnerships, we use the term data partnership to describe *collaborative organizational arrangements in which actors from different sectors attempt to cooperatively address a sustainable development problem through the exchange, processing, and analysis of big data.* A common defining characteristic among data partnerships for sustainable development is the involvement of partners from both the global North and the global South (Bull & McNeill, 2007; Hilbert, 2013). Prior research on cross-sector partnerships has explored partners' motives and strategies and has brought valuable insights to understand how to organize new types of collaborations (Buse & Hamrer, 2007; Austin & Seitanidi, 2012; van Tulder et al., 2016). Also, critical research has studied the challenges that such partnerships face to achieve the social

improvement they set out to address (Selsky & Parker, 2005; Utting & Zammit, 2009; Le Ber & Branzei, 2010). While this research has brought significant knowledge about potentials and challenges of partnerships for sustainable development in general, only limited attention has been given to the particular challenges that big data provides for such partnerships.

Data partnerships for sustainable development change the way in which development-related data is collected, analysed, and distributed. So far, international organizations and national governments primarily use survey-based data to create, monitor, and evaluate development policies (Ginsberg et al., 2009). The UN and the World Bank frequently use household surveys for collecting information on populations in developing and emerging economies (The World Bank, 2004; United Nations, 2005). These surveys provide data on a variety of topics, ranging from poverty to healthcare and education. Samman (2013) argues that such surveys are still the main "workhorse" of data collection for international development. However, she also recognizes that there are limits to survey-based data collection. Surveys have been proven to deliver results that are costly and delayed (Deaton, 2000; McAfee & Brynjolfsson, 2012). Further, surveys are often restricted to the head of the household, making it difficult to gather data on some topics (e.g., in-home violence). Data partnerships address some of these shortcomings (e.g., they usually provide results much quicker; Flyverbom et al., 2017), but they also face problems on their own, some of which are highlighted in this paper.

We distinguish between open and closed data partnerships. This distinction is an analytical one and it reflects two ideal types (Doty & Glick, 1994) of collaborative agreements. It is therefore important to highlight that the underlying difference between open and closed data partnerships is based on a continuum. In practice, the openness of partnerships is less dichotomous and more nuanced, depending on how exactly partnerships behave with regard to the criteria that are outlined below. For instance, the Flowminder/WorldPop partnership follows an open model when it comes to data souring and the sharing of insights, but the collaboration

itself is limited to an agreement between the University of Southampton and the Flowminder Foundation. We base our distinction between open and closed data partnerships on the more general discussion of open/closed data (see e.g., Janssen et al., 2012; Open Data Institute, 2013; Sa & Grieco, 2016). According to Dietrich et al. (2018), open data can be described as "data that can be freely used, re-used and redistributed by anyone – subject only, at most, to the requirement to attribute and share-alike." This definition emphasizes an ideal type of open data – that is, data that can be completely freely used and (re)distributed without any restrictions. However, the openness of data within partnerships is often restricted, also because such collaborative agreements do not reflect fully open systems in which no organizational boundaries would exist.

In order to better understand in what ways data partnerships can differ regarding their openness, we discuss three criteria. These criteria rest on a categorization of insights derived from recent discussions of such partnerships in the literature (see e.g., Madsen et al., 2016; Flyverbom et al., 2017; Susha et al., 2017; Verhulst & Sangokoya, 2015). First, the role of participation matters in the context of data partnerships. In some cases, everyone is able to access and distribute the data that is used and produced in the partnership, while in other cases the collaboration is limited to a few selected partners (Berrone et al., 2016; Susha et al., 2017). This shows that the openness of a data partnership depends on the way in which *membership* is regulated (i.e. on what grounds can organisations be included in the partnership, and who has decision authority?). Second, an initiative's openness is also influenced by the way in which the data, that is necessary for the partnership to work, can be accessed by participants (Madsen et al., 2016; Janssen et al., 2012). In a fully open data partnership, the partners would make their data freely available, so that all participating organizations have unrestricted and equal access to the data that is used to produce insights for sustainable development. This shows that

the openness of a data partnership also depends on the way in which the relevant *data is sourced* (i.e. is data only provided to and from selected organizations or is it an open sourcing process?).

Finally, the openness of data partnerships also depends on whether the results, which are created in and through the partnership, can be freely distributed and re-distributed (Flyverbom et al., 2017; Poel et al., 2015). For instance, Dietrich et al. (2018) highlight that open data implies that "the data must be provided under terms that permit re-use and redistribution including the intermixing with other datasets." This emphasizes that the openness of a partnership is also influenced by the way in which *insights are shared* (i.e. are the resulting insights made openly available to the public or is it only a few organisations that get to apply the information?). Data partnerships are formed not only to enable the baseline sharing of relevant data, but also for joint processing and analysis, and the resulting output in the forms of data and insights are usually validated, analyzed, aggregated, and otherwise modified beyond the source data. Table 1 gives an overview of eight exemplary data partnerships and positions them as open or closed vis-à-vis the three criteria discussed above.

===================

Insert Table 1 About Here

===================

Although both types of data partnerships differ in various ways (e.g. in terms of their way to source data), we believe that it is possible and beneficial to compare them. First, both types reflect partnership arrangements that were created for similar purposes – that is, they reflect collaborative agreements in which actors from different sectors cooperatively address sustainability problems through the exchange, processing, and analysis of big data. Although open partnerships involve more actors and follow a more accessible approach towards data sourcing and analysis, there is still a high degree of cooperation among different partners. While the current literature does not explicitly distinguish between open and closed partnerships, it

still discusses both types as belonging to a common phenomenon: collaborative initiatives for producing and disseminating development data (see e.g., Flyverbom et al., 2017; Susha et al., 2017). Second, we believe that the differences between both types of partnerships (e.g., their scale) makes a comparison vis-à-vis criteria to assess their democratic legitimacy valuable. It is the distinctive nature of both types of partnerships that influences how their legitimacy is either strengthened or impeded. In the following, we further characterize open and closed data partnerships and discuss some of the illustrative examples, which are listed in Table 1.

Characterizing Open Data Partnerships. Open data partnerships designate collaborative agreements in which *membership* is rather less regulated and where data and insights are less frequently restricted by bilateral or network-based agreements. For instance, the Humanitarian Data Exchange (HDX), an increasingly independent project within the UN Office for Coordination of Humanitarian Affairs (OCHA), reflects a partnership where at the time of writing 924 organisations have openly shared over 6,000 datasets and indicators on 224 locations (www.data.humdata.org). The partnership operates under less strict inclusion criteria and hence more liberal access policies, as it is open to all interested organizations (HDX & UN OCHA, 2017; Kessler & Hendrix, 2015). Not much different, the Global Open Data on Agriculture and Nutrition (GODAN) partnership is open to all interested parties as long as the organizations commit to a joint Statement of Purpose (GODAN, 2018).

In terms of *data sourcing*, open data partnerships offer participating organizations the possibility to share/post data that is subsequently used to generate relevant insights. For instance, all members of the Humanitarian OpenStreetMaps (e.g., Google, Planet Labs, the US Humanitarian Information Unit) can contribute to the partnership and have unrestricted access to the data sources (Palen et al., 2015). In this sense, the partnership is based on an open sourcing process of the relevant data. The Flowminder/WorldPop partnership works in a similar way with the partnership being set up between the non-profit organisation Flowminder and the

WorldPop research team at the University of Southampton. The partnership produces geospatial open data on population and demographics in collaboration with organisations such as the Bill & Melinda Gates Foundation, the World Bank, and UN agencies that provide country data (e.g. survey data) and project funding. While the country level or regional data integration and production is project funded, the resulting outputs are made open access and publicly available. In terms of *sharing of insights*, open data partnerships are designed in a way that the generated insights can be shared freely with other organizations or interested individuals. All four exemplary initiatives listed in Table 1 share their insights without many restrictions. While HDX allows to share all created insights within the networks of participating members, the other three initiatives even follow a policy where results can be distributed to any interested organization (and can be also redistributed by these organizations).

Characterizing Closed Data Partnerships. Closed partnerships are collaborative agreements that usually apply formal inclusion criteria and strictly regulate who participates. *Membership* is therefore deliberately regulated. For instance, the partnership between Facebook and the Columbia Center for International Earth Science Information Network (CIESIN) rests on an explicit agreement between the company, Columbia University, selected UN agencies, and International Committee of the Red Cross (ICRC) (Tiecke et al., 2017). Not much different, the data partnership between Twitter and the UN is only restricted to these two players (Crowell, 2016). These data partnerships therefore do not work with an open access policy. This is also reflected in how these partnerships *source* relevant data. The data that underlies these partnerships either comes from bilateral agreements between organizations (under strict contracts and confidentiality) or is shared within a network of organizations or through a trusted intermediary. For instance, the UN-Twitter partnership rests exclusively on anonymized data that is sourced from Twitter. Also, the GSMA Big Data for Social Good partnership only sources data that is provided by telecommunication operators who participate in the initiative.

With regard to *sharing of insights*, the results that are generated by these partnerships are either exclusively accessible to organizations that are part of the initial contractual agreements or the distribution of results to other organizations is tightly regulated. For instance, the Facebook-CIESIN partnership produces population maps with higher resolution than any previous estimates (CIESIN, 2016). These maps allow to gather more accurate information on where people are living, which can be used by selected agencies in the UN system (including The World Bank) when designing development initiatives and disaster response strategies. Although some of the results can be publicly accessed (https://ciesin.columbia.edu/data/hrsl/), the detailed methodology, the input data and deeper insights are retained within the partnership. Not much different, the UNICEF Magic Box partnership distributes the created insights only to selected UN agencies through UNICEF's Office of Innovation.

TOWARDS A FRAMEWORK TO STUDY THE
SOURCES OF LEGITIMACY OF DATA PARTNERSHIPS

The Input and Output Legitimacy of Data Partnerships

Our study views data partnerships as governance arrangements that are in need of democratic legitimacy, regardless of whether they are open or closed. If partnerships are supposed to play an important role within the global sustainable development agenda (United Nations, 2013), we need to have a framework to assess their democratic legitimacy. Although analytically related, our research does *not* focus on the legitimacy of the organizations participating in these partnerships. We aim to develop a framework to study the legitimacy of the partnership itself, even though we acknowledge that potential legitimacy spillover effects between individual participants and the partnership as a whole exist (Haack et al., 2014). Research, which reaches beyond an analysis of the legitimacy of individual participants, remains scarce to date. Rueede and Kreutzer (2015, p. 42) confirm this perspective by stating: "Existing research has been

conducted from the perspective of the participating organizations with the organization as the unit of analysis […] but up to now research on legitimacy in CSSP [cross-sector social partnerships] has fallen short of analysing how the partnership itself gains legitimacy on a partnership level."

In its most general sense, legitimacy refers to the socially shared belief that "the actions of an entity are desirable, proper, or appropriate within some socially constructed system of norms, values, beliefs, and definitions." (Suchman, 1995, p. 574). Democratic legitimacy, more specifically, refers to the acceptance or recognition of authority of a governing body by others (Wolf, 2005). We follow a process-based understanding of democratic legitimacy (Suddaby et al., 2017) and therefore emphasize that legitimacy is not something that data partnerships can simply possess or adopt (e.g., by aligning with dominant social values). Rather, the legitimacy of data partnerships is constantly negotiated between different actors, for instance among those actors that participate in the partnership and among relevant actors in the organizational environment. We deem such a process view to be an appropriate lens to study the legitimacy of data partnerships, as relevant initiatives operate transnationally and are therefore not bound to any particular society or even nation state that could act as a reference point for legitimization (see e.g., Ruef & Scott, 1998). Data partnerships can thus not easily align with an established shared background of community values that exist in their environment. Because such initiatives operate in a transnational environment characterized by a plurality of values and voices, their legitimization depends on *ongoing processes of interactions* among relevant social actors (see also Palazzo & Scherer, 2006).

Based on the literature discussing the democratic legitimacy of transnational governance arrangements (Bäckstrand, 2010; Risse, 2004; Scharpf, 1997), interactive processes of legitimization can have two distinct reference points. On the one hand, it is possible to look at the input side of relevant legitimization processes to study whether data partnerships are

perceived as legitimate in terms of the design of underlying communicative structures. Prior research has referred to such a perspective as *input legitimacy* (Scharpf, 1997). On the other hand, it is possible to look at whether the outputs that arise from relevant interactions are seen to effectively solve those problems that are targeted by the partnership. Prior research has referred to such a perspective as output legitimacy (Mena & Palazzo, 2012). Whereas input legitimacy is concerned with governance by the people, output legitimacy is concerned with governance for the people. The distinction between input and output legitimacy is crucial, because partnerships with a high participatory quality do not necessarily need to be effective in terms of addressing a certain issue.

We first develop criteria for assessing the input and output legitimacy of data partnerships. We then demonstrate why each criterion is relevant and which particular characteristics of open and closed partnerships can be expected to influence the analysis. Table 2 summarizes our analysis.

===================

Insert Table 2 About Here

===================

Input Legitimacy

Input legitimacy is concerned with the question of whether the authentic preferences of those who are affected by a governance arrangement are actually represented in the arrangement in some form. It refers to the belief that "decisions are derived from the preferences of the population in a chain of accountability linking those governing to those governed" (Mayntz, 2010, p. 10). It thus mostly deals with the decision-making processes and governance structures that underlie data partnerships. We suggest three criteria to judge the input legitimacy of data

partnerships: *the existence of participatory decision-making, the transparency of the partnership,* and its *deliberative quality.* As our paper aims at opening a debate for the evaluation of the legitimacy of data partnerships, we do not argue that this list of criteria is necessarily exhaustive.

Participatory Decision-Making. Participation is a key dimension of input legitimacy. It is widely accepted within the literature that legitimate governance arrangements need to include different stakeholder voices; ideally all those affected by relevant decisions (Bäckstrand, 2010; Dingwerth, 2007, Mena & Palazzo, 2012; Wolf, 2005). Theoretically speaking, the idea that higher levels of stakeholder inclusion positively affect input legitimacy rests on deliberative democracy thinking (Habermas, 1998; Young, 2000). This yields a number of important questions in the data partnership context: What are the patterns of exclusion and inclusion when partnerships make decisions? Are there some actors that dominate the partnership? Are there barriers to participation? In the data partnership context, participation refers to whether those organizations, which participate in a partnership and are affected by it, can also participate in relevant decisions (e.g. which sustainability issues to focus on). High levels of participation do not *ipso facto* yield high levels of input legitimacy, as the quality of the underlying deliberations are also important (see below).

Participation is a relevant assessment criterion for data partnership's input legitimacy, as it connects the scale of participation in a partnership to discussions around representation in decision-making structures. At first glance it seems that since open data partnerships have many participating organizations from different societal sectors and countries, it can be expected that such set-up favours the creation of inclusive participatory structures. For instance, HDX acknowledges that "[a]ll user groups will be considered when designing solutions, products and services." (UN OCHA, 2016 p. 3) However, in practice such inclusiveness is limited by the large-scale nature of most collaborative agreements. Research in participatory decision-making

has shown that organizational structures, which contain a high number of heterogenous members, find it harder to set up high-quality engagement structures (Börzel & Risse, 2005; Boström, 2006; Goodin, 2000). An increase in the organizational complexity that underlies the partnership can therefore imply (a) that it becomes more difficult to couple participants and decision-makers, (b) that possibilities for participation are limited, and (c) that it becomes more difficult to move participation towards justification and beyond a mere expression of opinions (Parkinson, 2003; Pingree, 2006; Walzer, 1999).

Participation in open data partnerships can therefore be expected to function mostly through advocates and hence indirect representation. For instance, Humanitarian OpenStreetMaps has chosen to create participatory structures through the election of voting members – these are "people in the HOT [Humanitarian OpenStreetMap Team] community who have shown commitment to the HOT mission." (HOTOSM, 2018). Not much different, the GODAN partnership works with a Steering Committee consisting of selected representatives. While such representatives can ensure a certain level of inclusiveness, open data partnerships usually involve heterogenous sets of actors (e.g. organizations of different sizes, cutting across societal sectors and countries), which change frequently and rapidly (due to low entry barriers). These conditions can limit the degree of representation, which can be achieved through advocates, as it can make the identification of adequate representatives more difficult.

Closed data partnerships have fewer organizational participants and usually depend on contractual agreements between a few actors. This puts limits on their ability to involve a representative set of participating actors when making decisions, especially as many closed data partnerships do not have formal governance structures (e.g., the Facebook-CIESIN partnership lacks formal governance and appears to be managed on a contractual basis; CIESIN, 2016, 2018; Tiecke, 2016). Closed data partnerships can therefore either adopt a strategy of setting

up structures for stakeholder representation in their decision-making (which would enhance their perceived legitimacy) or to limit participation to just include those organizations that actually contribute to the collaboration (which would impede their legitimacy). The latter option can be expected to favour a homogeneity of ideas, as small group deliberation often impedes the deliberative ideal that numerous views and opinions are expressed and considered (Ryfe, 2005).

Transparency. We understand transparency as a partnership's ability to make its "behaviour and motives readily knowable to interested parties" (Hale, 2008, p. 75). It covers access to information by stakeholders and also the general disclosure of relevant information. For data partnerships two aspects of transparency are particularly important: (1) transparency around how data was sourced, analysed, and distributed and (2) transparency around the decision-making structures that the partnership agrees to. For data partnerships this yields interesting questions, such as: What are the levels of transparency that the partnership agrees to in terms of sourcing and analysing data? What accountability structures are created? What information is accessible and by whom is it accessible? Transparency can be expected to facilitate stakeholder participation and hence strengthen democratic governance and accountability (Christensen & Schoeneborn, 2017; Scharpf, 1997). If important aspects concerning data partnerships (e.g. how decisions are reached) are transparent, affected stakeholders can better judge whether their preferences were respected. Whether or not data partnerships can be viewed as appropriate providers of insights into sustainable development challenges thus also depends on whether it is possible to evaluate their activities from the outside (Bernstein & Cashore, 2007; Black, 2008).

The openness of data partnerships can impact in how far relevant initiatives are transparent about how they source, analyse, and distribute development data. While the open data literature usually assumes that open data implies higher levels of transparency (Janssen et

al., 2012; Sa & Grieco, 2016), we caution that this does not need to be the case in the context of data partnerships. Although the push towards open data within government has resonated with calls for more transparency and accountability, one can have open data initiatives without much transparency (Lourenço, 2015). Just because data partnerships offer participating organizations the possibility to provide, access, use and share development data, they do not automatically have to enjoy high levels of transparency. First, the mere availability of open data says little about where the data comes from exactly (e.g. it could have been reposted) and how the analysis was performed (e.g. it is common that different datasets are combined within analyses; Hilbert, 2013). For instance, although most datasets provided by participating organizations in HDX are transparent, data providers cannot always be held accountable to insights created based on their data, as datasets are usually mixed within the analysis. Second, even if a data partnership is transparent regarding how data was sourced and analysed, it can still lack transparency regarding its governance structure. If key decision-making processes (e.g. how it was decided what sustainability issues the partnership wants to address) remain opaque, it becomes harder for outside audiences to judge whether the partnership really can be seen as an appropriate way to address sustainable development problems (see also Schouten et al., 2012).

Deliberative Quality. The input legitimacy of a data partnership also depends on the quality of the communicative processes that underlie participation. Procedural demands are usually conceptualized as the *deliberative quality* of governance arrangements (Bäckstrand, 2010; Ryfe, 2005). Such procedural elements are important, as data partnerships can be inclusive in terms of involving relevant parties but may not have processes in place that guarantee the fairness of argumentative procedures. The deliberative quality of data partnerships can be assessed by asking questions such as: Do participating stakeholders have a valid voice in decision-making processes? To what extend does a partnership live up to the

ideal of an unconstrained dialogue? Are participating actors able to achieve a consensual solution to the problem they are addressing? The deliberative quality of data partnerships is vital, as participating organizations usually come from different societal sectors and hence need to move beyond "bargaining solutions" that are likely to be influenced by power differences (Schouten et al., 2012).

Deliberative theorists have focused on how deliberation ought to look like and what influences the successful creation of action-guiding consensus (see e.g., Dryzek, 2010; Karpowitz & Mendelberg, 2007; Risse, 2004). Two issues are particularly relevant in the context of data partnerships. First, deliberative quality is influenced by whether power differences between actors can be neutralized as much as possible (Habermas, 1998). Power, here, can be understood either as direct coercion of other actors or as the manipulation of what counts as the accepted boundaries of argumentation (Fleming & Spicer, 2014). While power differences will occur in open and closed partnerships, we can expect that the visibility and consequences of unequal power relations are particularly relevant in the context of closed partnerships that are based on bilateral agreements (e.g. Twitter and UN Global Pulse; Facebook and CIESIN). In such partnerships, there is a strong dependence on single actors, which makes it more difficult to neutralize power differences between the participating organizations. As Emerson (1962, p. 32) suggested: "power resides implicitly in the other's dependency." In such a situation, it will be harder to ensure that argumentative rationality rests on being persuaded by the better argument (Habermas, 1990; Hale, 2008). Rather, it is more likely that "argumentative power" is replaced with "bargaining power" in the sense that material resources and the threat to exit the partnership influence the mode of communication (Risse, 2004).

Second, the scale of deliberations can have an influence on the ability of participants to create action-guiding shared knowledge structures (Albrecht, 2006). Since deliberative theory

assumes that there is a direct exchange of arguments between different groups (Mansbridge et al., 2012), a high number of participants makes it more difficult to create argumentative procedures that result in a common understanding of problems and solutions (Goodin, 2000; Parkinson, 2003). Prior research has observed that participants in small-scale deliberations often find it easier to detect information or cues that enables them to create shared knowledge and eventually consensus (Ryfe, 2005). We can therefore expect that even though closed data partnerships may find it harder to set up deliberative processes in which numerous views and opinions are expressed, they should find it easier to create a framework for identifying common knowledge among participants. However, research has also shown that the existence of such consensual agreement can impede the conditions for future dialogue that legitimizes relevant decisions (e.g. because it may hamper future dialogue to reconsider or update decisions; Friberg-Fernros & Schaffer, 2014).

Output Legitimacy

Output legitimacy refers to the problem-solving capacity of a partnership and the results it creates; it couples the legitimacy of a governance arrangement to its ability to effectively address the problems that it claims to address (Levine et al., 2005; Wolf, 2005). Output legitimacy links legitimacy to the consequences of deliberations and therefore complements input legitimacy's focus on elements of institutional design. One key strength of assessing the democratic legitimacy of data partnerships in this way is that it acknowledges that although some initiatives may show good degrees of input legitimacy, their problem-solving capacity might be disappointing (*and vice versa*). Based on insights from the partnership literature (Bekkers & Edwards, 2016; Schouten et al., 2012), we believe that two dimensions of output legitimacy are particularly relevant for judging the legitimacy of data partnerships: their outputs and their outcomes. Data partnerships' outputs refer to the insights on sustainability problems

that are created, while their outcomes refer to the observed effects that these insights make to address the problem that motivated the creation of the collaboration (Gulbrandsen, 2005; Young, 2014).

Outputs. The output dimension puts the analytical focus on whether data partnerships have the required activities and resources in place to produce relevant outcomes (Bäckstrand, 2010). In other words, the focus is on whether the interactions among participants actually produce insights that help to address sustainability problems. For instance, a data partnership that provides insights to policy makers without following up on whether these insights help to shape decisions would not show a very high problem-solving capacity. In order to judge whether the outputs of a data partnership strengthen or limit its output legitimacy it is necessary to focus on questions such as: Does the partnership have the required resources in place to create relevant insights to policy makers? Does the partnership have feedback mechanisms in place to adjust its outputs? Are the created insights really used by others (e.g. to change policies)? This puts the focus on (a) how data partnerships produce insights and (b) whether these insights can be potentially translated into problem-solving outcomes (Mena & Palazzo, 2012). While a partnership may produce a lot of insights into sustainable development issues, these insights may be inefficacious with regard to creating outcomes that benefit people or the natural environment. Alternatively, partnerships may produce viable insights, but international organizations or national governments may not use these insights (e.g. because they distrust big data as a source of expert knowledge; Flyverbom et al., 2017).

What kind of outputs are created by open and closed data partnerships remains an empirical question. However, we can expect that the capacity of both types of partnerships to create non-predetermined outputs differs. Open data partnerships allow interested organizations to use the provided data to create insights that they deem useful vis-à-vis sustainability issues. Additional insights and innovative perspectives on a particular issue can be created, depending

on which user group accesses the data. This, in turn, means that the output of such partnerships is usually not predetermined. For instance, datasets about Nepal, which were stored on the HDX platform, were used by actors as diverse as the Red Cross and MapAction (Wilson, 2015). This is also why systems theory deems open systems to have a higher capacity to address problems than closed systems (Surowiecki, 2004). Although open partnerships may not necessarily provide better outputs than closed initiatives, we can expect that they often have the ability to create more heterogenous (and sometimes even unexpected) outputs that emerge from the diverse interactions of various participants.

By contrast, closed data partnerships are usually designed for creating solutions that fit more narrowly (and usually predefined) purposes, also because the partnering organizations are likely to use the provided data to create knowledge of interest for their own purposes and potentially disfavouring other relevant interests. For instance, the Facebook-CIESIN partnership produces insights that are fit for the declared purpose; in this case to assist disaster response and humanitarian planning to the development of communications infrastructure (Tiecke et al., 2017). However, the commercial purpose of the partnership (e.g. long-term market creation for Facebook) cannot be neglected, and there is a long-term risk that the existence of such commercial interests could undermine the output legitimacy of data partnerships, as well as hinder the evolution of long-term sustainable access models for private data sources for public good (Klein & Verhulst, 2017).

Outcomes. A partnership's output legitimacy also depends on whether the created insights actually create outcomes that make a difference for final beneficiaries (Wolf, 2005). Outcomes usually relate directly to changes in actors' behaviour. For instance, the Flowminder/WorldPop partnership provided unique information from mobile and satellite data analytics in support of the Nepal 2015 earthquake response (Shakya, 2015; Sneed, 2015). This data showed peoples' movement after the disaster and hence enabled relief agencies to provide

food, shelter, and medicine at the right places. Focusing on outcomes implies to link a partnership's legitimacy to the question of whether the created insights really help to address the problem that is supposed to be addressed and whether final beneficiaries are impacted. A partnership may be able to produce insights that allow for creating outcomes, but the outcomes may be less relevant when trying to solve the underlying sustainable development problem. This situation has been described as one of misfit (Galaz et al., 2008). The problem of misfit between a partnership's outputs and its anticipated outcomes is often seen as impeding higher degrees of output legitimacy (e.g. because results are produced with a Western context in mind; Vogel, 2010). However, research has also shown that partnerships' impact on outcomes can vary over time. Some initiatives start out as relatively ineffective tools to address problems but then gain strength through learning processes (Young, 2014).

We can expect that open data partnerships are better equipped to overcome the misfit problem, particularly in those situations where achieving fit requires a testing of different insights. The ability of open partnerships to create non-predetermined insights (see above) enables these initiatives to launch experimentation processes that can align output and outcomes over time. Such experimentation increases the variety of available results and thus the likelihood that at least some of these results make a difference for final beneficiaries (an insight known from the open innovation literature; see West et al., 2014). Also, the speed of the availability of results increases the likelihood that there is a fit between the produced insights and relevant issues. For instance, HDX data was used as a springboard for a collaboration between UN OCHA and the World Food Program to visualize household food consumption in conflict-torn Yemen (United Nations, 2016). The swift availability of insights allowed the partnership to produce outcomes that were matched with the needs of relevant data users. Given that one key problem of current development data is that it is only available with significant

time lags (e.g., when conducted via household surveys; Deaton, 2000), the swift availability of results can enhance outcomes.

While we can expect that closed data partnerships also deliver development data swiftly (because they use similar analytical techniques; Hilbert, 2013), they would overcome the fit problem through careful planning of specific data needs and integration of the competences of participants. For instance, the Facebook-CIESIN partnership addresses a very specific challenge – that of disaggregated population density estimates for developing countries (Tiecke et al., 2017) – an area where CIESIN is a highly regarded pioneering organisation and where Facebook has the relevant access and competences. In fact, Facebook was modelling developing country populations internally for several years as part of their mission to provide internet (and Facebook) access to more people globally (Hempel, 2018).

DISCUSSION AND FUTURE RESEARCH

To discuss the implications of our analysis, we first show in how far our framework differs from legitimacy assessments of partnerships for sustainable development that are *not* tied towards big data analyses (hereafter "traditional partnerships"). This discussion explores the boundary conditions of our framework and thereby also discusses its generalizability. Based on this, we outline a number of future research challenges that arise from our study, both in terms of possible topics and appropriate methods.

The Legitimacy of Traditional Cross-Sector Collaborations and Data Partnerships

What differences and commonalities occur when comparing how to assess the democratic legitimacy of data partnerships and more traditional collaborations? While we believe that some insights of our analysis can also be applied to a discussion of the legitimacy of traditional (i.e. non-data driven) cross-sector partnerships (see e.g., Bäckstrand, 2010; Rueede & Kreutzer,

2015), the legitimacy assessment of data partnerships contains a number of unique characteristics.

One key difference relates to the scale of partnerships. Most traditional partnerships operate as closed collaborative agreements, either on a bilateral basis or between a few selected participants (Selsky & Parker, 2005). The open structure of some data partnerships therefore introduces a new, and so far unexplored, characteristic into the debate around cross-sector collaboration. Our analysis shows that the scale of data partnerships can strengthen their legitimacy (e.g., because it may allow for higher degrees of experimentation and hence the production of more viable outputs), while, at the same time, it can also undercut it (e.g., because it may impede the creation of participatory structures). Given that partnerships with such an open structure start to appear outside of the sustainable development domain – e.g. the idea of open government has witnessed the emergence of a number of collaborations (von Lucke & Große, 2014) – we believe that our insights are important and timely for an analysis of future types of cross-sector collaborations.

The legitimacy assessment of data partnerships also differs from traditional partnerships because we can expect that the level of philanthropic engagement differs. Traditional partnerships are often motivated by concerns for creating a "business case" – that is, the collaborative agreement should yield a financial return to the participating firms (Hartman & Dhanda, 2018). Although it is likely that such motivations also matter with regard to data partnerships, at least indirectly when collaboration is thought to enhance a firm's brand value, the role of philanthropic engagement can expected to be higher. The vast majority of firms donate their anonymized data and thus engage in "data philanthropy" (Kirckpatrick, 2011). Such philanthropic framing can change firms' basis for argumentation in relevant interactions and thus enhance the deliberative quality of data partnerships. Börzel & Risse (2005), for instance, observed that traditional partnerships often need to live with "lowest common

denominator solutions", as interests between business and non-business actors are poorly aligned. While we should not expect an unproblematic overlap of interests in the context of data partnerships, as it would be naïve to assume that large IT firms engage without any strategic motives, we can at least expect that philanthropic considerations help to move deliberations within such partnerships from a bargaining mode (which is focused on self-interest) to a more argumentative mode (which is focused on the common interest) (Risse, 2004).

One area where we see commonalities in the legitimacy assessment of data partnerships and traditional collaborative agreements regards the involvement of international organizations. International organizations like the UN or The World Bank have participated in numerous cross-sector collaborations, both in the context of data partnerships (e.g., the UNICEF Magic Box and Twitter-UN Global Pulse) as well as more traditional partnership arrangements in support of sustainable development (Reed & Reed, 2009). Critics have pointed out that such mingling of the private sector with international organizations can contribute to a creeping commercialization of world politics (Nolan, 2005). Such fears can also be relevant in the context of data partnerships that involve international organizations, and it may threaten or even undercut their legitimacy in the long run, especially as the power of large IT giants like Twitter and Facebook is controversially discussed.

The legitimacy assessments of data partnerships and more traditional collaborations also share another commonality: the difficulty to ensure that multiple actors participate in relevant decisions. One key critique brought up against traditional partnerships is their lack of inclusiveness – that is, a situation in which legitimacy is negatively affected by the inclusion of only selected participants into relevant decisions (Boström, 2006). Our discussion shows that we can expect open and closed data partnerships to face similar struggles; open partnerships because of the complexities involved in setting up participatory structures among a large set of actors and closed partnerships because of their more exclusive nature, lack of formal

governance, and limited consideration of actors outside of the small group of participants. We can therefore assume that data partnerships will find it hard to close the "participatory gap" (Börzel & Risse, 2005, p. 212) that has been observed with regard cross-sector collaborations in support of sustainable development.

Future Research on the Legitimacy of Data Partnerships

We need further insights into the dynamics of this rapidly evolving phenomenon to understand better its potentials and limits. While our analysis has unpacked some of the complexities surrounding the legitimacy of data partnerships, we see the need for future research in at least three key areas.

First, we need to understand better how the legitimacy of individual organizations impacts the entire data partnership. How do legitimacy challenges of individual participants (e.g., firms or NGOs) impact a partnership's legitimacy? Here, the literature on legitimacy spillover effects (Kostova & Zaheer, 1999) can provide important insights and a theoretical yardstick. Will public organizations still continue partnership agreements with companies that see their legitimacy challenged in significant ways (e.g. as Facebook did recently with the Cambridge Analytica scandal)? Would the legitimacy of a partnership as a whole be negatively influenced by such a situation? Such research needs to differentiate between positive and negative spillover effects (Haack et al., 2014), as highly trusted public organizations (e.g., the UN; Barnett & Finnemore, 2008) can also enhance the legitimacy of those organizations that it decides to partner with.

Second, while our theoretical framework made a clear distinction between input and output legitimacy, there are likely to be interaction effects between both. Surprisingly little research has discussed such possible effects (for an exception see Bernauer et al., 2016). For instance, it would be relevant to know whether both sources of legitimacy can substitute each

other (at least in part). Does high input legitimacy make evaluators more supportive of rather poor outcomes? Or, conversely, does knowledge about superior outcomes make evaluators more tolerant towards restrictions with regard to a partnership's input legitimacy? Research can also explore whether trade-offs between both sources exist. For example, "all-inclusive" partnerships may lead to reduced efficiency and may limit the problem-solving capacity of a collaboration (Börzel & Risse, 2005). Although research in this direction can rest on different methods, we believe that ethnographic techniques (e.g., via participant observation) are particularly suited to gain an in-depth understanding of how both sources of legitimacy interact in the context of a specific data partnership. Ethnographic inquires allow for a direct immersion in the lifeworld of those who participate in partnerships (Bourdieu, 1990) and thus emphasize the contextual embeddedness of those criteria that influence input and output legitimacy. For instance, finding out whether and in what ways power was exercised to influence important decisions often requires to study peripheral actors as well as everyday experiences (Rasche & Chia, 2009).

Finally, future research needs to put more emphasis on studying the outcomes of data partnerships. Current research focuses a lot on the outputs of relevant initiatives (see e.g. Ginsberg et al., 2009 and Wesolowski et al., 2015), while it remains unclear what the observed effects of the outputs for final beneficiaries are. This type of research is rare as it faces one significant methodological challenge: it has to isolate those outcomes that can be attributed to the activities of a data partnership from those outcomes that cannot. In other words, there is a risk that the observed outcomes may not actually be due to the insights produced by a data partnership. This problem is further enhanced because many data partnerships also feed outputs into larger multi-stakeholder decision processes that depend on a number of information sources and stakeholders, creating the challenge of disaggregating the outcome attributable to a specific data source. In order to have robust assessments of the democratic legitimacy of data

partnerships, this problem needs to be addressed, as otherwise there is a risk that output legitimacy is under- or overvalued.

We therefore suggest to more strongly consider counterfactual evaluation designs (White, 2010) when studying outcomes. Such an approach would compare a partnership's outcomes in the light of a (often-hypothetical) no-partnership counterfactual (i.e. a situation in which the partnership would have not addressed the issue at all). The main challenge is that counterfactuals can seldom be observed directly and hence need to be approximated (e.g. with reference to another group of beneficiaries). Approximations should be possible in the context of data partnerships, because relevant insights can often be created for particular regions or groups of people due to the granularity of the data (Hilbert, 2013). A counterfactual outcome analysis would then compare such data with groups or regions that did not face the intervention.

CONCLUSION

Data partnerships for sustainable development have created much hope among policy-makers, businesses, civil society organizations, and not least beneficiaries (United Nations, 2011). While many partnerships are still in emergence and we should not rush into conclusions at this stage, our analysis make two key contributions: (a) it introduces and justifies the distinction between open and closed data partnerships for sustainable development and (b) it develops a theoretical framework for analysing their input and output legitimacy. While our analysis makes no empirical claims regarding the legitimacy of relevant initiatives, it emphasizes which particular characteristics of open and closed partnerships can be expected to influence an analysis of their input and output legitimacy. We view this debate as necessary and timely. In 2013, the UN-based High-Level Panel on the SDGs called for the creation of global partnerships on development data and thereby put further focus on these new collaborative agreements. At the same time, the Panel also recognized that without increased legitimacy and

accountability, data partnerships will not be able to change the way international organizations, governments, and NGOs make use of development data (United Nations, 2013).

REFERENCES

Austin, J. E., & Seitanidi, M. M. (2012). Collaborative value creation: A review of partnering between nonprofits and businesses. Part 2: Partnership processes and outcomes. Nonprofit and Voluntary faustSector Quarterly, 41(6), 929-968.

Albrecht, S. (2006). Whose voice is heard in online deliberation?: A study of participation and representation in political debates on the internet. *Information Communication and Society*, *9*(1), 62–82.

Barnett, M., & Finnemore, M. (2008). Political Approaches. In T. G. Weiss & S. Daws (Eds.), *The Oxford Handbook of the United Nations* (pp. 41–57). Oxford, England: Oxford University Press.

Bäckstrand, K. (2010). From Rhetoric to *Practice*: The Legitimacy of Global Public-Private Partnerships for Sustainable Development. In M. Bexell & U. Mörth (Eds.), *Democracy and Public-Private Partnerships in Global Governance* (pp. 145–166). London, England: Palgrave Macmillan UK.

Bekkers, V., & Edwards, A. (2016). Legitimacy and Democracy: A Conceptual Framework for Assessing Governance Practices. In V. Bekkers, G. Dijkstra, A. Edwards, & M. Fenger (Eds.), *Governance and the Democratic Deficit: Assessing the Democratic Legitimacy of Governance Practices* (pp. 35–60). London et al., England: Routledge.

Bernauer, T., Mohrenberg, S., & Koubi, V. (2016). *How Relevant are Input and Output Legitimacy in International Environmental Governance?* National Centre of Competence in Research (NCCR) Challenges to Democracy in the 21st Century, Working Paper No. 94. Zurich.

Bernstein, S., & Cashore, B. (2007). Can Non-State Global Governance Be Legitimate? An Analytical Framework. *Regulation & Governance*, 1(4), 347–371.

Berrone, P., Ricart, J. E., & Carrasco, C. (2016). The open kimono: Toward a general framework for open data initiatives in cities. *California Management Review*, *59*(1), 39–70.

Bitzer, V., & Glasbergen, P. (2015). Business-NGO partnerships in global value chains: Part of the solution or part of the problem of sustainable change? *Current Opinion in Environmental Sustainability*, *12*, 35–40.

Black, J. (2008). Constructing and Contesting Legitimacy and Accountability in Polycentric Regulatory Regimes. *Regulation & Governance*, *2*(2), 137–164.

Börzel, T. A., & Risse, T. (2005). Public-Private Partnerships: Effective and Legitimate Tools of Transnational Governance? In E. Grande & L. W. Pauly (Eds.), *Complex Sovereignty: Reconstituting Political Authority in the Twenty-First Century* (pp. 195–216). Toronto et al., Cananda: University of Toronto Press.

Boström, M. (2006). Regulatory Credibility and Authority through Inclusiveness: Standardization Organizations in Cases of Eco-Labelling. *Organization*, *13*(3), 345–367.

Bourdieu, P. (1990). The Logic of Practice. Cambridge, England: Polity Press.

Bull, B., & McNeill, D. (2007). Development issues in global governance: Public–private partnerships and market multilateralism. New York: Routledge.

Buse, K., & Harmer, A. M. (2007). Seven habits of highly effective global public–private health partnerships: practice and potential. *Social science & medicine*, 64(2), 259-271.

Center for International Earth Science Information Network. (CIESIN) (2016). CIESIN Teams with Facebook to Develop Open, Improved Settlement Data. Accessed (August 27, 2018) at: http://blogs.ei.columbia.edu/2016/11/29/ciesin-teams-with-facebook-to-develop-open-improved-settlement-data/

Center for International Earth Science Information Network. (CIESIN) (2018). High Resolution Settlement Layer, Accessed (August 27, 2018) at: https://www.ciesin.columbia.edu/data/hrsl/

Christensen, L. T., & Schoeneborn, D. (2017). The Corporate Construction of Transparency and (In)transparency. In A. Rasche, M. Morsing, and J. Moon (Eds.), *Corporate Social Responsibility: Strategy, Communication, Governance* (pp. 350–370). Cambridge et al., England: Cambridge University Press.

Crowell, C. (2016). Twitter and UN Global Pulse announce data partnership. Accessed (August 27, 2018) at: https://blog.twitter.com/official/en_us/a/2016/twitter-and-un-global-pulse-announce-data-partnership.html

Deaton, A. (2000). *The analysis of household surveys*. Baltimore, Maryland: Johns Hopkins University Press.

Dietrich, D. et al. (2018). *The Open Data Handbook*. Cambridge, England: Open Knowledge Foundation.

Dingwerth, K. (2007). *The New Transnationalism: Transnational Governance and Democratic Legitimacy*. Basingstoke, England: Palgrave Macmillan.

Doty, D. H., & Glick, W. H. (1994). Typologies as a Unique Form of Theory Building: Toward Improved Understanding and Modeling. *Academy of Management Review*, *19*(2), 230–251.

Dryzek, J. S. (2010). *Foundations and Frontiers of Deliberative Governance*. Oxford, England: Oxford University Press.

Einav, L. & Levin, J. (2014). Economics in the age of big data, *Science*, Vol. 346 (6210). DOI: 10.1126/science.1243089

Emerson, R. M. (1962). Power-Dependence Relations. *American Sociological Review*, *27*(1), 31–41.

Etter, M., Ravasi, D., & Colleoni, E. (2017). Social media and the formation of organizational reputation. *Academy of Management Journal*, forthcoming (online November 2017).

Fleming, P., & Spicer, A. (2014). Power in Management and Organization Science. *Academy of Management Annals*, *8*(1), 237–298.

Flyverbom, M., Madsen, A. K., & Rasche, A. (2017). Big data as governmentality in international development: Digital traces, algorithms, and altered visibilities. *The Information Society*, 33(1), 35–42.

Friberg-Fernros, H., & Schaffer, J. K. (2014). The consensus paradox: Does deliberative agreement impede rational discourse? *Political Studies*, *62*(S1), 99–116

Galaz, V., Hahn, T., Folke, C., & Svedin, U. (2008). The Problem of Fit among Biophysical Systems, Environmental and Resource Regimes, and Broader Governance Systems: Insights and Emerging Challenges. In O. R. Young, L. A. King, and H. Schroeder (Eds.).

Institutions and Environmental Change: Principal Findings, Applications, and Research Frontiers (pp. 147–186). Cambridge, MA: MIT Press.

Ginsberg, J., Mohebbi, M. H., Patel, R. S., Brammer, L., Smolinski, M. S., & Brilliant, L. (2009). Detecting influenza epidemics using search engine query data. *Nature,* 457, 1012–1014.

Global Open Data on Agriculture and Nutrition (GODAN) (2018). Statement of Purpose. Accessed (August 27, 2018) at: https://www.godan.info/pages/statement-purpose

Goodin, R. E. (2000). Democratic deliberation within. *Philosophy and Public Affairs*, 29, 81–109.

Global Mobile Industry Association (GSMA) (2016). *The Mobile Economy*. London, England: GSMA.

Gulbrandsen, L. H. (2005). The effectiveness of non-state governance schemes: A comparative study of forest certification in Norway and Sweden. *International Environmental Agreements: Politics, Law and Economics*, 5(2), 125–149.

Haack, P., Pfarrer, M. D., & Scherer, A. G. (2014). Legitimacy-as-Feeling: How Affect Leads to Vertical Legitimacy Spillovers in Transnational Governance. *Journal of Management Studies*, 51(4), 634–666.

Habermas, J. (1998). *The Inclusion of the Other: Studies in Political Theory*. Cambridge, MA.: MIT Press.

Habermas, J. (1990). The New Conservatism. Cambridge, MA: MIT Press

Hale, T. N. (2008). Transparency, Accountability, and Global Governance. *Global Governance*, 14(1), 73–94.

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

HDX & UN OCHA (2017). About the Humanitarian Data Exchange. Accessed (August 27, 2018) at: https://data.humdata.org/faq.

Hempel, Jessie (2018) What happened to Facebook´s grand plan to wire the world? Wired, May 17. Accessed 27 August, 2018) at: https://www.wired.com/story/what-happened-to-facebooks-grand-plan-to-wire-the-world/

Hilbert, M. (2013). Big Data for Development: From Information- to Knowledge Societies (Social Science Research Network (SSRN) Working Paper). Retrieved from http://ssrn.com/abstract=2205145.

Humanitarian OpenStreetMaps HOTOSM (2018), Voting Members, Accessed (August 27, 108) at: https://www.hotosm.org/voting-members

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

Karpowitz, C. F., & Mendelberg, T. (2007). Groups and Deliberation. *Swiss Political Science Review*, *13*(4), 645–662.

Kessler, C., & Hendrix, C. (2015). The Humanitarian eXchange Language: Coordinating disaster response with semantic web technologies. *Semantic Web*, *6*(1), 5–21.

Kirckpatrick, R. (2011, September 16). "Data Philanthropy: Public and Private Sector Data Sharing for Global Resilience, Blogpost." Retrived August, 8, 2018 from https://www.unglobalpulse.org/blog/data-philanthropy-public-private-sector-data-sharing-global-resilience

Klein, T. & Verhulst, S. (2017). Access to New Data Sources, Business Models and Incentives for the Corporate Sector, (Discussino Paper no. 10). Paris: Partnership in Statistics for Development in the 21st Century (PARIS21).

Kostova, T., & Zaheer, S. (1999). Organizational Legitimacy Under Conditions Of Complexity: The Case Of The Multinational Enterprise. *Academy of Management Review*, *24*(1), 64–81.

Le Ber, M. J., & Branzei, O. (2010). Towards a critical theory of value creation in cross-sector partnerships. *Organization*, 17(5), 599-629.

Levine, P., Fung, A., & Gastil, J. (2005). Future Directions for Public Deliberation. *Journal of Public Deliberation*, *1*(1), 271–288.

Lohr, S. (2013, August 7). Searching Big Data for "Digital Smoke Signals." Accessed (August 27, 2018) at: http://www.nytimes.com/2013/08/08/technology/development-groups-tap-big-data-to-direct-humanitarian-aid.html?_r=0

Lourenço, R. P. (2015). An analysis of open government portals: A perspective of transparency for accountability. *Government Information Quarterly*, 32(3), 323–332.

Madsen, A. K., Flyverbom, M., Hilbert, M., & Ruppert, E. (2016). Big Data: Issues for an International Political Sociology of Data Practices. *International Political Sociology*, *10*(3), 275–296.

Mansbridge, J., Bohman, J., Chambers, S., Christiano, T., Fung, A., Parkinson, J., Thompson, D. F., & Warren, M. E. (2012). A systemic approach to deliberative democracy. In J. Par- kinson & J. Mansbridge (Eds.), *Deliberative Systems* (pp. 1-26). Cambridge, England: Cambridge University Press.

Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C. and Byers, A.H. 2011). "Big Data: The next frontier for innovation, competition, and productivity". Report McKinsey Global Institute, Accessed (August 27, 2018) at: https://www.mckinsey.com/business-functions/digital-mckinsey/our-insights/big-data-the-next-frontier-for-innovation

Mayer-Schönberger, V., & Cukier, K. (2013). *Big Data: A Revolution that will Transform How We Live, Work and Think.* London, England: John Murray.

Mayntz, R. (2010). *Legitimacy and Compliance in Transnational Governance.* Working Paper 10/5. Cologne, Germany: Max Planck Institute for the Study of Societies.

McAfee, A., and Brynjolfsson, E. (2012). Big Data: The management revolution. *Harvard Business Review,* 90, 60–68.

Melamed, C. (2014) *A World That Counts: Mobilizing the Data Revolution for Sustainable Development*. Report prepared at the request of the United Nations Secretary General, by the Independent Expert Advisory Group on a Data Revolution for Sustainable Development, November.

Mena, S., & Palazzo, G. (2012). Input and Output Legitimacy of Multi-Stakeholder Initiatives. *Business Ethics Quarterly*, *22*(3), 527–556.

Nolan, J. (2005). The United Nations' Global Compact with business: Hindering or helping the protection of human rights? *University of Queensland Law Journal*, *24*, 445–466.

Open Data Institute (2013). What makes data open. Accessed (August 27, 2018) at: https://theodi.org/article/what-makes-data-open/

Palazzo, G., & Scherer, A. G. (2006). Corporate Legitimacy as Deliberation: A Communicative Framework. *Journal of Business Ethics*, *66*(1), 71–88.

Palen, L., Soden, R., Anderson, T. J., & Barrenechea, M. (2015). Success & Scale in a Data-Producing Organization: The Socio-Technical Evolution of OpenStreetMap in Response to Humanitarian Events. *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems - CHI '15*, 4113–4122.

Parkinson, J. (2003). Legitimacy problems in deliberative democracy. *Political Studies*, *51*(1), 180–196.

Pingree, R. J. (2006). Decision structure and the problem of scale in deliberation. *Communication Theory*, *16*(2), 198–222.

Poel, M., Schroeder, R., Treperman, J., Rubinstein, M., Meyer, E., Mahieu, B., … Svetachova, M. (2015). *Data for Policy: A study of big data and other innovative data-driven approaches for evidence-informed policymaking*. Oxford, England: Oxford Internet Institute & Centre for European Policy Studies.

Rasche, A., & Chia, R. (2009). Researching strategy practices: a genealogical social theory perspective. *Organization Studies*, *30*(7), 713–734.

Reed, A., & Reed, D. (2009). Partnerships for Development: Four Models of Business Involvement. *Journal of Business Ethics*, *90*, 3–37.

Risse, T. (2004). Global Governance and Communicative Action. *Government and Opposition*, *39*(2), 288-313

Rueede, D., & Kreutzer, K. (2015). Legitimation Work Within a Cross-Sector Social Partnership. *Journal of Business Ethics*, 128(1), 39–58.

Ruef, M., & Scott, W. R. (1998). A Multidimensional Model of Organizational Legitimacy: Hospital Survival in Changing Institutional Environments. *Administrative Science Quarterly.*, *43*(4), 877–904.

Ryfe, D. M. (2005). Does Deliberative Democracy Work? *Annual Review of Political Science*, *8*(1), 49–71.

Sa, C., & Grieco, J. (2016). Open Data for Science, Policy, and the Public Good. *Review of Policy Research*, *33*(5), 526–543.

Samman, E. (2013, June 10). *Using household surveys to start a data revolution and tackle social inequality*. Retrieved August, 28, 2013 from http://www.theguardian.com/global-development-professionals-network/2013/jun/10/mdgs-household-surveys-data-revolution

Scharpf, F. W. (1997). Economic Integration and the Welfare State. *Journal of European Public Policy*, 4(1), 18–36.

Schouten, G., Leroy, P., & Glasbergen, P. (2012). On the deliberative capacity of private multi-stakeholder governance: The Roundtables on Responsible Soy and Sustainable Palm Oil. *Ecological Economics*, *83*, 42–50.

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.

Shakya, A. (2015) Where are we? Ncell partners with Flowminde to track movements of Nepalis post-earthquake. Nepali Times, 29 May-05 June, #760. Accessed (August 27, 2018) at: http://archive.nepalitimes.com/article/nation/Ncell-Flowminder-track-movement-of-nepalis-post-earthquake,2278

Sneed, A. (2015, August 05) The Open Source Maps that Make Rescues in Nepal Possible. Wired. Accessed (August 27, 2018) at https://www.wired.com/2015/05/the-open-source-maps-that-made-rescues-in-nepal-possible/

Steele, J., Sundsøy, P., Pezzulo, C., Alegana, V., Bird, T., Blumenstock, J., Bjelland, J., Engø-Monsen, K., de Montjoye, Y., Iqbal, A., Hadiuzzaman, K., Lu, X., Wetter, E., Tatem, A. & Bengtsson, L. (2017) *Mapping poverty using mobile phone and satellite data*, *Journal of Royal Society Interface*, Vol. 14(127), 1-10. DOI: 10.1098/rsif.2016.0690

Suchman, M. C. (1995). Managing Legitimacy: Strategic and Institutional Approaches. *Academy of Management Review*, *20*(3), 571–610.

Suddaby, R. O. Y., Bitektine, A., & Haack, P. (2017). Legitimacy. *Academy of Management Annals*, *11*(1), 451–478.

Surowiecki, J. (2004). *The Wisdom of Crowds: Why the Many Are Smarter Than the Few and How Collective Wisdom Shapes Business Economies, Societies and Nations*. London, England: Doubleday

Susha, I., Janssen, M., & Verhulst, S. (2017). Data Collaboratives as a New Frontier of Cross-Sector Partnerships in the Age of Open Data: Taxonomy Development. *Proceedings of the 50th Hawaii International Conference on System Sciences*, 2691–2700.

The World Bank (2004). *Monitoring and Evaluation: Some Tools, Methods and Approaches.* Washington D.C.: The World Bank.

Tiecke, T. G., Liu, X., Zhang, A., Gros, A., Li, N., Yetman, G., … Dang, H.-A. H. (2017). Mapping the world population one building at a time. Accessed (August 27, 2018) at: https://arxiv.org/abs/1712.05839

Tiecke, T. G. (2016) *Open population datasets and open challenges* Accessed (August 27, 2018) at: https://code.facebook.com/posts/596471193873876/open-population-datasets-and-open-challenges/

United Nations (2005). *Household Sample Surveys in Developing and Transition Countries*. Department of Economic and Social Affairs (ST/ESA/STAT/SER.F/96). New York: United Nations.

United Nations. (2011). *Secretary-General's remarks at General Assembly briefing on the Global Pulse initiative.* Retrieved October 15, 2017, from: http://www.un.org/sg/statements/?nid=5668).

United Nations (2013). *A New Global Partnership: Eradicate Poverty and Transform Economies through Sustainable Development,* (The report of the high-level panel of eminent persons on the post-2015 Development agenda). New York: United Nations.

United Nations (2016). UN Agencies Boost Partnership on Visualization of Food Security Data for Yemen. Accessed (August 27, 2018) at:

https://news.un.org/en/story/2016/05/528392-un-agencies-boost-partnership-visualization-food-security-data-yemen.

UN OCHA, United Nations Office for the Coordination of Humanitarian Affairs (2016). *Connecting People and Data to Improve Lives*, New York.

UN Global Pulse (2014) *Mining Indonesian Tweets to Understand Food Price Crises*, Methods Paper, Accessed (August 28, 2018) at: http://www.unglobalpulse.org/sites/default/files/Global-Pulse-Mining-Indonesian-Tweets-Food-Price-Crises%20copy.pdf.

UN Global Pulse and Crimson Hexagon (2011) *Twitter and perceptions of crisis-related stress: Methodological white paper*. New York: United Nations.

Utting, P., & Zammit, A. (2009). United Nations-business partnerships: Good intentions and contradictory agendas. Journal of Business Ethics, 90(1), 39-56.

von Lucke, J., & Große, K. (2014). Open Government: Opportunities and Challenges for Public Governance. In M. Gascó-Hernández (Ed.), *Open Government Collaboration* (pp. 189–204). New York: Springer New York.

van Tulder, R., Seitanidi, M. M., Crane, A., & Brammer, S. (2016). Enhancing the impact of cross-sector partnerships. Journal of Business Ethics, 135(1), 1-17.

Verhulst, S., & Sangokoya, D. (2015, April 22). Data Collaboratives: Exchanging Data to Improve Peoples' Lives. Accessed (August 27, 2018) at: https://medium.com/@sverhulst/data-collaboratives-exchanging-data-to-improve-people-s-lives-d0fcfc1bdd9a

Vogel, D. (2010). The Private Regulation of Global Corporate Conduct: Achievements and Limitations. *Business & Society*, *49*(1), 68–87.

Waddock, S. A. (1991). A typology of social partnership organizations. *Administration & Society*, *22*(4), 480–515.

Walzer, M. (1999). Deliberation and what else? In S. Macedo (Ed.), Deliberative politics: Essays on democracy and disagreement (pp. 58–69). New York: Oxford University Press.

Wesolowski, A., Buckee, C., Bengtsson, L., Wetter, E., Lu, X. & Tatem, A. (2015). *Commentary: Containing the Ebola outbreak – the potential and challenge of mobile network data*, PLoS Current Outbreaks. Edition 1 accessed (August 27, 2018) at: doi: 10.1371/currents.outbreaks.0177e7fcf52217b8b634376e2f3efc5e.

West, J., Salter, A., Vanhaverbeke, W., & Chesbrough, H. (2014). Open innovation: The next decade. *Research Policy*, *43*(5), 805–811.

White, H. (2010). A contribution to current debates in impact evaluation. *Evaluation*, *16*(2), 153–164.

Wilson, M. (2015, April 30). The Candy Crush of Data is Saving Lives in Nepal. Retrieved December 18, 2017, from: https://www.fastcodesign.com/3045699/how-the-candy-crush-of-data-is-saving-lives-in-nepal.

Wolf, K. D. (2005). Private actors and the legitimacy of governance beyond the state. In A. Benz & I. Papadopoulos (Eds.), *Governance and Democratic Legitimacy: Conceptual Outlines and Empirical Explorations* (pp. 200–227). London, England: Routledge.

Young, O. R. (2014). The effectiveness of international environmental regimes: Existing knowledge, cutting-edge themes, and research strategies. *Advances in International Environmental Politics*, 108(50), 273–299.

Young, I. M. 2000. *Inclusion and democracy*. Oxford, England: Oxford University Press.

Table 1: Open and Closed Data Partnerships for Sustainable Development

		Description	Membership *(Who participates in the data partnership?)*	Sourcing of Data *(Who can access data that the partnership rests on?)*	Sharing of Insights *(Who gets access to the produced insights?)*
Open Data Partnerships for Sustainable Development	**UN OCHA HDX**	Foundation: 2014 Scope: Global Focus area: diverse humanitarian data	Open application, liberal signup criteria (open)	All members have unrestricted access to data sources (open)	Unrestricted sharing of insights (open)
	GODAN (Global Open Data on Agriculture and Nutrition)	Foundation: 2013. Scope: Global. Focus area: agriculture and nutritional data	All interested organisations can join (open)	All members have unrestricted access to data sources (open)	Unrestricted sharing of insights (open)
	Humanitarian OpenStreetMaps (HOTOSM)	Foundation: 2010 Scope: Global Focus area: map data in the context of crises responses	Open for individuals and partner organisations (open)	All members have unrestricted access to data sources (open)	Unrestricted sharing of insights (open)
	Flowminder/WorldPop	Foundation: 2010 Scope: Global Focus area: diverse humanitarian data	Uni. Southampton and Flowminder Foundation, selected other partners (closed)	Open data (WorldPop), proprietary telco data, proprietary UN agency data (open)	Unrestricted sharing of insights through WorldPop (open)
Closed Data Partnerships for Sustainable Development	**Twitter-UN Global Pulse**	Foundation: 2016 Scope: Global Focus area: diverse humanitarian data	Bilateral agreement between UN and Twitter (closed)	Sourced from Twitter only (closed)	To UN agencies and other humanitarian agencies (closed)
	GSMA Big Data for Social Good	Foundation: Launched in 2017. Scope: Global Focus area: mobile data for disaster responses	Telecommunication firms and selected UN agencies only (closed)	Telecommunication data from participants only (closed)	To selected UN agencies and other humanitarian agencies (closed)
	UNICEF Magic Box	Foundation: 2016 Scope: Global Focus area: diverse humanitarian data	UNICEF & selected data partners; Facebook, Telefonica, IATA (closed)	Data partners provide data to be analysed in Magic Box (closed); some public sources are used	To community of UN agencies (incl. their stakeholders) through UNICEF (closed)
	Facebook-CIESIN High Resolution Settlement Layer (HRSL)	Foundation: 2016. Scope: Global Focus area: high resolution population estimates	Facebook, selected UN agencies, ICRC & Columbia University (closed)	Sourced from Facebook only (closed)	To selected UN agencies (incl. their stakeholders), ICRC & Columbia University (closed)

Table 2 Input and Output Legitimacy of Data Partnerships

Dimension	Criterion	Definition	Key Questions	Expected Influence of Open/Closed Nature of Data Partnership on Criterion
Input Legitimacy	Participation	Data partnerships ability to include stakeholders affected by the outcomes of the partnership into relevant structures and decision-making processes	What are the patterns of exclusion and inclusion when partnerships make decisions? Are there actors that dominate the partnership? Are there barriers to participation?	• Open data partnerships should find it more difficult to couple participants and decision-makers because of their underlying complexity • Representatives can ensure certain level of inclusiveness in open data partnerships, but suffer from heterogenous set of actors that change frequently • Closed data partnerships can be expected to face limits in terms of their ability to involve a diverse set of representative actors
	Transparency	Data partnerships ability to make their behaviour knowable to interested parties, both in the sense of answering inquires and also with regard to the general disclosure of relevant information	What are the levels of transparency that the partnership agrees to? What accountability structures are created? What information is accessible and by whom is it accessible?	• Open data partnerships do not necessarily need to offer high levels of transparency with regard to data sourcing and data analysis • Transparency may be blurred by unclarity about data sources (e.g., is the data reposted) and the combination of different datasets within analyses • Both, open and closed data partnerships can suffer from non-transparent decision-making processes (regardless of whether the underlying data sourcing and analysis is transparent)
	Deliberative Quality	Data partnerships ability to facilitate a fair process of argumentation in which different stakeholders have a valid voice in relevant decision-making processes	Do participating stakeholders have a valid voice in decision-making processes? To what extend does a partnership live up to the ideal of an unconstrained dialogue? Are participating actors able to achieve a consensual solution to the problem they are addressing?	• Power differences between actors can be expected to be most visible in closed data partnerships where a strong dependence on single actors exists • Closed data partnerships may therefore replace "argumentative power" with "bargaining power" • Small-scale deliberations in closed data partnerships can be expected to find it easier to create shared knowledge structures (because information cues can be exchanged easier)
Output Legitimacy	Outputs	Data partnerships ability to produce outputs that are perceived as relevant by policy makers	Does the partnership have the required resources in place to create relevant insights to policy makers? Does the partnership have feedback mechanisms in place to adjust its outputs? Are the created insights really used by others (e.g., to change policies)?	• The output of open data partnerships is less predetermined and hence can be more heterogeneous and even unexpected. This can potentially improve the a partnership's problem-solving capacity (although outputs themselves may not necessarily be better). • Closed data partnerships usually provide insights on pre-defined and more narrow purposes. This can potentially limit the problem-solving capacity of a partnership (although outputs may be of high quality).
	Outcomes	Data partnerships ability to create outcomes that help to address the issue that the partnership was set up to alleviate	Does the partnership manage to produce outcomes that address the issue which the partnership aims to address?	• Problem of misfit – i.e. a partnership's outputs do not produce the desired outcomes – is overcome by open data partnerships through experimentation that can align outputs and outcomes • Closed data partnerships are likely to address misfit through more careful planning of data needs and competences