C4DC
CONTRACTS FOR DATA COLLABORATION

DEMYSTIFYING DATA SHARING: INSIGHTS FROM KEY STAKEHOLDERS

GOVLAB

UNIVERSITY *of* WASHINGTON

TR=NDS
Thematic Research Network
on Data and Statistics

WORLD ECONOMIC FORUM

TABLE OF CONTENTS

BACKGROUND INFORMATION AND ACKNOWLEDGEMENTS

Contracts for Data Collaboration (C4DC) (UNSDSN 2019) aims to shed light on the opportunities and challenges inherent to data sharing practices. The project will work to provide policy makers, development and humanitarian organisations, and private companies with a range of tools for them to better understand formalized data sharing practices that are underpinned by written agreements. Project outputs will include an online repository of data sharing agreements (DSAs) by the end of 2019, together with an analytical framework explaining the relevance of each part of a DSA. In the medium-term, the project hopes to also produce and publish a series of case studies documenting when and how DSAs are, and can be, most effectively applied for maximum impact and with minimum risk within development and humanitarian contexts. This insight report draws extensively on interviews with partners and key stakeholders. It aims to break down the complexities and opacity of DSAs as much as possible and to raise awareness of their value and potential, using terms from project partners and development and humanitarian professionals.

Demystifying data sharing: Insights from key stakeholders

1. THE RISKS AND OPPORTUNITIES OF DATA SHARING

The transfer of data from one set of hands to another raises a whole host of questions around data rights, ownership, use, control and risk that can seem overly complicated or off-putting when first encountered. This is for good reason as the risks involved in the sharing of data, especially personal or sensitive data, are very real. In 2013 - 2014 for instance, "errors in data sharing and disclosure accounted for 17% of complaints to the UK Information Commissioner's Office (ICO) and approximately 45% of ICO enforcement action [i.e. sanctions and fines]" (Lloyd-Jones 2015).

Notwithstanding the risks, the benefits of data collaboration are immense. Whole industries now rely on the sharing, mashing-up, (re)processing and analysis of data to generate value and wealth, with benefits that span both the private and public sectors. In the private sector, many of the world's most valuable companies now rely on data sharing in one form or another as key to their success (Statista 2018). Within the public sector, the use of data to drive evidence-informed policy making is now well established too (Jones 2012).

This balancing act between risk and opportunity informs the data activities of organizations and individuals acting within the sustainable development and humanitarian sectors as well. The need for focused research and work on data sharing in these spaces is widely recognized, and TReNDS' flagship report, Counting on the World (UNSDSN 2017) identifies a need for inclusive partnerships, principles and standards to drive better data sharing practices in support of the Sustainable Development Goals (SDGs). In relation to the SDGs especially, data sharing has the potential to unlock information that can strengthen humanitarian responses and inform their attainment (United Nations 2015). In pursuit of these opportunities, governments, private companies, researchers and development practitioners all have to navigate similar processes as they strive to reach agreements on how to share data between and among themselves. Navigating contractual negotiations and all the considerations that have to be kept in mind during that process – the who, what, where, when, why and how of data sharing – is intricately complex and is informed by a combination of technological, institutional, and ultimately, human factors.

EXAMPLES OF DATA SHARING AGREEMENTS IN ACTION

In Colombia, the Centro de Pensamiento Estratégico Internacional (CEPEI) has piloted an innovative project with the Bogota Chamber of Commerce to reconcile local data sources. The result has been a supply of relevant data on economic growth, infrastructure and industrialization now available to the National Administrative Department of Statistics (DANE). Although the collaboration has been a success, securing the necessary arrangements proved more difficult than initially anticipated. CEPEI was able to analyse the Chamber of Commerce data in less than two months, but the process of negotiating a one-and-a-half-page agreement to enable them to do so took over six months. Similarly, Development Gateway, which promotes data-driven development solutions, regularly finds that it takes three to four months to negotiate data sharing agreements with its partners: "developing the document, getting buy-in from the different stakeholders, the different country officials, that all takes time." (Beverley

Hatcher-Mbu, Development Gateway). Most recently, Flowminder took a year to negotiate a three-way data sharing agreement in Ghana between themselves, Vodafone and Ghana Statistical Services. What the three examples share is an understanding that any data sharing, or partnership agreement that involves data being shared, must be founded in trust which takes time to develop through discussions and negotiations, and is supported and underpinned by broader regulatory and legal frameworks.

Ultimately, as the sustainable development and humanitarian communities promote developmental efforts informed by data, data partnerships must be supported by an increased understanding of the sorts of legal arrangements that can maximize impact while also ensuring the responsible and safe use of data.

This report aims to shed light on the potential of DSAs in this space, drawing on insights from the C4DC project's partners and the perspectives gleaned from interviews with staff at Development Gateway, Flowminder and CEPEI; three organisations with extensive data sharing experience across the development and humanitarian sectors.

2. WHAT ARE DATA SHARING AGREEMENTS AND HOW ARE THEY USED?

The term 'data sharing agreement' has broad applications and can mean different things depending on who you are talking to and what part of the world you are in. It is important to distinguish legal and everyday definitions of DSAs. This is because in certain jurisdictions, DSAs can have very specific legal meanings that vary greatly. For the purposes of the C4DC project, a broad-stroke and inclusive approach has been taken. DSAs are deemed to include any form of written agreement, be it a legal contract, memorandum of understanding, non-disclosure agreement, or other mechanism that sets out the terms for how data is shared between parties.

In the sustainable development sector, many organisations find DSAs useful for project planning, implementation and keeping track of who is accountable for what within a project.

"For me, data sharing means that we need to have rules and regulations to ensure that different stakeholders that are sharing and using data are cognizant of the risks that exist when data is shared. I refer back to our data sharing agreements all the time. The MOU [memoranda of understanding] we use to set the terms for sharing help both partners in the project keep each other accountable"

– Charlene Migwe (Development Gateway, Kenya)

In other contexts, DSAs form part of larger, more comprehensive agreements and can relate to very specific issues – such as confidentiality.

> "A data sharing agreement relates to how two or more entities agree to share data within a legal framework, and the conditions that they attach to that sharing. The most important clauses in my view are around the use of data, and how it is connected with confidentiality and privacy issues. We call them 'confidentiality agreements' and they are typically agreed after the general MOU that contains the terms of the overall partnership"
>
> – *Fredy Rodriguez (CEPEI)*

And in some jurisdictions, they are required by law and need to be treated as a compliance issue for the protection of individuals' rights in relation to their data.

> "I first heard the term 'data sharing agreement' properly in the context of the General Data Protection Regulation [GDPR] sometime last year. Due to my focus being on working with personal data, I initially think of the data processing agreements that need to be in place for the sharing of personal data but I'm also aware that you may have data sharing agreements for non-personal data"
>
> – *Tracey Li (Flowminder)*

Across these areas flows a common thread that is an essential ingredient of any good agreement – the need for trust. "The concept of 'trust' sits right at the heart of a data sharing agreement" as Bill Hoffman from WEF puts it (Hoffman 2019). More broadly, trust is "fundamental […] to the necessity to encourage greater cooperation between public and private entities and confidence in each other's respective roles and responsibilities" as Jessica Espey of TReNDS explained in a recent blog post. (Espey 2019). Within the project,

> "We have deliberately avoided having a definition of a 'good' data sharing agreement because the concept varies according to interest and situation. I think perhaps the better route is to describe the issues that the parties had to negotiate around and help contextualise their decisions."
>
> – *Hayden Dahmm (TReNDS)*

A variety of use scenarios for DSAs in the development and humanitarian sectors exist:

Use Category	Explanation	Example
Filling gaps in national data	National Statistics Offices (NSOs) or other governmental bodies can use DSAs to negotiate access to privately held data that can enhance official statistics.	In Colombia, DANE has gained access to Sustainable Development Goal (SDG)-relevant data, enabled by an agreement between CEPEI and the Bogota Chamber of Commerce (Fredy Rodríguez, 2019).
Accessing private data during emergencies	In the event of an emergency, having an existing agreement can help private data holders provide information that leads to better emergency responses.	Flowminder has worked in multiple countries across the world to access telecom data for humanitarian and development purposes (e.g. for disaster response, and studying disease transmission). Agreements may take months or years to realize, but lead to valuable insights that can inform decision-making once in place. (Tracey Li, 2019)
Sharing data between nonprofit and public sectors	Non-profit organizations that gain special access to public sector data can improve the quality of their community services. (https://www.utah.gov/pmn/files/334857.pdf)	The Utah State Board of Education and the group Utahans Against Hunger developed a DSA to share school-level data and improve student breakfast services.
Private-public data sharing	Private companies can share data with the public sector to improve planning and policy.	The private companies Esri and Waze have shared transportation data with city governments around the world. (The GovLab 2019ii).
Public-to-public data sharing	Without a legal framework, even sharing data within government can be challenging. Agreements or MOUs can provide the necessary structure.	Development Gateway has done extensive work on improving data sharing within governments and between ministries in developing countries (Beverley Hacher, 2019).
Trans-national data sharing	Sharing data across borders can help improve international cooperation but may require special considerations in an agreement.	Jersey and Germany signed an agreement to share tax data to improve service delivery. (https://www.gov.je/Government/PlanningPerformance/Pages/MinisterialDecisions.aspx?showreport=yes&docid=1165D270-71FF-4A86-81DB-E0C04CE71FA6)

3. THE CONTRACTS FOR DATA COLLABORATION (C4DC) PARTNERSHIP

While each partner to the C4DC partnership has come to the project with different objectives – ranging from how barriers to public data sharing can be lowered, through to how private sector data sharing for public purposes can be incentivized, there is a common understanding between all partners that the risks involved in data sharing cannot be overcome in isolation. A DSA is an artifact of a given data sharing de-risking exercise among its parties. Reviewing a body of agreements will reveal commonalities that may facilitate the exercise of solving multiple interdependent data risks, and doing so in ways that individual parties cannot achieve alone.

The priorities for UNSDSN-TReNDS include supporting statisticians, government policy makers and knowledge managers to better understand the issues involved in data sharing, helping these actors to have

> "As a project, we quickly identified that the legal and institutional arrangements hindering public-private data partnerships were immense. Many of the collaborations that we heard of were being seriously held-up by the fact that the partners couldn't reach a consensus on things like who was the data owner, who had the right to access certain information under what kinds of conditions, and so on."
>
> – Jessica Espey (TReNDS)

THE C4DC DATA SHARING ANALYTICAL FRAMEWORK

One of the main products from the C4DC partnership to date is the analytical framework (see Annex A) that is being used to analyse DSAs and break them down into building blocks in preparation for their addition to the planned online repository. The analytical framework has been collectively built by all partners. At the broadest frame, the analytical framework has been informed by the GovLab's work on Data Collaboratives, first defined in 2015 (The GovLab 2015) as: "a new form of collaboration, beyond the public-private partnership model, in which participants from different sectors - including private companies, research institutions, and government agencies - can exchange data to help solve public problems." The intention has been to build a tool that not only aids analysis within the project but can also be used as a checklist once complete. This has been further honed by partners at University of Washington (UW).

> "In many contexts the most useful product from this project might not be the contracts themselves – for instance in places where there is no enforcement – but rather the talking points that flow from the framework. Using the framework as a checklist I think is going to be one of the most valuable dimensions of this project and the value of the checklist is independent of whether you eventually enter into a contract or not."
>
> – Scott David (UW)

At the highest level of analysis, the framework draws on the GovLab's Contractual Wheel of Data Collaboration developed by Stefaan G. Verhulst and Andrew Young (Govlab 2019iii), as part of their data collaboratives program and splits out the components of a DSA into six parts: Where, Why, What, Who, How and When. Depending on the use case, context and jurisdiction, a typical DSA should contain provisions that cover most, if not all these dimensions. The complete framework can be found on the Contracts for Data Collaboration website.

- **Why data is being shared:** what is the context and purpose for which data is being shared? DSAs should provide details of the purpose(s) for which data are being shared.

- **What kinds of data are being shared:** DSAs should aim to be as specific as possible about what data are being shared (and what are not). DSAs should ideally specify the standards, formats and rules that apply to the data in question, including the source of the data, the metadata, and other technical requirements.

- **When data should be shared:** DSAs should be clear on timelines, from the dates that the agreement starts and ends through to any important milestones throughout the agreement period; as well as details of how long data should be stored and used following the end of the agreement, right up to the point of safe disposal or archiving.

- **Who is involved in the sharing:** fundamental to a DSA is an explanation of who the parties to the agreement are. Describing parties' legal status (i.e. natural persons, corporate persons, government bodies, charities, etc.) helps to establish that they have the authority to enter into the agreement in the first place, the authority or right to share data, and conversely the authority to receive it. Are there any other parties other than those making the agreement who have an interest in the agreement (so-called 'Third Parties')? The agreement should set out roles and responsibilities.

- **How data is being shared:** In addition to specifying the standards, formats and rules that should be used to determine what data should be shared, DSAs should also specify how they should be shared. The DSA should specify the data security and confidentiality rules that are in place that will govern how data is shared.

- **Where data is being shared from and to:** The question of 'where' is incredibly important in DSAs as it relates to jurisdiction. If data are being shared within one single jurisdiction, it is important that the DSA abides by all applicable laws and regulations. If data is being shared across jurisdictions, it is important that the DSA specify which jurisdiction applies to the contract. In cases of inter-jurisdictional data sharing, regional or international laws such as the EU's General Data Protection Regulation (GDPR) may specify additional compliance requirements.

INSIGHTS FROM THE PROJECT TO DATE

C4DC partners have analyzed some 30 agreements to date using the analytical framework. The initial sample has included concise MOUs and extensive legal agreements. The considered agreements span Africa, Europe, North America and Latin America and cover data describing issues from climate change to poverty. They concern data being shared between businesses, governments, and civil society and capture arrangements at the local, national and international levels. These agreements have been broken down into their constituent parts to understand and compare how the aspects of collaboration are being addressed throughout the data ecosystem. Additional work is required before specific conclusions can be made, but after undertaking this initial scoping and analytical work, it is now possible to reflect on some of the emerging trends that run through the project. Despite a number of initial reflections highlighting difficulties around accessing DSAs and negotiating them, there is also substantial support for the planned repository and the value it can add.

Many informal partnerships rely on non-binding agreements

Out of the initial pool of agreements that the team has located and analyzed, a significant number are in fact legally non-binding documents. MOUs, confidentiality agreements and similar documents appear to be common tools in this space. This is consistent with comments heard during interviews with practitioners. It suggests that parties may need to be more flexible when establishing partnerships. Although a legal contract can ensure a level of certainty, they can also cause delays (such as in disaster response and public health settings), and contracts are more costly, which can also create a disproportionate amount of risk for some. In those cases, a non-binding document can support some level of shared understanding to help guide the parties performance together, and help introduce the trust that is foundational to any agreement and partnership.

"Depending on the local context, a contract isn't always appropriate. Because it goes through a different approval process, it might go through a different department, partners may become concerned about the idea of being legally obligated to provide certain things. So MOUs are definitely very helpful as a middle ground to say – we are committed to sharing data, but we don't want to go through a contracting process."

– Beverley Hatcher-Mbu (Development Gateway)

Data sharing agreements take time to negotiate and vary greatly

Managing and transferring data assets between entities can be incredibly complex and this is reflected in the length of time it can take to negotiate and agree on DSA terms. Often, there are chains of responsibility that all touch upon the data in question and require multiple levels of approval before any agreement can be finalized. The length of time it can take to negotiate a DSA can also depend on the degree to which data governance processes are already formalized within organizations. Where governance processes are undocumented or lax, it can often take significantly longer to navigate the negotiation process than instances where formal policies and processes are set out clearly. Changes in legal and regulatory regimes can also impact on timelines as new compliance issues emerge and have to be addressed by the negotiating parties.

> "Our time estimates for entering into data sharing agreements vary greatly, it is hard to foresee what hiccups might arise along the way. In some recent negotiations we have also been held up by the introduction of the GDPR. When it came in, nobody understood how it should be translated into practice, so that caused a lot of delays. It was a good experience in the end, and we have a clearer understanding of the process now, but it was very challenging at the time."
>
> — *Tracey Li (Flowminder)*

4. INITIAL REFLECTIONS ON THE VALUE OF DSAS TO THE DEVELOPMENT AND HUMANITARIAN COMMUNITIES

There are two general reflections that sustainable development and humanitarian practitioners agree on that reinforce the relationship between the practice of sharing data and the broader context of the data revolution for sustainable development data:

DSAs should reflect good data management and governance practices

A key insight supported by practitioners is that DSAs are part of a broader regulatory and legal landscape. There is an interdependence between the need for legal frameworks in which DSAs can be contextualized but at the same time, in places where those frameworks are weak, DSAs themselves can help to inform their development and use.

"There's a lot of discussion around sharing, but not a lot of concrete connection between national strategies and work plans and how to actually implement them. While a lot of data sharing takes place within countries and government bodies, much of that is friend-to-friend or colleague-to-colleague. Sharing is not necessarily driven by an articulated need to have that data available for reporting, decision-making or other things. In practice, data sharing agreements are often not as powerful as legal frameworks for data sharing."

– Paige Kirby (Development Gateway)

Access to DSAs should promote responsible and ethical data sharing

There is increased recognition that data can be highly personal and that their sharing and use should be regulated and governed. As governments, businesses and individuals become more aware of the risks, the demand for greater protection and more ethical and legally regulated data sharing activity has grown. By opening up discussions on DSAs, and how and when they should be used, it is envisaged that data sharing practices generally will also improve over time, positively impacting individuals' rights over their personal data.

"From our experience there is a trend towards recognition of data being used for social good and the opportunities there. Within that context, there is increased recognition of the need to protect data. It is important to mention how the integration of these concepts works. You have to advocate for leadership to share data but clearly state that personal data is not going to be released."

– Fredy Rodriguez (CEPEI)

Overall, within the sustainable development and humanitarian sectors, the use of DSAs is crucial to the encouragement of more responsible and rights-based data sharing practices.

> "Regardless of whether or not it's a legal requirement, if you are working to help vulnerable people, then you have an ethical responsibility to ensure that the way you work with data respects the privacy of the people that the data is it is coming from. We have to minimise the risk of harm that can be done and while privacy regulations might be a compliance headache, at their heart they are very necessary to protect people's rights."
>
> – Tracey Li (Flowminder)

For the C4DC project, the aim is to find the balance between ensuring that data is used to inform evidence-based decision-making in pursuit of the 2030 Agenda, while at the same time making certain that data collaborations emerge in a safe and responsible environment. To this end, it is hoped that the C4DC project will help to demystify the practice of data sharing, enabling and expediting innovative data collaborations that also appropriately address and consider sensitive issues in their supporting agreements.

REFERENCES

Dahmm, Hayden; Espey, Jessica. (UNSDSN-TReNDS). 2019. Interview by Tom Orrell.

David, Scott. (UW). 2019. Interview by Tom Orrell and Hayden Dahmm.

Davis, Taryn; Hatcher-Mbu, Beverley; Kirby, Paige; Migwe, Charlene. (Development Gateway). 2019. Interview by Tom Orrell and Hayden Dahmm.

Espey, Jessica (2019). Data: The New Currency of Davos. Available at: https://www.sdsntrends.org/blog/2019/1/29/wef-data-currency-davos-espey

Hoffman, William. (WEF). 2019. Interview by Tom Orrell and Hayden Dahmm.

Jones, Harry (2012). Promoting evidence-based decision-making in development agencies. Overseas Development Institute. Available at: https://www.odi.org/publications/6240-promoting-evidence-based-decision-making-development-agencies

Li, Tracey (Flowminder). 2019. Interview by Tom Orrell and Hayden Dahmm.

Lloyd-Jones, Heledd. "Data sharing agreements: avoiding common pitfalls". Privacy and Data Protection Volume 15, issue 6 (2015).

Rodriguez, Fredy (Cepei). 2019. Interview by Tom Orrell and Hayden Dahmm.

Statista (2018). The 100 largest companies in the world by market value in 2018 (in billion U.S. dollars). Available at: https://www.statista.com/statistics/263264/top-companies-in-the-world-by-market-value/

The GovLab (2015). Data Collaboratives. Available at: https://datacollaboratives.org/introduction.html

The GovLab (2019i). Data Collaboratives Explorer. Available at: http://datacollaboratives.org/explorer.html

The Govlab (2019ii). Esri and Waze Open Data-Sharing for Governments. Available at: http://datacollaboratives.org/cases/esri-and-waze-open-data-sharing-for-governments.html

United Nations (2015). About the Sustainable Development Goals. Available at: https://www.un.org/sustainabledevelopment/sustainable-development-goals/

UNSDSN-TReNDS (2017). Counting on the World. Available at: http://unsdsn.org/resources/publications/counting-on-the-world/

UNSDSN-TReNDS (2019). Partnerships founded on trust: Introducing Contracts for Data Collaboration (C4DC). Available at: https://www.sdsntrends.org/research/2019/4/24/partnerships-trust-c4dc

Verlhust, Stefaan G. (2019iii). "Data Collaboratives: The Emergence of Public Private Partnerships around Data for Social Good." Presentation to the AI for Social Good Conference, Doha, February 17, 2019. https://qcai.qcri.org/wp-content/uploads/2019/03/Data_Collaboratives__Stefaan_Verhulst__GovLab.pdf.