Laying the Foundation for Effective Partnerships: An Examination of Data Sharing Agreements

By Hayden Dahmm, UN Sustainable Development Solutions Network's Thematic Research Network on Data and Statistics (SDSN TReNDS)

Abstract

In the midst of the COVID-19 pandemic, data has never been more salient. COVID has generated new data demands and increased cross-sector data collaboration. Yet, these data collaborations require careful planning and evaluation of risks and opportunities, especially when sharing sensitive data. Data sharing agreements (DSAs) are written agreements that establish the terms for how data are shared between parties and are important for establishing accountability and trust. However, negotiating DSAs is often time consuming, and collaborators lacking legal or financial capacity are disadvantaged. Contracts for Data Collaboration (C4DC) is a joint initiative between SDSN TReNDS, NYU's GovLab, the World Economic Forum, and the University of Washington, working to strengthen trust and transparency of data collaboratives. The partners have created an online library of DSAs which represents a selection of data applications and contexts. This report introduces C4DC and its DSA library. We demonstrate how the library can support the data community to strengthen future data collaborations by showcasing various DSA applications and key considerations. First, we explain our method of analyzing the agreements and consider how six major issues are addressed by different agreements in the library. Key issues discussed include data use, access, breaches, proprietary issues, publicization of the analysis, and deletion of data upon termination of the agreement. For each of these issues, we describe approaches illustrated with examples from the library. While our analysis suggests some pertinent issues are regularly not addressed in DSAs, we have identified common areas of practice that may be helpful for entities negotiating partnership agreements to consider in the future.

Policy Significance Statement

Data collaboratives bring together data from across government and other sectors to inform policy. Data sharing agreements (DSAs) describe the terms under which collaboratives are formed. This report offers insights and considerations for policymakers and other actors when developing these agreements. Our analysis demonstrates that DSAs are generally not addressing the full spectrum of pertinent issues, and expectations on how data should be handled have not yet been standardized. This suggests that further discussion on these topics is needed among policymakers to ensure more effective data partnerships. This analysis highlights key considerations for data partnerships to get these conversations started.

1. Introduction

The COVID-19 pandemic has heightened the demand for data, including new types of data to understand the different dimensions of the crisis. As a result, many countries are engaging in emergency data sharing arrangements (DSAs) to gain access to mobile network operator (MNO) data to help track and understand the effects of the crisis. In particular, there has been widespread use of MNO data to track population

movement and to quantify the impact of social distancing measures (Sibande, 2020). These activities reflect some of the more recent examples of data collaboration that have emerged as part of the wider data revolution for sustainable development (Espey et al., 2019).

Within the data ecosystem, National Statistics Offices (NSOs), as well as other government and multi-lateral actors have been increasingly turning to cross-sector collaborations to mobilize the vast data resources now available. For instance, the UN Environment Programme (UNEP) partnered with Google to create measures of surface water (Dahmm, 2019), Uber released data on traffic to aid transportation planners and city officials (Dwoskin and Siddiqui, 2017), and the telecommunications company, Airtel, shared data with the World Health Organization (WHO) to help combat Tuberculosis in India (GSMA, 2018). Such initiatives are providing public officials and decision-makers with detailed, timely data that is generating insights that were likely not possible or even affordable with traditional data methods.

Yet, data sharing activities during the pandemic have highlighted both new and existing challenges and concerns about data collaborations. Even before the crisis, data sharing prompted questions of security, privacy, and consent. These issues have become increasingly salient over recent years as the public learns more about how their data are being handled. New data protection laws, such as the EU's General Data Protection Regulation (GDPR), have not only significantly impacted both private and public sector practices, but they have also raised public awareness of privacy issues (Deloitte, 2018). Data sharing can expose the parties involved to liability, including risks to organizations or individuals whose data has been shared. For instance, in 2018, it was revealed that Facebook had shared user data with other technology giants, including Amazon, Microsoft, and Apple (Dance et al., 2018), despite already facing class-action lawsuits for having failed to protect its users' personal information during the Cambridge Analytical scandal (Davis, 2018). And concerns have extended beyond data sharing within the private sector. For instance, the University of Chicago shared historical medical data with Google as part of its research into artificial intelligence to predict medical events, but a former patient has sued the University in 2019, claiming that data was not appropriately de-identified (Wakabayash, 2019). Likewise, a new arrangement between the governments of the United States and the United Kingdom to secure mutual access to tech company data for criminal investigations has raised apprehension within the legal community about privacy protections (Rodriguez and Fischer, 2020). Doubts have also been raised about governments' access to sensitive data and the possibility of continuing access long after the pandemic (Timberg and Harwell, 2020). In particular, there are misgivings that MNO data sharing may in some instances go against human rights legislation (Human Rights Watch, 2020). Standards and appropriate protocols are needed to establish confidence and security.

Recent events have also underscored barriers to collaboration. The continued absence of large, multi-national data sharing systems in the healthcare sector has limited the ability of researchers to analyze COVID-19 treatments and monitor the spread of the virus (Cosgriff et al., 2020). Additionally, although the use of MNO data for tracking COVID-19 has grown, there have only been scattered implementation efforts and no broad internationally concerted efforts (Oliver et al., 2020). Many governments also lack the experience, technology, and capacity to utilize MNO data, and the data are often difficult to access because companies are reluctant to provide access to what they view as commercial assets (Oliver et al., 2020). The World Economic Forum (WEF) has noted how the intersection of risks around legal, social, technical, ethical, and commercial issues has hampered innovation and limited the use of private data for the public good (Hoffman et al., 2019).

DSAs are an essential element in managing collaborative relationships and navigating the associated risks. Although the term has broad applications, here we define a DSA as covering any form of written agreement - be it a legal contract, memorandum of understanding (MOU), non-disclosure agreement, or other mechanisms - that establishes terms for how data are shared between parties. The WEF describes DSAs in its holistic data governance framework as "essential tools for ensuring that stakeholders are held accountable to governing principles and that they understand the frontline risks, realities, expectations, roles, penalties, duties, and responsibilities as they embark on data-collaboration efforts" (Hoffman et al., 2019). While governments, private companies, researchers, and development practitioners usually negotiate these agreements to reflect their particular needs, there is a risk of them becoming imbalanced. Powerful and data-rich providers, like Google and Facebook, are often better placed to negotiate such agreements because they have more capacity and resources, which can result in them accruing more data rights and ownership. Many countries, meanwhile, are still gaining experience on when and how to create DSAs. The Philippines' NSO, for example, has experience with sharing data with researchers and other entities, including the NGO sector, but it is still working to access private sector data (Bersales, 2019). Furthermore, in low and middle-income countries, NSOs regularly lack the financial and technical capacity to meet the demand for data by domestic decision-makers (Sethi and Prakash, 2018), and according to officials we have spoken with, many lack legal counsel to negotiate DSAs. Negotiating DSAs often takes many months, and this multi-month process is often the most time-consuming and burdensome step throughout the entire process. These negotiation challenges, conflicting priorities, and sensitivities often at play, underscore the need for an open discussion on how these agreements are created to ensure more ethical and effective data collaborations.

Contracts for Data Collaboration (C4DC) is a joint initiative that seeks to strengthen the trust, transparency, and accountability of cross-sector data collaboratives. To better understand the legal conditions that can enable effective data collaboration, the Sustainable Development Solutions Network's Thematic Research Network on Data and Statistics (SDSN TReNDS), the Governance Lab at New York University (GovLab), the World Economic Forum (WEF), and the University of Washington's Information Risk Research Initiative have developed an online library of redacted DSAs used in a variety of settings and countries that are accompanied by associated use cases, analysis, and resources that demonstrate how DSAs are constructed and applied. These agreements cover a range of data applications and consider a host of legal issues in the use of data. Governments that might be under-capacitated can access examples; companies that might not have prior experience with data sharing can see what is possible; and data collaborators can be more empowered with the legal knowledge to ensure more fair and balanced negotiations between parties. Ultimately, standard practices and expectations could emerge.

This report serves to introduce the C4DC initiative, its DSA library, and provide collaborators with a more comprehensive understanding of how DSAs are applied in practice. First, we explain the process taken in analyzing the individual agreements. We then consider how six key issues are addressed by different agreements in the library. For each of the highlighted issues, we describe observed approaches illustrated with examples. From these observations and examples, we draw conclusions and present recommendations for how the DSA library can continue to support the data community.

2. The DSA Framework

DSAs can take on many forms, ranging from highly specialized to general and less formal. However, there is ongoing momentum within the data community toward standardizing data sharing activities. According to the Chief Data Officer of the U.S.-based Urban Institute, which undertakes regular data sharing efforts, each data sharing project is different and unique (MacDonald, 2019). When working with companies interested in contributing data for the public good, though, the Institute will often guide them to certain standards (MacDonald, 2019), and they have published guidance on the elements addressed in their private data sharing agreements. Additional entities have also shared guidance to help lower barriers to data sharing. For example, Microsoft recently published a set of draft DSAs (Microsoft, 2019). U.S. government agencies are also working to create template DSAs with standard terms and components; this stemmed from the realization that it takes longer for them to negotiate agreements than to analyze the data, and agreements that are not constructed correctly can create difficulties later on (Madans, 2020).

A DSA addresses several issues, ranging from generic contracting matters to questions that are highly specific to data. Founded on our combined experience and expertise, the C4DC team iteratively defined an analytical framework to help guide conversations around these issues. The GovLab's Contractual Wheel of Data Collaboration (Verhulst, 2019) was used as a framing structure for our analysis, separating the components of a DSA according to six overarching questions: (1) Why is the data being shared?; (2) What kinds of data are being shared?; (3) When should the data be shared?; (4) Who is involved in the data sharing?; (5) How is the data being shared?; and (6) Where is the data being shared from and to? 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, covering a total of 49 detailed questions, can be found on the Contracts for Data Collaboration website.

Why Is The Data Being Shared?

DSAs should provide details of the context and purpose for why the data are being shared. Particularly when parties are balancing competing incentives, forming a common value proposition can aid in the delivery of the collaboration (Hoffman et al., 2019). Establishing clear goals can also guide actions and help to justify the partnership to other parties.

What Kinds Of Data Are Being Shared?

DSAs should aim to be as specific as possible about what types of data are and are not being shared between parties. For example, does the DSA involve individual observations, summary statistics, geolocation data, or other categories? It should also clarify if data about individuals will be shared. Personal data that can be used to identify an individual, sensitive personal data that cover special categories of data about individuals, and derived data produced by processing other datasets each involve different types of information, but all require special procedures to protect privacy. Moreover, DSAs should ideally specify the standards, formats, and rules that apply to the data in question, including the specific source of the data, the metadata, and other technical requirements.

When Should The Data Be Shared?

DSAs should be clear on timelines, including the agreement start and end dates, as well as important milestones throughout the agreement period. They should also provide details on how long the data should be stored and used following the end of the agreement, up to the point of safe disposal or archiving.

Who Is Involved In The Data Sharing?

An explanation of the relevant parties is fundamental to a DSA. This includes a description of the legal status of parties, their authority, who is providing and receiving the data, and other roles and responsibilities. Additionally, it should identify third parties that will be impacted by the data sharing. The DSA might also allow for further sharing of the received data with other parties, and any further sharing of the data is usually carefully considered before entering into an agreement and is negotiated at the outset.

How Is The Data Being Shared?

DSAs should also specify the procedures by which the data should be shared. Among other items, it should lay out the data security and confidentiality rules that will govern how data are handled and protected, as well as describe how the data will be hosted, analyzed, and the logistical arrangements.

Where Is The Data Being Shared From And To?

The location where the data sharing activities take place can be significant to the formation of a DSA. For example, if the data are shared within a single jurisdiction, it is important that the DSA abides by all applicable laws and regulations. If data are being shared across jurisdictions, then the DSA should clarify which jurisdiction applies to the agreement. In cases of inter-jurisdictional data sharing, regional or international laws such as the GDPR may specify additional compliance requirements.

3. DSA Library

C4DC's online library of DSAs documents how these questions are being answered in practice. As of June 2020, the library features 41 agreements gathered from public sources and contributed by other organizations. While we plan to build out a more diverse collection, the sample represents countries from five continents, data from a number of key sectors (Table 1), and types of data sharing collaborations from public-private to public-public and public-NGO (Table 2).

Partly a consequence of which agreements are openly available, the majority of the collection features agreements between governments sharing data with the private sector and NGOs, as well as with other government agencies. The library is intended to be a tool that can help parties think through what they need their own DSAs to cover and consider connections with broader data governance issues. However,

the library is intended to serve only as a learning tool and we do not offer this as actual legal guidance. It is not our intention for any of the hosted agreements to be used as templates by others without first seeking proper legal advice.

Below, we explore how DSAs are addressing some of the most significant questions facing the data community. As the complete 49-question framework demonstrates, the number of considerations that should form the basis of a DSA is significant, but we have chosen to focus on six additional questions for this report that were highlighted as especially relevant during our conversations with key stakeholders. These questions include: (1) How is data use limited?; (2) Who else can access the data?; (3) How are data breaches managed?; (4) How are proprietary issues handled?; (5) How will the activities be publicized?; and (6) Does the data need to be deleted when the agreement is terminated?

For each of these questions, we have reviewed the agreements in our library and present initial insights. The analysis draws on agreements from a range of collaborations, but our overarching aim is to provide replicable guidance for public actors looking to negotiate an agreement with other actors in a way that expands the usage of new data sources.

Table 1: Represented Sectors in the C4DC Library*

Sector	Number of Agreements
Health	12
Economic Development	9
Environment	6
Telecommunications	5
Education	4
Criminal Justice	2

*Note: Some DSAs touch on more than one sector.

Table 2: Types of Collaborations included in the C4DC library

		Data Contributor				
		Academia	Government	NGOs and International Organizations	Private Sector	Not Specified
Data User	Academia	1	2	1	1	-
	Government	2	11	7	6	-
	NGOs and International Organizations	1	9	3	7	-
	Private Sector	-	2	1	-	-
	Not Specified	-	2	-	-	2

***** Some agreements involve more than one category of contributor or user, and they are counted twice in this case.

3.1 How Is Data Use Limited?

When data are shared, there is always a risk of it being used for reasons other than the stated purpose. To avoid abuse or unforeseen consequences, many agreements will outline limitations. Often, this involves stating specific limitations on what kinds of data cannot be shared and explaining how the data can and cannot be used. Some agreements may take a general form, allowing the data to be used in a variety of ways with only select restrictions, while others place more extreme constraints. For example, when the U.S. National Center for Health Statistics is forming an agreement, it has a particular interest in ensuring that the data is only used for statistical purposes and how this is reflected in writing (Madans, 2020). Additionally, Facebook has a Data for Good program that provides data from its platform to aid in health and natural disaster situations, usually according to the terms of a template data license agreement that it created (McGorman, 2020). Reflecting the natural sensitivity of Facebook user data, this agreement only permits the sharing of aggregated, de-identified data. Facebook allows partners to download data and create derivative products, which they can share publicly, but there are purpose limitations included in their agreements around use of this information for social impact purposes. Articulating the limitations on data use is necessary to a thoughtful data sharing arrangement, and our library demonstrates the variety of ways to undertake these issues.

A Simple Limitation

Data use limitations are sometimes simply stated with a generic line. In short, any uses that do not support the purpose of the data collaboration are not allowed. This is the case with a DSA from Scotland that involves public housing data, which included the following clause:

Text from the Agreement: *"Save as provided for in this Agreement, the Parties agree not to use any Data disclosed in terms of this Agreement in a way that is incompatible with the Purpose."*

Although this clause establishes the expectation that the parties will use the data responsibly, it remains broad and potentially open to interpretation. The parties could realize a new application for the data that wasn't previously considered that is still in keeping with the initial purpose. Such a flexible constraint still provides the parties with the leeway to explore new applications.

More Specific Limitations

At the same time, not defining what is "incompatible with the purpose" could create unnecessary risk or uncertainty. Some other agreements clarify very explicit limitations. This is especially common when the DSA relates to people in vulnerable situations. For example, an agreement between the Samoa Bureau of Statistics (SBS) and UNICEF that references data on children and the disabled carefully elucidates restrictions on the potential use of individual data. First, the agreement text states that data can only be used for determining information about groups of people, and it specifies that the data cannot be used for identifying individuals, families, or businesses. It then goes a step further and requires that if such an identification should happen by accident, the information cannot be used, and the event must be reported to the SBS.

Text from the Agreement: *"The data will be solely for reporting of aggregated information, and not for investigation of specific individuals or organizations. No attempt will be made to identify any individual person, family, business, enterprise, or organization. If such a unique disclosure is made inadvertently, no use will be made of the identity of any person or establishment discovered and full details will be reported to SBS. The identification will not be revealed to any other not included in the Data Access Agreement."*

Agreements involving personal medical data also often take additional steps to address the clear sensitivities. Health agreements generally prevent the identification of patients, the publication of their data, or contacting individual patients. For instance, a template agreement from the Health Data Coalition (HDC) in Canada does not allow for the use of personal health data. It states that the parties will work to avoid the sharing of such data in the first place. Furthermore, the data holder commits to manage its wider database in a way that protects individual privacy.

Text from the Agreement: *"The Parties acknowledge and agree that it is not the intention of either Party to share patient personal information or personal health information in identifying form via the HDC application, and HDC will use all reasonable efforts to ensure no such information is submitted to the HDC central server; To clarify further, HDC's application will not use or share any personal health information, and HDC will assess the impact of HDC's national database operations on individual privacy if any research project or Quality Improvement project creates a risk of re-identification or involves the use of identifying personal health information to create linkages with other data sources or other data access requests for disclosure of record level patient data*."

By placing definitive bounds on the ways data can be used, these agreements strengthen the security of individual data subjects. While this may restrict positive applications of the data, ensuring its safe and responsible use are prioritized.

Limitations With Exceptions

Unanticipated circumstances and data needs can arise, so the ability to make special allowances is helpful. A few agreements in the library describe use limitations, but also include some flexibility by creating a review process and allowing the data provider to make possible exceptions. One sample agreement between a state Department of Corrections (YDOC) and a non-profit (X) prohibits data about recent prisoners to be used for various legal documents unless the department provides written permission. The non-profit is responsible for considering other requests from an established re-entry mapping group (REMG) to access the data and ensuring that uses are consistent with the stated purposes.

Text from the Agreement: *"Organization X shall: b. Review requests for maps and analysis results from REMG members to ensure that the requests are consistent with REMG's goals of identifying reentry patterns and resource gaps. Requests that are not consistent with these goals will not be pursued. ... VI. Re-dissemination Data and information furnished by YDOC to Organization X may not be used for any legal documents, such as search warrants, affidavits, or any document that may become public, or re-disseminated or used for any purpose other than that for which it was requested, without the written permission of YDOC."*

The agreement provides a mechanism for exceptions to be made, but it also states clear expectations for how the use of personal data should be limited. While the ability to respond to unexpected data requests is important, outlining procedures for how the parties will uphold the initial purpose of the agreement is also integral.

Commercially-Sensitive Limitations

Beyond sensitive personal information, data sharing with the private sector may impact commercially-sensitive information as well. Agreements can address this sensitivity and outline use limitations. For example, an agreement that collects information from health providers in the Bay of Plenty, New Zealand recognizes the significance of commercial data. It also prevents the sharing of commercial data unless deemed necessary, and in the case that it is shared, it limits its use.

Text from the Agreement: *"Constraints on Use: The Parties recognize that each will hold commercially sensitive data and that inclusion of such data in a Data Sharing Arrangement may disadvantage the sharing Party. The Parties agree: i) Commercially sensitive data shall be excluded from Data Sharing Arrangements unless its inclusion has been specifically agreed in writing by the Parties; ii) Where a Party agrees to share data that may be of commercial value to it, the other Party shall not access or use that data except for the agreed purposes of its sharing."*

When dealing with private sector data sharing, concerns about commercial sensitivity may be an obstacle to collaboration. By specifically excluding commercially-sensitive data, these misgivings can be relieved.

Discussion

Establishing data use limitations are important to building trust when sharing data. There are several ways that limits can be developed to align with the different priorities of data providers and receivers, as well as the interests of individual data subjects. Less strict limitations can expand the potential opportunities for data receivers, whereas stricter limitations can address commercial or security issues. The library highlights a spectrum of ways in which these limitations can be formulated and included in agreements. Yet according to our analytical framework, less than two-thirds of the DSAs currently in the library (26 out of 41) address use limitations in some manner, and at least eight of the agreements that do not set limitations involve potentially sensitive information, including data about students, housing, and mobile networks. Although the practice is relatively common, language about data use limitations is not yet standardized or universal.

3.2 Who Else Can Access The Data?

Beyond determining how the primary parties will share and access the data, a DSA should address if other parties have access. Providing opportunities for wider sharing can help leverage the value of data, especially for the public good. For instance, the Facebook agreement described in section 3.1 was signed by the NGO, Direct Relief, and it enabled the organization to share aggregated, de-identified maps of estimated population movement with partner organizations during disaster situations (Schroeder, 2020). As a result, local public authorities were able to use this information to respond to wildfires in California and hurricanes in Texas. However, when dealing with data of a confidential nature, there are many important issues to consider around third-party access. The library presents a scope of circumstances and paths that can be taken.

Making The Data Open

If the data being shared does not contain attributes that could lead to the identification of individuals, and if the information does not involve security risks, parties have the option to make the data available to the public. The Open Data Charter, a set of international principles adopted in 2015, calls for data to be open by default in a way that is accessible and usable (The International Open Data Charter, n.a.). This approach was taken in a data sharing arrangement between Google and UNEP in 2018 (Dahmm, 2019). As part of the arrangement, Google provided access to measures of surface water, and the two partners created an open platform to establish a global indicator of surface water that has supported the work of UNEP and member countries and has even led to official Sustainable Development Goals (SDGs) reporting. This intention was laid out in their MOU.

Text from the Agreement: *"To ensure holistic monitoring of the environmental dimension of sustainable development, the Parties also agree to collaborate in terms of data dissemination and visualization. Under this area of collaboration, the Parties agree to make derived data products of the collaboration public and freely available (when possible based on the license of input datasets, other partner restrictions… the spirit of open data, and using standard formats for distribution when possible)."*

Opening up data can maximize the benefits and lead to important insights that might not otherwise be possible. Committing to this in writing is an important step to realizing these potential insights, and in effect, it expands the scope of the agreement.

Third-Party Sharing With Prior Approval

When making data free and open to the public is not possible, extending access to third-parties on a case-by-case basis might be an option. In this situation, the original data provider is often required to provide prior approval. The third-party may also, in effect, become an additional signatory to the agreement. For example, The Rhode Island Department of Health, formed a DSA with a local NGO to share health data, and it establishes both of these conditions.

Text from the Agreement: *"2. The data recipient will not release data to a third-party without prior approval from the data provider. … 6. Any third-party granted access to data, as permitted under condition #2, above, shall be subject to the terms and conditions of this agreement. Acceptance of these terms must be provided in writing by the third-party before data will be released."*

Allowing for additional recipients to join the existing agreement creates the opportunity for the data to be employed in more ways with minimal hassle. Yet, unlike an open data arrangement, this practice recognizes the sensitivity of the data and allows the provider to use discretion when expanding access. Furthermore, by holding any additional recipients to the original terms of the agreement, the same levels of accountability are extended along with the access.

Limiting The Type Of Data That Can Be Accessed

An alternative to requiring a third-party to abide by all the terms of the agreement is to place strict constraints on the type of data that can be shared with third-parties. This has the potential to protect the confidentiality of the data while still permitting useful insights to be more widely circulated. For instance, Louisiana's Department of Education formed a DSA with a scholarship organization (STO) regarding student data. The agreement expressly forbids STO from sharing the department's student data, unless the data are aggregated and students cannot be identified.

Text from the Agreement: *"STO cannot publish any document, whether in hard copy or electronic form, or otherwise disclose to any third-party any student-level data or information in any form whatsoever in data sets and/or cell sizes of less than ten (10) or under any circumstances which would directly or indirectly make a student's identity easily traceable."*

This allows for valuable information and insights to still be extracted from the data and circulated, while also maintaining individuals' privacy. Collaborators may find this to be a worthwhile solution for both supporting and protecting the communities they serve.

Restricted Data Access

Some agreements do not include conditions for third-parties to access data. Indeed, data can be regarded as so sensitive that agreements restrict which individuals can access the data and where this can be done. Similar to the Louisiana agreement, Oakland California's school district and a local community college formed a DSA around student data. However, the agreement is far more restrictive. Student data are required to be kept in a secure location, and it is only available to authorized personnel.

Text from the Agreement: *"Procedures and systems that ensure all student records are kept in secured facilities and access to such records is limited to personnel who are authorized to have access to said data under this section of the MOU."*

This restriction reflects a prioritization of data security. And while third-parties with legitimate interests might not be able to easily benefit from the data collaboration, the signatories have chosen to emphasize protective procedures. When negotiating an agreement, collaborators have to navigate a range of options around access, including weighing the benefits and trade-offs between enabling third-party applications and fundamental data protections. The eventual decision is a function of both the type of data being exchanged and the interests of the parties involved.

Discussion

Providing open data supports the equitable and democratic distribution of information, and this can be done in a conscientious way that aims to solve public problems, while protecting the privacy of individuals (Badiee, 2020). For example, the government of Paraguay has used open health data to construct an early warning system for detecting dengue fever, and Namibia's government worked with its largest MNO to produce open data for combatting malaria (Badiee, 2020). The library shows how a commitment to open data can be established in a DSA. In the case of sensitive data, the parties might decide that no third-party should be allowed any form of access, but alternative arrangements can be negotiated. From requiring that third-parties become signatories to the agreement to only permitting access for aggregated data, there are constructive ways that the data can be made more widely available while also protecting security. Within the current library, 25 of the 41 agreements address the question of who else can access the data, which highlights that it is an issue that is regularly considered but may not always be resolved within data collaborations.

3.3 How Are Data Breaches Managed?

Even with established measures to limit data use, manage third-party access, and maintain security, there are risks to placing data with another party. 13 of the 41 agreements currently in the library, address what should happen in the event of a data security breach. Several define this as being "any act or omission that compromises either the security, confidentiality, or integrity" of the data. Notably, the agreements that include language on breaches deal with data that could be of a personal or sensitive nature.

Notification Of A Breach

In the event of a data breach, there are some common expectations of the parties. At a minimum, the data receiver is required to promptly inform the provider when a breach does occur. For example, the International Organization for Migration (IOM) has an agreement for transferring personal data of its beneficiaries to other parties and specifies what should happen in the event of a data breach.

Text from the Agreement: *"Immediately notify IOM in writing upon becoming aware of any data breach, in particular if the data breach is likely to result in personal injury or harm to the data subjects."*

Providing an immediate notification enables the provider to respond promptly and may help to mitigate any potential damage.

Shared Breach Response

However, some agreements, place added responsibilities on the data receiver to minimize the risk of a breach. Common terms include notifying the provider within a defined time window, specifying the type of information that must be communicated about the possible breach, and assisting the provider with an investigation of the breach. For example, as part of a 2018 DSA, MNO data from an operator was shared with an intermediary for a government in Sub-Saharan Africa to use in its health system, and it included stringent procedures.

Text from the Agreement: *"VII. Security Breach Procedures: The [INTERMEDIARY] agrees to fully cooperate with the [DATA CONTROLLER] in any investigation, litigation or other action deemed necessary by the [DATA CONTROLLER] in connection with the security, use, and disclosure of Data, including Personal Information. The [INTERMEDIARY] shall report any confirmed or suspected Security Breach to the [DATA CONTROLLER] immediately upon discovery, both orally and in writing, but in no even more than five (5) Business Days after the [INTERMEDIARY] becomes aware that a Security Breach has or may have occurred. The [INTERMEDIARY]'s Security Breach report above, shall identify: (i) the nature of the unauthorized access, use, or disclosure, (ii) the Personal Information accessed, used, or disclosed, (iii) the person(s) who accessed, used, and disclosed and/or received Personal Information (if known), (iv) what the [INTERMEDIARY] has done or will do to mitigate any deleterious effect of the unauthorized access, use or disclosure, and (v) what corrective action the [INTERMEDIARY] has taken or will take to prevent future unauthorized access, use, or disclosure. The [INTERMEDIARY] shall provide such other information, including a written report, as reasonably requested by the [DATA CONTROLLER]. In the event of a suspected Security Breach, the [DATA CONTROLLER] shall keep the [INTERMEDIARY] informed regularly of the progress of its investigation until the matter/issue is resolved."*

Establishing clear standards and procedures in the event of an emergency is important for minimizing confusion and helping to manage potential risks. A structured plan included in the agreement allows for the data receiver to have a clear understanding of actions to take should there be a breach. And mobilizing both the data receiver and the data provider in response efforts could improve the outcome. Additionally, it may hold receivers more accountable and give data providers the necessary confidence to permit data access.

Discussion

From the library, terms about data breaches appear to be introduced only when unauthorized access would threaten sensitive information, but there seems to be more standardization in how to respond than on other issues. Several DSAs use the same definition of a breach, and similar responsibilities are placed on the parties. One variable is the level of commitment from the data receiver to simply inform the provider of a breach or to actively aid in the response efforts and investigation. Such measures may require more advanced planning and a certain level of capacity on the part of the receiver. Yet, less than a third of the DSAs currently in the library actively consider data breaches, and more may need to be done to prepare the data community for such an event.

3.4 How Are Proprietary Issues Handled?

When data are shared between two parties, unlike a transaction of normal goods, both parties hold the data. This raises a host of questions about who then owns the data. One of the most important topics that arises in a DSA is data ownership and associated intellectual property rights.

Ownership Of Data

Not all agreements in the library address the issue of data ownership. When they do, though, it is usually to ensure the continued ownership rights of the data provider. The WHO has a template DSA for gathering data on tuberculosis from countries and other organizations. In the agreement, the WHO ensures that its data providers maintain full ownership.

Text from the Agreement: *"The [INSERT ACRONYM OF TRANSFERRING PARTY] retains all ownership rights to the data. The WHO Central Database will not claim any rights of ownership in any data submitted, but in order to allow the WHO Central Database Management Committee to operate, it requires your permission to use the data for the purposes set out in this agreement."*

Protecting existing ownership may assuage data contributors' concerns, as they are not giving up their rights and can maintain ultimate control of their product. While this might curtail the actions of the receiver, such an assurance could widen the scope of datasets available.

Ownership Of Analysis

Proprietary issues extend beyond the provided data. Agreements are created so that data can be used and manipulated, and some agreements also address the ownership of the resulting analysis and products produced from the collaboration. This ownership can be assigned to either party, or both. For example, the U.S. Council of Large Public Housing Authorities created an agreement to facilitate data sharing between housing authorities and intermediaries, such as researchers. When these intermediaries develop a product based on the shared data entirely on their own, they have undivided ownership of the product. However, if the product is produced in conjunction with the authority, the ownership is shared:

Text from the Agreement: *"4. Intellectual Property (a) Any and all tangible materials, analysis and reports, regardless of format, delivered by and developed or created solely by [Intermediary Organization] ("Work Product") shall belong to [Intermediary Organization]. [Intermediary Organization] shall grant the Authority a perpetual, nonroyalty bearing, world-wide license to use, reproduce, publish, and distribute the Work Product in accordance with this Agreement. The Authority agrees to recognize its use of the Work Product by including an attribution in a prominent location within publications, reports, or other materials that acknowledges the contribution of [Intermediary Organization] and/or use of the Work Product.; (b) Any reports developed or created by both [Intermediary Organization] and the Authority will be jointly owned."*

Addressing ownership of outputs from the onset provides clarity and can minimize conflict. This example gives added protection to the receiver as well, by guaranteeing that it has sole ownership of any analysis it produces on its own.

Discussion

Based on the available sample of DSAs in the library, terms about proprietary issues are often focused on confirming the data provider's ownership rights. They may also guarantee ownership rights for outputs from the collaboration. However, less than half (20 out of 41) of the DSAs address proprietary issues. And although many do not address the concern, ownership rights can become highly complex and intersect with local laws and regulations. Even if ownership of data is not explicitly conferred, for example, a data receiver may be able to claim database rights. This suggests that a more systematic consideration of data ownership may be needed in collaborations going forward.

3.5 How Will The Activities Be Publicized?

Related to the above issues of ownership and third-party access, having the right to publish the results of data analysis can be critical for data recipients (MacDonald, 2019). Even if data or the results of the analysis are not strictly available for publication, agreements may include additional stipulations about the publication process.

Approval Before Publication

The majority of the library's DSAs (25 out of 41) make a point of specifying how the results of a data analysis can be published. It is almost universal to require that the results first be shared with the data provider for approval. A health data agreement from Rhode Island states this simply.

Text from the Agreement: *"The data recipient will not share, publish, or otherwise release any findings or conclusions derived from analysis of data obtained from the data provider without prior approval from the data provider."*

Other agreements describe more particular procedures, including timelines, review boards, and the type of changes the data provider can make before publication. For example, the 2016 agreement from the U.S. Council of Large Public Housing Authorities includes the following conditions.

Text from the Agreement: *"Publication: (a) [Intermediary Organization] agrees to provide the Authority with an advance copy of any publication resulting from the Scope of Work not less than thirty (30) days prior to the submission or disclosure of the publication, to permit the Authority to reasonably comment, update, redact, or otherwise propose modifications or edits to the draft publication, and to ensure there is no disclosure of Confidential Information."*

Requiring advanced approval allows the provider to confirm that the data has been analyzed accurately and appropriately. Such an oversight procedure can also be to the advantage of the data receiver because the provider has a unique understanding of how the data was produced and what it represents. Moreover, advanced approval allows the provider to confirm that the terms of the agreement are being followed before the relationship is publicized.

Providing Credit And Disclaimers

Data recipients are usually obligated to cite the data provider in their publications or presentations, and some data providers require a disclaimer about their responsibility for the analysis. An agreement in 2016 between Canada's Health Data Coalition (HDC) and its clinical data contributors includes both of these requirements.

Text from the Agreement*: "Unless directed otherwise, HDC must be acknowledged in any publication or presentation using HDC data, and the following disclaimer must appear on any materials developed for public distribution with data used under this DSA: 'The views expressed herein do not necessarily represent the views of HDC.'"*

Data providers generally appreciate recognition for their efforts, and this could indeed be an underlying motivation for participation. At the same time, they may want to mitigate the risks associated with outputs that they did not co-produce. An agreement can address both of these areas.

Discussion

Publishing the results from a collaboration can support transparency and generate knowledge, even if access to the underlying data is restricted. DSAs can outline procedures for publication and give parties the assurances that results will be both vetted and made public. Additionally, most of the language on publications focus on the interests of the data provider, guaranteeing recognition, and the right to prior approval.

3.6 Does The Data Need To Be Deleted When The Agreement Is Terminated?

Collaborators will need to consider how to handle data in the long term and if the data should be retained for future use or deleted. The Centre for Humanitarian Data has provided guidance on this issue from a humanitarian perspective and advises that "following the retention principle, humanitarian data should be retained as long as its foreseeable potential value outweighs risks associated with retention. Sensitive data should only be retained for the time that is necessary for the specified purpose" (OCHA, 2019). If the balance of data risk and utility favors deletion, then data must be thoroughly removed from all devices.

Only 14 of the 41 library's DSAs address if and how data should be deleted, and there is some variation in the strictness of requirements.

Retaining Data

An agreement may allow the parties to retain data that has already been shared. This is the case in a 2019 agreement from Kings County, California that facilitated data sharing between partner agencies for a Homeless Management Information System (HMIS), managed by a private company.

Text from the Agreement*: "If this Agreement is terminated, the County and all participating Partner Agencies maintain their rights to the use of all client information previously entered into the HMIS, subject to the terms of this Agreement and other applicable rules, regulations, and agreements…Upon any such termination of this Agreement, the Agency may request and receive one export copy of all data entered by it into the HMIS from the Effective Date up to the date of termination. If such a copy is requested, the Partner Agency will be responsible for reimbursing the County for the costs associated with producing the report."*

This approach recognizes that while the partnership might not continue for various reasons, the need for the data will continue. Even though the recipient will not receive updated data, retaining the data that was already provided will help achieve the purpose of the agreement.

Returning And Deleting Data

Alternatively, the agreement can delineate rules for returning data to the provider and deleting any remaining copies. For example, the Baltimore Education Research Consortium (BERC) created an agreement about student data that requires the data to be deleted upon termination of the agreement.

Text from the Agreement: *"Upon the conclusion of this MOU or any succeeding MOU with regard to BERC, all parties with City Schools data shall deliver to the City Schools, upon the school district's request, all of the data obtained (including all documents, technology, software and all copies thereof) in whatever medium that contains the data provided as a result of the BERC MOU."*

Requiring that data be deleted establishes protection for the provider. While this means that the receiving party will not be able to analyze the data once the agreement is terminated, it gives the provider added security, knowing that its data are not outside of its control.

Deleting Data While Retaining A Copy

Even with strong return and deletion rules, the data recipient might be allowed to retain a copy of the data, as demonstrated in an agreement with the Texas Statewide Data Exchange Compact (TSDEC). Under the terms of the agreement, a receiving agency can retain data if necessary, but it would still be responsible for following the terms of the agreement.

Text from the Agreement: *"6.8. Upon termination of the Scope of Information Exchange, receiving agency will return or destroy confidential information received from disclosing agency to the extent reasonably*

feasible and permitted by law. If receiving agency is required by law or litigation to retain confidential information beyond the termination of the Scope of Information Exchange, receiving agency will continue to safeguard the confidential information in accordance with this TSDEC."

Discussion

With governments worldwide pursuing unprecedented access to MNO data and other sensitive information as part of the COVID-19 pandemic response, advocates worry that the data could be retained or accessed well past the end of the crisis (Human Rights Watch, 2020). And a conversation about how collaborators handle data in the long-term has never been more pressing. The C4DC library offers insights into the ways that DSAs have recently addressed these issues. As noted, some agreements allow recipients to retain data indefinitely, although this presumably depends on the level of trust between the collaborating parties. Other DSAs require that all data be completely deleted by the receiving party, while some permit a copy to be retained for record purposes. Generally, the agreements that discuss retention or deletion involve the exchange of sensitive data. However, with only a third of the library's DSAs actively addressing this issue, our sample suggests that language is not common or standardized.

4. Conclusion

The demand for data sharing and collaboration continues to grow at an unprecedented rate, and cross-sector partnerships have gained considerable attention in recent years (Hoffman et al, 2019). While these partnerships promise novel insights, innovation, and humanitarian benefits, the data community should invest in managing these relationships responsibly. Data sharing agreements are key to initiating and guiding data collaboratives, but the investments of time and uncertainty around negotiating and preparing these agreements suggest that a wider conversation is needed. Amid concerns around data privacy and use, demonstrating how these agreements are being structured in practice can help guide action and ensure more effective partnerships moving forward.

As demonstrated, the C4DC library is documenting how DSAs are being formed across a variety of contexts, and this report highlights the types of insights, guidance, and examples gleaned from the library to help inform future collaborators when negotiating their agreements. Although we have profiled six core questions, not all of the agreements in the library address these questions. Nor do any of the 41 agreements address all 49 questions from the analytical framework. Furthermore, not every topic is essential to a comprehensive DSA, and this often depends on context. However, this does indicate that current data sharing practices are generally not addressing the full spectrum of issues they might encounter.

While many of the underlying issues are common, this report demonstrates that there is no "universal" DSA. The circumstances faced by one team of data collaborators are often unique, and the level of detail given in agreements varies. As expected, agreements involving data on sensitive topics or vulnerable individuals generally take more precautions. In particular, greater limitations are placed on the use of data about refugees, students, or medical patients. However, based on the limited sample, it does not appear as though collective approaches to regular issues are emerging. We highlighted two agreements that are

sharing individual student data, one of which does allow for the further distribution of restricted data, while the other does not make such an allowance. Yet, our analysis shows that there are identifiable clusters of approaches. With further development, these could lead to a selection of standard terms to expedite negotiations and establish uniform expectations and accountability among parties.

The issues of third-party access, ownership, and publication are also interconnected. They each address the extent to which the control of data is extended to the receiving party or maintained by the provider. At one extreme, the receiving party may have considerable independence, including the ability to share with third-parties, claim ownership of the analysis, or publish with fewer restrictions. At the other end, the data provider may have different constraints on these activities, such as requiring prior approval or forbidding certain actions in the agreement itself. These questions should be considered in concert, and the library highlights a range of available practices.

Furthermore, the responsibility for managing data extends beyond the planned length of the project. Agreements need to address the long-term issues, including what will happen after the agreement is finished and what will be done with the data at that time. Only 14 of the library's agreements address this issue, and it is unclear how the data will be managed in the long-term in the majority of cases. In the DSAs presented, student data needed to be deleted and returned, while data on the homeless could be retained indefinitely. This is not to criticize either arrangement, but based on the sample, there does not appear to be standardized approaches for how sensitive data should be managed after the agreement is completed.

Although the library is already beginning to document and offer insights on data sharing practices across the sustainable development space, there is a clear need to expand the platform to capture a wider sample. Out of the 41 available agreements, 20 are from North America, six are from Europe, and only two were not in English. The local legal context can be significant to how a DSA is negotiated and implemented, and achieving greater geographic diversity will be necessary to better document practices and inform the community. Also, there are major categories of data, such as satellite imagery, that are not yet represented in the library, which would allow for broader guidance. By continuing to gather DSAs and document these experiences, we can foster a more open dialogue on data sharing practices and ensure that future collaborations are effective, secure, and underpinned by trust.

Acknowledgments

This report is a pre-print and was written by Hayden Dahmm (Analyst, SDSN TReNDS), with input from Jessica Espey (Director, SDSN TReNDS) and Tom Orrell (Director, DataReady).

We are grateful to the many data sector experts who generously shared their insights and experiences. In particular, we thank Graham MacDonald (Chief Data Scientist, Urban Institute), Jennifer Madans (Associate Director for Science, National Center for Health Statistics), Laura Walker McDonald (Senior Director, Digital Impact Alliance), Lisa Grace Bersales (Professor of Statistics, University of the Philippines), Laura McGorman (Policy Lead, Data for Good at Facebook), and Andrew Schroeder (Vice President of Research and Analysis, Direct Relief).

DSAs in the C4DC library were analyzed by Hayden Dahmm, Scott David (Director of Policy, Center for Information Assurance and Cybersecurity, University of Washington), and Mamali Mohapatra (Intern, SDSN TReNDS). We also thank Alyson Marks (Communications Manager, SDSN TReNDS) for her assistance with editing.

Bibliography

Badiee, S. (2020, April 9). Is Open Data at Odds with Citizens' Privacy?. Apolitical. Retrieved from https://apolitical.co/en/solution_article/is-open-data-at-odds-with-citizens-privacy

Bersales, L. (2019, November 9). Personal Interview.

Carroll, M.W. (2015). Sharing Research Data and Intellectual Property Law: A Primer. *PLoS Biology, 13*(8). https://doi.org/10.1371/journal.pbio.1002235

Dahmm, H. (2019, September 12). UN Environment and Google Partnering to Monitor Global Surface Water: A Case Study by SDSN TReNDS for C4DC. Contracts for Data Collaboration. Retrieved from https://contractsfordatacollaboration.org/static/files/un-environment-and-google-case-study_3.pdf

Dance, G. J. X, LaForgia, M., & Confessore, N. (2018, December 18). As Facebook raised a privacy wall, it carved an opening for tech giants. *The New York Times*. Retrieved from https://www.nytimes.com/2018/12/18/technology/facebook-privacy.html

Dahmm, H. Laying the Foundation for Effective Partnerships: An Examination of Data Sharing Agreements. Open Science Framework [pre-print]. June 10, 2020.

Davis, K. (2018, April 6). Facebook users sue over personal data breach in Cambridge Analytica case. *Los Angeles Times.* Retrieved from https://www.latimes.com/local/lanow/la-me-facebook-lawsuit-20180406-story.html

Dwoskin, E. and Siddiqui, F. (2017, January 8). Uber is finally releasing a data trove that officials say will make driving better for everyone. *The Washington Post.* Retrieved from https://www.washingtonpost.com/news/the-switch/wp/2017/01/08/uber/

Espey, J., Badiee, S., Dahmm, H., Appel, D., & Noe, L. (2019). Counting on the World to Act: A Roadmap for Governments to Achieve Modern Data Systems for Sustainable Development.

Gooch, P., Luysterbourg, E., Sponselee, A., Frank, D. P., Dewitt, B., Sehgal, M., & Batch, D. (2018). A New Era for Privacy: General Data Protection Regulation Six Months On. Deloitte. Retrieved from https://www2.deloitte.com/content/dam/Deloitte/uk/Documents/risk/deloitte-uk-risk-gdpr-six-months-on.pdf

GovLab (2016, October 13). *Esri and Waze Open Data-Sharing for Governments.* Retrieved from http://datacollaboratives.org/cases/esri-and-waze-open-data-sharing-for-governments.html

GSMA. (2018, September 21). Airtel case study: Harnessing mobile Big Data to identify tuberculosis hotspots in India. Retrieved from https://www.gsma.com/betterfuture/resources/airtel-case-study-harnessing-mobile-big-data-to-identify-tuberculosis-hotspots-in-india

Hoffman, H., Boral, A., Olukoya, D., Henke, N., Bick, R., Rifai, K., Roth, M., Youldon, T., Jurgens, J., O'Halloran, D., Morhard, R., Samans, R., Toth, A., Scarpino, S., Harrington, m., McGornan, L., Wetter, E., Verhulst, S., Tan, V., Hwang, C., Stonier, J., Deland, E., Shetty, S., Espey, J., David, S., & Berens, J. (2019

April 4). Data Collaboration for the Common Good: Enabling Trust and Innovation Through Public-Private Partnerships. World Economic Forum. Retrieved from http://www3.weforum.org/docs/WEF_Data_Collaboration_for_the_Common_Good.pdf

Hudgins, V. (2019, October 22). US-UK Data Sharing Agreement Might Not Soothe Law Firms' Cloud Concerns. *Legaltech News.* Retrieved from https://www.law.com/legaltechnews/2019/10/22/us-uk-data-sharing-agreement-might-not-soothe-law-firms-cloud-reluctance/?slreturn=20200509100928

Human Rights Watch. (2020, May 13). Mobile Location Data and Covid-19: Q&A. Retrieved from https://www.hrw.org/news/2020/05/13/mobile-location-data-and-covid-19-qa

International Open Data Charter. (2015). Retrieved from https://opendatacharter.net/

MacDonald, G., (2019, October 8). Personal Interview.

MacDonald, G. & Lin, J. (2018, August 7). Urban's Process for Creating Private-Sector Data Use Agreements. Medium. Retrieved from https://medium.com/@urban_institute/urbans-process-for-creating-private-sector-data-use-agreements-e8c9e0376561

MacDonald, L. (2019, October 8). Personal Interview.

Madans, J. (2020, January 2). Personal Interview.

McGorman, L. (2020, January 3). Personal Interview.

Microsoft. (2019). Removing barriers to data innovation. Retrieved from https://news.microsoft.com/datainnovation/

Rodriguez, K. & Fischer, C. (2019, October 4). A Race to the Bottom of Privacy Protection: The US-UK Deal Would Trample Cross Border Privacy Safeguards. Electronic Frontier Foundation. Retrieved from https://www.eff.org/deeplinks/2019/10/race-bottom-privacy-protection-us-uk-deal-would-trample-cross-border-privacy

Schroeder, A. (2020, February 18). Personal Interview.

Sethi, T., & M. Prakash. (2018). Counting on Statistics: How can national statistical offices and donors increase use? Williamsburg, VA: AidData at William & Mary.

Sibande, R. (2020, April 28). Using Mobile Network Operator Data for COVID-19 Response. Digital Impact Alliance. Retrieved from https://digitalimpactalliance.org/using-mobile-network-operator-data-for-covid-19-response/

Timberg, C. & Harwell, D. (2020, January 19). Government efforts to track virus through phone location data complicated by privacy concerns. *Washington Post.* Retrieved from https://www.washingtonpost.com/technology/2020/03/19/privacy-coronavirus-phone-data/

U.S. Department of Health and Human Services. (2018, September). The State of Data Sharing in the U.S. Department of Health and Human Services. Retrieved from https://www.hhs.gov/sites/default/files/HHS_StateofDataSharing_0915.pdf
UNOCHA & Centre for Humanitarian Data. (2019, March), Data Responsibility Guidelines - Working Draft. Retrieved from https://centre.humdata.org/wp-content/uploads/2019/03/OCHA-DR-Guidelines-working-draft-032019.pdf

Verhulst, S G. (2019, February 17). Data Collaboratives: The Emergence of Public Private Partnerships around Data for Social Good. Presented at the AI for Social Good Conference, Doha, Qatar. Retrieved from https://qcai.qcri.org/wp-content/uploads/2019/03/Data_Collaboratives__Stefaan_Verhulst__GovLab.pdf.

Wakabayashi, D. (2019, June 26), Google and the University of Chicago Are Sued Over Data Sharing. *The New York Times.* Retrieved from https://www.nytimes.com/2019/06/26/technology/google-university-chicago-data-sharing-lawsuit.html