MOVING FROM IDEA TO PRACTICE WITH DATA SHARING AGREEMENTS

The Contractual Wheel of Data Collaboration 2.0

Andrew J. Zahuranec, Hannah Chafetz, and Stefaan Verhulst

THE**GOV**LAB

Open.Data.Policy.Lab

This paper is part of a series—"Data Sharing Agreements: Moving from Idea to Practice." This series describes the ways organizations can use data-sharing agreements to facilitate data collaborations that serve both the public and institutional interests. To view the other papers in this series, visit our webpage at dsa.opendatapolicylab.org.

The world is becoming increasingly urbanized. As of 2020, an estimated 56.2% of people lived in urban areas, with the United Nations projecting this number to increase to 68% by 2050. These changes require cities to be at the forefront of modern problem-solving efforts, to develop new ways of addressing social, environmental, political, and economic challenges. However, cities cannot develop these approaches if they do not understand what is happening within their borders. They need data to inform them about local needs.

Unfortunately, as the Open Data Policy Lab's Open Data Action Labs with TrustRelay outlined, the process of securing this data is rarely easy. The most useful and relevant data is rarely held by a singular actor but instead diffused across a variety of public, private, and community-based groups with their own interests and motivations.[1] The gap between city data holders and data users can be bridged by well-drafted data sharing agreements—written agreements that establish the terms for how data is shared between parties—but developing these agreements can be expensive and time-consuming.
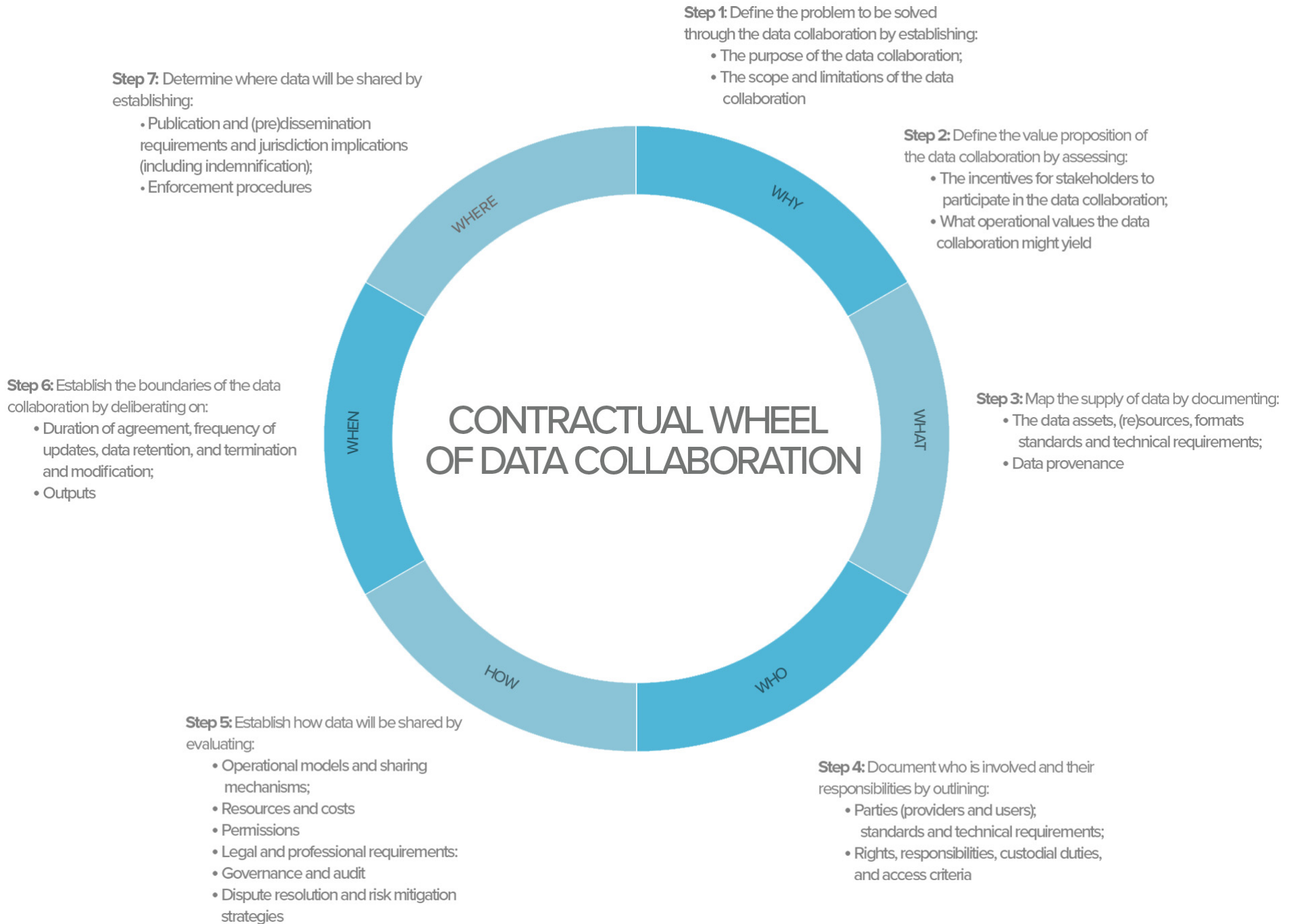
This document seeks to address that problem. After using the readiness matrix to assess the readiness of all partners to collaborate and the principle-negotiation framework to discuss each party's interests, this **Contractual Wheel of Data Collaboration 2.0** provides smart city data collaborative organizers with a step-by-step framework on how to assemble a data sharing agreement that can operationalize meaningful work and answer core questions regarding the collaboration.

---

[1] The Open Data Action Labs (ODAL) is an effort from the Open Data Policy Lab (ODPL) to support organizations by helping to define new strategies and tools that can resolve the policy challenges they face. These solutions are built around addressing gaps and asymmetries that slow data innovation and hinder progress toward realizing the promise of the Third Wave of Open Data. ODALs are built around a series of workshops (called "studios") which give experts and stakeholders an opportunity to define the problems facing them and then ideate possible solutions to those problems in a collaborative setting. In February and March 2023, the ODPL and TrustRelay co-hosted conversations with experts in law, data, and smart cities. This mechanism is based on the findings from these conversations.

As the name suggests, this document is an update of our prior [Contractual Wheel of Data Collaboration](#), first developed in 2019 by The GovLab. This document used discussions with real-world data providers and users and a [library of data-sharing agreement clauses](#) to outline the kinds of provisions that need to appear in agreements. It ultimately provided the readers with [six categories of questions](#). These are:

1. **Why** is data being shared? What is the context and purpose?
2. **What** kinds of data are being shared? What are the sources, formats, and other technical requirements?
3. **Who** is party to the agreement? Who will be providing and using data resources? Are there any other 'third parties' that are also involved? Who has certain rights and duties?
4. **How** is data being shared? How is the relationship managed? How will issues such as security, privacy, and risk be handled?
5. **When** will data actions take place? At what point does the agreement start and end?
6. **Where** is data being shared to and from? Are there jurisdictional issues to consider? Are there any international laws that apply?

Inspired by feedback from a recent studio workshop, The Open Data Policy Lab has expanded on this framework. The **Contractual Wheel 2.0** outlines _how_ those setting up data collaborations can answer each of these questions. It provides specific steps that organizers can follow in order to develop a data sharing agreement that meets the needs of all partners.

Step 1: Define the problem to be solved through the data collaboration by establishing:
- The purpose of the data collaboration;
- The scope and limitations of the data collaboration

Step 2: Define the value proposition of the data collaboration by assessing:
- The incentives for stakeholders to participate in the data collaboration;
- What operational values the data collaboration might yield

Step 3: Map the supply of data by documenting:
- The data assets, (re)sources, formats standards and technical requirements;
- Data provenance

Step 4: Document who is involved and their responsibilities by outlining:
- Parties (providers and users); standards and technical requirements;
- Rights, responsibilities, custodial duties, and access criteria

Step 5: Establish how data will be shared by evaluating:
- Operational models and sharing mechanisms;
- Resources and costs
- Permissions
- Legal and professional requirements:
- Governance and audit
- Dispute resolution and risk mitigation strategies

Step 6: Establish the boundaries of the data collaboration by deliberating on:
- Duration of agreement, frequency of updates, data retention, and termination and modification;
- Outputs

Step 7: Determine where data will be shared by establishing:
- Publication and (pre)dissemination requirements and jurisdiction implications (including indemnification);
- Enforcement procedures

CONTRACTUAL WHEEL OF DATA COLLABORATION

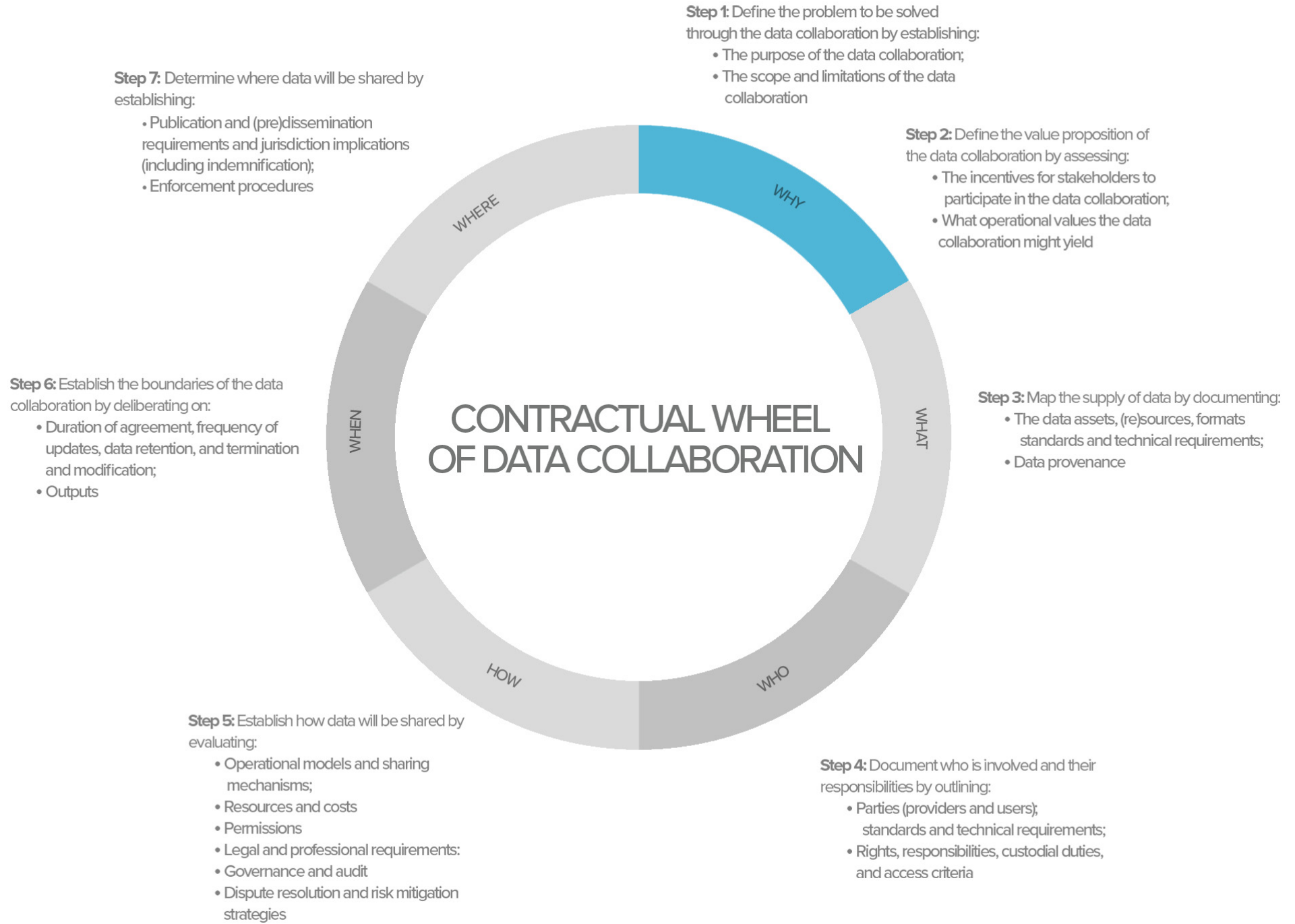WHY

WHAT

WHO

HOW

WHEN

WHERE

While this graphic alone may provide clarity for some, in the following sections, we explain what each category means and what kinds of actions each step entails. This work is done in a table format:

- **Category** lists the type of question that needs to be answered.
- **Steps** outline actions that can be taken to address that category of question.
- **Questions to Answer** provides provides a checklist of questions that partners should answer among themselves to ensure they can appropriately achieve the aforementioned steps.
- **Example Clauses,** finally, provides examples of clauses from real-world data-sharing agreements that organizations might use or be inspired by to develop their own agreements.
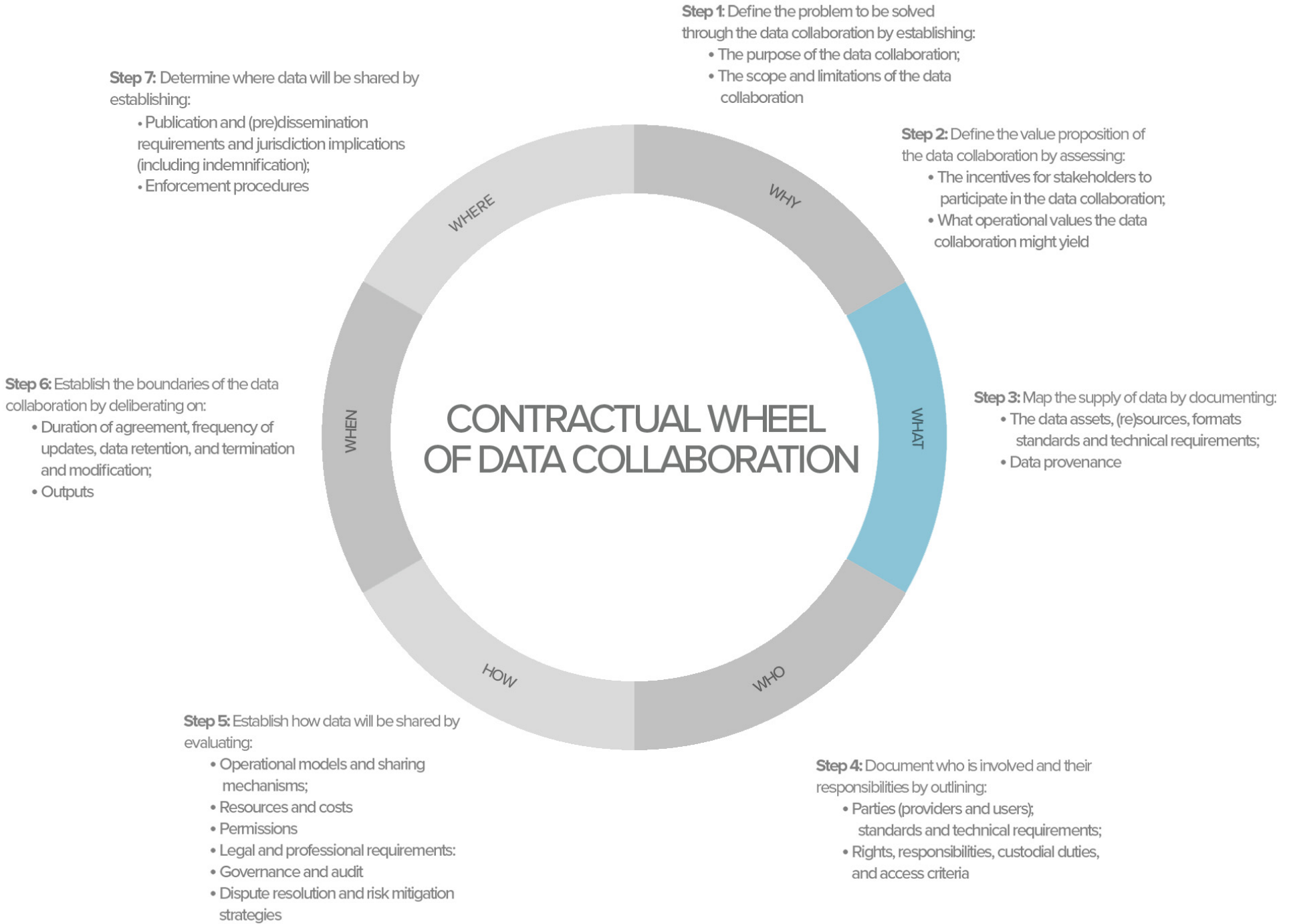
Additionally, at the end of the document, we include a table providing **Useful Resources** that may be helpful in operationalizing the steps outlined. These tables provide organizers with a list of considerations and a sample of the different ways these considerations can be addressed.

While each data-sharing agreement is different and no document can be truly comprehensive, this document aspires to address the most common challenges that organizers come across when trying to actualize the Contractual Wheel of Data Collaboration. We hope to iterate this resource in response to regular feedback and use.
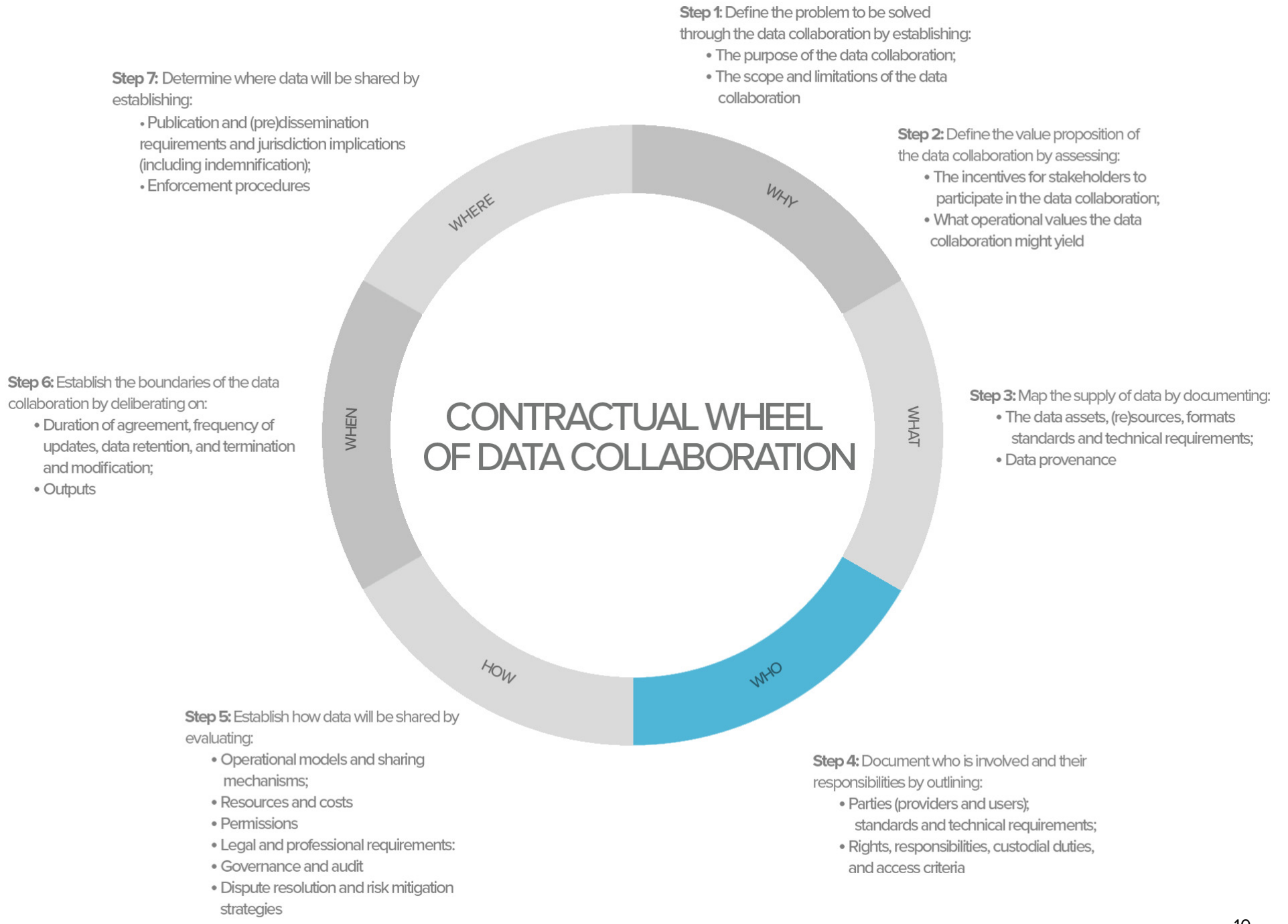
Step 1: Define the problem to be solved through the data collaboration by establishing:
- The purpose of the data collaboration;
- The scope and limitations of the data collaboration

Step 2: Define the value proposition of the data collaboration by assessing:
- The incentives for stakeholders to participate in the data collaboration;
- What operational values the data collaboration might yield

Step 3: Map the supply of data by documenting:
- The data assets, (re)sources, formats standards and technical requirements;
- Data provenance

Step 4: Document who is involved and their responsibilities by outlining:
- Parties (providers and users); standards and technical requirements;
- Rights, responsibilities, custodial duties, and access criteria

Step 5: Establish how data will be shared by evaluating:
- Operational models and sharing mechanisms;
- Resources and costs
- Permissions
- Legal and professional requirements:
- Governance and audit
- Dispute resolution and risk mitigation strategies

Step 6: Establish the boundaries of the data collaboration by deliberating on:
- Duration of agreement, frequency of updates, data retention, and termination and modification;
- Outputs

Step 7: Determine where data will be shared by establishing:
- Publication and (pre)dissemination requirements and jurisdiction implications (including indemnification);
- Enforcement procedures

CONTRACTUAL WHEEL OF DATA COLLABORATION

WHY
WHAT
WHO
HOW
WHEN
WHERE

Why		
Steps	**Questions to Answer**	**Example Clauses**
Step 1: Define the problem to be solved through the data collaboration. This includes establishing: ● **The purpose of the data collaboration:** Evaluating how the collaboration could impact different stakeholders, sectors, or initiatives and establishing a clear problem statement or question to guide the entire process. ● **The scope and limitations of the data collaboration:** Documenting the boundaries of the data collaboration. This work might include defining the type of collaboration, timeline, goals, and required resources.	☐ In what way will resolving this problem meaningfully help a select group or audience? What gap does it fill? ☐ What financial, human, and technological resources does the prospective partnership need to ensure it can make a meaningful difference? Does the data user and supplier have these resources? ☐ What limitations does the partnership face in pursuing this work?. Is there a hard time limit? Are there expectations or rules which you must abide by that will limit your capacity? ☐ Keeping the above in mind, what are 1-3 goals that the partnership considers to be achievable?	**Establishing the Purpose:** ● **Example #1**: "The goal of the affiliation is to advance research and discovery in genomics in order to enable safer and more effective drug therapies." ● **Example #2**: "1. Purpose: The purpose of this Agreement is to disclose school-level data for the purposes of evaluating federally funded programs 7 C.F.R., Part 210 and 7 C.F.R., Part 220 and to identify schools and LEAs that are effectively improving breakfast access and participation. This agreement does not involve the disclosure or use of Personally Identifiable Student Data." ● **Example #3**: "PURPOSE: To facilitate the health of Indian tribes and Individual American Indians and Alaska Natives X and Y are entering into an agreement which will allow the exchange of data and clarification of data access and utilization. Y will provide data collected to X for the purposes of." **Documenting the scope and limitations:** ● **Example #1**: "2.3 Constraints on Use: The Parties recognise 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;" ● **Example #2**: "This MOU becomes effective on the date of the last signature of all the Participants and is expected to continue through the end of the [project agreement] [date of agreement expiration]. However, the Participants may decide, in writing, to extend this period." ● **Example #3**: "Health Data Coalition is funded by GPSC (General Practice Services Committee). HDC will provide endpoint software and hardware to be used within physician offices or within HDC managed data centres…"

| **Step 2:** Define the value proposition of the data collaboration. This involves assessing:

• **The incentives for stakeholders to participate in the data collaboration:** Understanding how the data collaboration could benefit data providers and evaluating what factors could motivate stakeholders participate.

• **Determining what operational values the data collaboration might yield:** Understanding the kinds of insights that a data collaborative can generate, whether that be descriptive (understanding the world); predictive (forecasting future activities); diagnostic (understanding the cause or effect of an action); or prescriptive (identifying what should or should not happen to address future needs). | ☐ Who does the partnership view as the target audience for this work? Who are the beneficiaries? What are they looking to achieve?

☐ Who are all actors involved in the data collaboration and what is their incentive to participate?

☐ What value statement would articulate the way the partnership could bring value to the target audience and the larger problem landscape?

☐ What kind of insight would be transformative to partnership members? How does the partnership foster this kind of insight? | **Aligning on incentives for stakeholders to participate in the data collaboration:**

• **Example #1:** "The Participants intend to focus joint activities around the following objectives: • Close coordination with development partners to ensure maximum convenience and practical utility of the [specific type of] data management tools including ongoing feedback regarding the system effectiveness and ease of use. • Capacity building of [Partner] staff to create and maintain this data set and updating of data management plan to include [specific type of] data production processes."
• **Example #2**: The Parties agree: i) To work together towards identification and exploration of data sharing opportunities that will benefit the care and outcome of their patients; … 2.2 Intended Use: The Parties agree that the data to be provided via the Data Sharing Arrangements is intended primarily for the purposes of advancing integrated healthcare and improved patient care and outcomes.
• **Example #3:** "The educational institutions who choose to participate in this MOU desire to evaluate and improve their respective educational programs through the analysis of academic performance data concerning students who have or who are now attending an educational institution."

Determining what operational values the data collaboration might yield:

• **Example #1:** "The parties are interested in joint efforts and information sharing to determine whether the information that is collected, processed and protected by the Chamber of Commerce of Bogotá, in the exercise of the functions that have been delegated by the Colombian regulations that are in place and in the activities that these develop, can meet the characteristics of open data, namely: - Open by default - Timely and complete - Accessible and usable - Comparable and inter-operable - To improve governance and citizen participation - For inclusive development and innovation AI"
• **Example #2:** ". To this end, HDC will develop, implement, and maintain a centrally hosted, web-based application to enable participating Clinical Data Contributors to collaborate and use aggregated Electronic Medical Record ("EMR") patient data (referred to as "HDC Data." |

Step 1: Define the problem to be solved through the data collaboration by establishing:
- The purpose of the data collaboration;
- The scope and limitations of the data collaboration

Step 2: Define the value proposition of the data collaboration by assessing:
- The incentives for stakeholders to participate in the data collaboration;
- What operational values the data collaboration might yield

Step 3: Map the supply of data by documenting:
- The data assets, (re)sources, formats standards and technical requirements;
- Data provenance

Step 7: Determine where data will be shared by establishing:
- Publication and (pre)dissemination requirements and jurisdiction implications (including indemnification);
- Enforcement procedures

Step 6: Establish the boundaries of the data collaboration by deliberating on:
- Duration of agreement, frequency of updates, data retention, and termination and modification;
- Outputs

CONTRACTUAL WHEEL
OF DATA COLLABORATION

WHERE
WHY
WHAT
WHEN
WHO
HOW

Step 5: Establish how data will be shared by evaluating:
- Operational models and sharing mechanisms;
- Resources and costs
- Permissions
- Legal and professional requirements:
- Governance and audit
- Dispute resolution and risk mitigation strategies

Step 4: Document who is involved and their responsibilities by outlining:
- Parties (providers and users); standards and technical requirements;
- Rights, responsibilities, custodial duties, and access criteria

WHAT		
Steps	**Questions to Answer**	**Example Clauses**
Step 3: Map the supply of data. This requires documenting: ● **The data assets, (re)sources, formats standards and technical requirements:** Understanding what data types, data sources, expertise, quality and standards are required in order to address the problem at hand ● **Data Provenance:** Understanding where the data came from (the origin or lineage) and the decisions that informed its creation.	☐ Have the partners documented what types and sources of data are needed to address the context at hand? ☐ Have the members of the partnership come to a consensus on the minimum requirements needed for the data assets to achieve the objectives of the data collaboration—including the provenance, timeframe, format, and interoperability with different systems? ☐ Have the partners established procedures for keeping track of how decisions are made and who is involved in them? Do these procedures allow partners to understand how certain actions occurred and who is accountable?	**Mapping the data assets and (re)sources and understanding formats and technical requirements:** ● **Example #1**: Data Subjects: [Familiar Name] data Processed for the Purpose may relate to the following categories of data subjects: •Consumer customers and business customers public bodies and corporate customers of the mobile network operators (MNOs) providing data to [Familiar Name] ● **Example #2:** "2. NATURE OF THE DATA TO BE SHARED The data to be shared is address level heat demand data for the specific areas selected by the Local Authority applicant and listed below. The data will be released in the form of X and Y co-ordinates for each address in that area, to enable the data to be mapped using a Geographic Information System (GIS)." ● **Example #3:** "1.3 A list of all the data files and variables required by DECC, and the timetable for data sharing is provided in Annex A." **Data Provenance:** ● **Example #1:** "In order to do this the WDP collects data from customers and businesses, and use this data for a variety of information and service delivery functions including planning and research activities." ● **Example #2**: "Project Management Committee - means the core team responsible for managing the WHO central database and maintenance of the data management system, collating data centrally, carrying out routine monitoring of the data, preparing bi-annual reports as well as coordinating communications with all members;"

Step 1: Define the problem to be solved through the data collaboration by establishing:
- The purpose of the data collaboration;
- The scope and limitations of the data collaboration

Step 2: Define the value proposition of the data collaboration by assessing:
- The incentives for stakeholders to participate in the data collaboration;
- What operational values the data collaboration might yield

Step 3: Map the supply of data by documenting:
- The data assets, (re)sources, formats standards and technical requirements;
- Data provenance

Step 4: Document who is involved and their responsibilities by outlining:
- Parties (providers and users); standards and technical requirements;
- Rights, responsibilities, custodial duties, and access criteria

Step 5: Establish how data will be shared by evaluating:
- Operational models and sharing mechanisms;
- Resources and costs
- Permissions
- Legal and professional requirements:
- Governance and audit
- Dispute resolution and risk mitigation strategies

Step 6: Establish the boundaries of the data collaboration by deliberating on:
- Duration of agreement, frequency of updates, data retention, and termination and modification;
- Outputs

Step 7: Determine where data will be shared by establishing:
- Publication and (pre)dissemination requirements and jurisdiction implications (including indemnification);
- Enforcement procedures

CONTRACTUAL WHEEL OF DATA COLLABORATION

WHY · WHAT · WHO · HOW · WHEN · WHERE

WHO		
Steps	**Questions to Answer**	**Example Clauses**
Step 4: Document who is involved and their responsibilities. This includes documenting: ● **Parties (providers and users):** Bringing together all stakeholders involved in the exchange–including data suppliers, users, and other collaborators ● **Rights, responsibilities, custodial duties, and access criteria:** Understanding what the expectations are for each party in the agreement, aligning on who will be responsible for upkeep and the management of the datasets provided, and finalizing which collaborators will be provided access the data and how they can use it	☐ Has the partnership surveyed the existing legal and regulatory landscape to understand the rights and responsibilities of different parties involved? ☐ Have the partnership organizers drafted a list of all parties who need to be involved in the data collaboration and when, specifically, they can contribute to the data lifecycle? Can the organizer determine if these individuals represent themselves or their institution?. ☐ Has the partnership outlined the obligations that the data collaboration has to those inside and outside the arrangement, including any duties to inform data subjects or relevant public sector authorities about ongoing activities? ☐ Have the partnership's leads engaged with community leaders to understand what kinds of activities are considered appropriate and legitimate to ensure that the collaborative has a social license to operate? ☐ Has the partnership determined if there are any entities who partners should be prohibited from sharing data with, such as law enforcement agencies?	**Parties:** ● **Example #1: "**The [partner1- Funding Partner] (hereinafter referred to as "partner1"), the [partner2- Government] (hereinafter referred to as "partner2"), and the [partner3- Executing Partner] ("herein after referred to as "partner3") have a common interest in the implementation of the [project description]." ● **Example #2:** "1. Parties and Purpose. This document constitutes an agreement between the U.S. Census Bureau and the[Name of other party]." ● **Example #3:** "This Agreement is by and between the [INSERT INSTITUTION NAME (acronym)] and the WHO Central Database for collection of safety data (AE and/or SAE) in the scope of the surveillance of aDSM of anti-TB drugs by World Health Organization, hereafter individually referred to as "Party" and collectively as the "Parties"" **Rights, responsibilities, custodial duties, and access criteria:** ● **Example #1:** "SEVENTH: JURISDICTION. The parties agree that for all purposes of this agreement, they shall by governed by the jurisdiction and competence of the judicial system of the Republic of Colombia with jurisdiction in the City of Bogotá, to the exclusion of any other jurisdiction or jurisdiction that may correspond." ● **Example #2:** "3. 1 Each Party will: 3.1.7 Acquire and use the Material in a manner that respects the scientific value of the Material and the legal rights of Parties and Associated Agencies."

CONTRACTUAL WHEEL OF DATA COLLABORATION

Step 1: Define the problem to be solved through the data collaboration by establishing:
- The purpose of the data collaboration;
- The scope and limitations of the data collaboration

Step 2: Define the value proposition of the data collaboration by assessing:
- The incentives for stakeholders to participate in the data collaboration;
- What operational values the data collaboration might yield

Step 3: Map the supply of data by documenting:
- The data assets, (re)sources, formats standards and technical requirements;
- Data provenance

Step 4: Document who is involved and their responsibilities by outlining:
- Parties (providers and users); standards and technical requirements;
- Rights, responsibilities, custodial duties, and access criteria

Step 5: Establish how data will be shared by evaluating:
- Operational models and sharing mechanisms;
- Resources and costs
- Permissions
- Legal and professional requirements:
- Governance and audit
- Dispute resolution and risk mitigation strategies

Step 6: Establish the boundaries of the data collaboration by deliberating on:
- Duration of agreement, frequency of updates, data retention, and termination and modification;
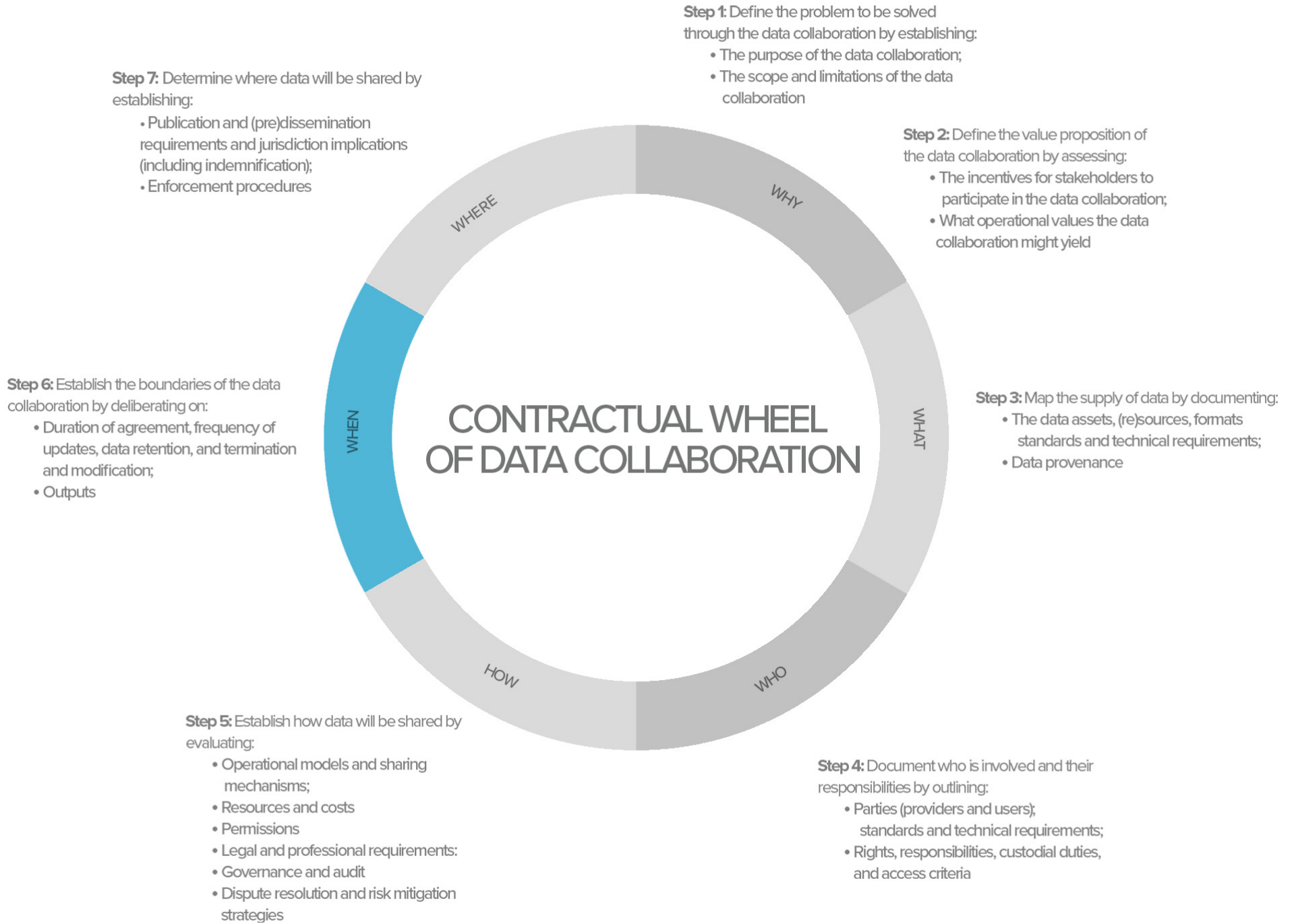- Outputs

Step 7: Determine where data will be shared by establishing:
- Publication and (pre)dissemination requirements and jurisdiction implications (including indemnification);
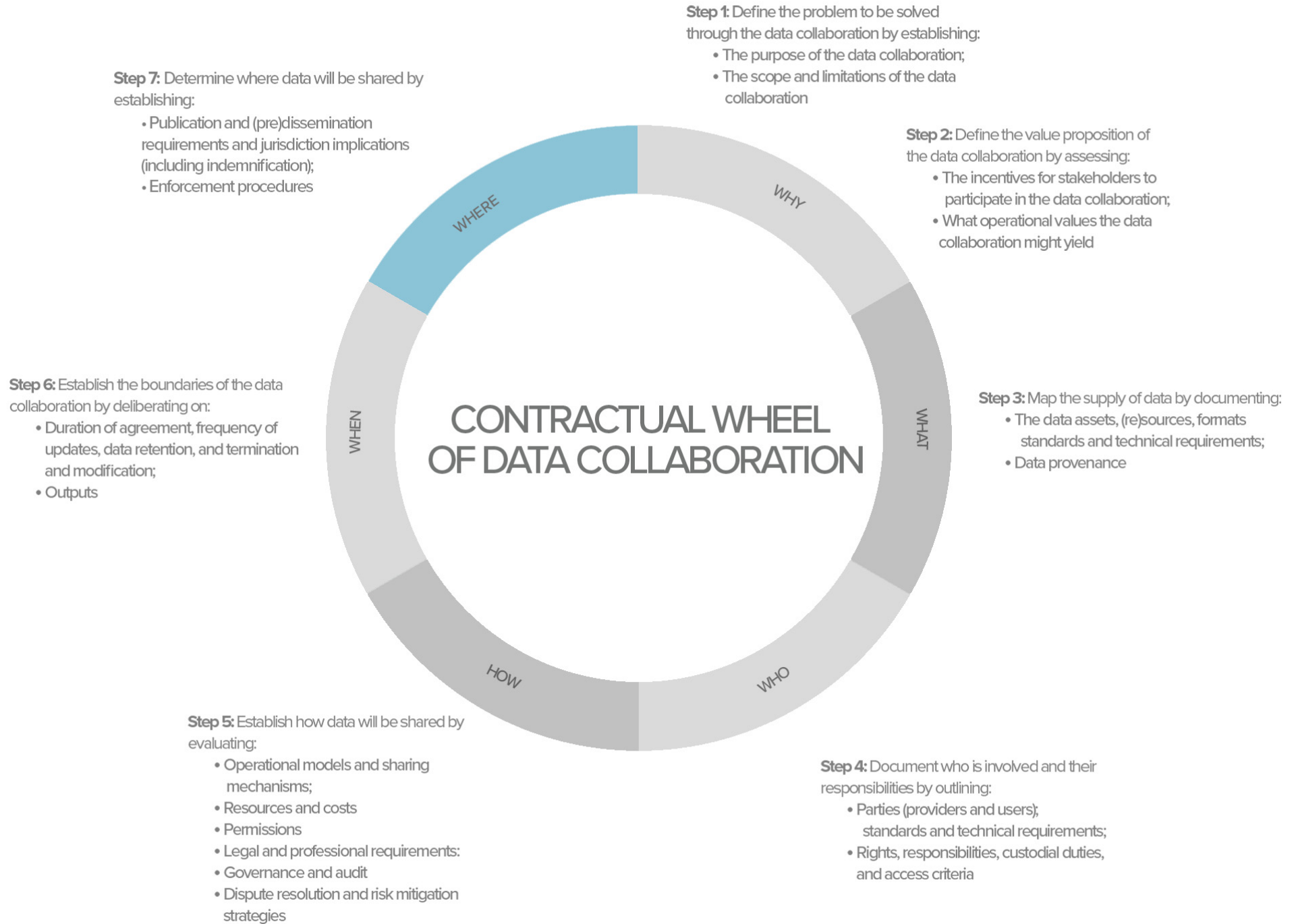- Enforcement procedures

Wheel labels: WHY, WHAT, WHO, HOW, WHEN, WHERE

HOW		
Steps	**Questions to Answer**	**Example Clauses**
Step 5: Establish how data will be shared. This involves evaluating: ● **Operational models and sharing mechanisms:** Defining the mechanics of the collaboration or data exchange ● **Resources and costs:** Establishing the budgetary considerations and provision ● **Permissions:** Assessing how parties can utilize the data provided and where credit needs to be established ● **Legal and professional requirements:** Identifying the actions to be taken based on industry- or domain-specific conditions affecting parties ● **Governance and audit:** Establishing decision-making and accountability processes ● **Dispute resolution and risk mitigation strategies:** Developing a plan of how potential challenges could be mitigated or addressed	☐ Has the partnership identified the data being shared, including the sensitivity of the datasets, the size of the datasets, and the systems needed to interpret it? ☐ Does each partner understand the unique security and legal requirements around sharing their collaborators' data, including whether data has been modified in any way to protect the privacy of subjects, whether data can only be accessed at a specific site or individuals, and whether data is regularly deleted? ☐ Has the partnership assessed the financial, human, and technical costs of using the data, including where resources are available and, if not, where they might be secured? ☐ Has the partnership come to a consensus on who will oversee implementation of the data collaborative (e.g. an intermediary, a governance board or committee set up specifically to manage it) and how disagreements between the parties can be resolved?	**Establishing operational models and sharing mechanisms** ● **Example #1:** "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." ● **Example #2:** "All data generated by this project shall be approved for dissemination by the Institutional Review Board and … Y shall be cited as the source of the data in all tables, reports, presentations, and scientific papers, and X shall be cited as the source of interpretations, calculations, and/or manipulations of the data." **Identifying Resources and Costs:** ● **Example #1:** "6. Sharing of Costs 6.1 The Parties agree that all costs incurred in relation to this MOU shall be the responsibility of each party. Where the parties agree to share costs in relation to activities initiated under this MOU that shall be set out in the relevant Schedule attached to this MOU or via a separate contractual agreement." ● **Example #2**: "The Project is being funded by two organisations Flora Hewlett Foundation ("FHF") and Vodafone Foundation. FHF is not party to this Agreement. … 3.2 Vodafone Foundation 3.2.1 Vodafone Foundation's responsibility under this Agreement is to make payments to Flowminder in accordance with the governance of clause 5 and the cost of funding table detailed in clause 9 of this Agreement. … 9 Cost of Funding 9.1 [Redacted financial details] 9.2 [Redacted financial details] 9.3 Except as otherwise stated in this Agreement, each party shall pay its own costs and expenses in relation to the negotiation, preparation, execution and carrying into effect and performance of this Agreement."

Outlining Permissions:
- **Example #1:** "10.5 The Parties agree that an Associated Agency may become a Party to this Agreement only after a recommendation is provided by the Administrator's Committee in consultation with the Coordinator's Committee."
- **Example #2:** "The Parties agree: iii) To work together to standardise data use consent processes and statements to ensure that identifiable personal data can be shared with the implicit or explicit knowledge and consent of the individual whose data is affected"

Describing legal and professional requirements
- **Example #1:** "PERSONS HAVING ACCESS TO DATA: All persons who will have access to data must complete a data privacy training through < specify >. Prior to the transfer of any data, staff members and researchers who will have access to the data shall sign the Use and Disclosure of Client Information, (signed copies shall be provided to X)."
- **Example #2:** ""Authorized User" means a person: ... (4) Who has completed training in privacy, security, and breach response and who has agreed to be bound by the terms of this TSDEC."

Establishing governance and audit mechanisms:
- **Example #1:** "SECTION 9 - DISPUTE RESOLUTION 9.1 The Parties agree to use all reasonable efforts to resolve a dispute amicably. 9.2 Unresolved disputes between specific Parties at the Coordinators Committee or Technical Committee will first be referred to the Climate Networks Coordinator for further discussion and clarification. If the dispute cannot be resolved, then it will be elevated to a Network Administrator Committee representative of the Parties in dispute for a decision."
- **Example #2:** "Any disputes arising from the terms of this Agreement shall be subject to adjudication in a New York State court of competent jurisdiction. Controversies or questions with respect to this Agreement shall be determined in accordance with the law of the State of New York. This Agreement shall be deemed to be made and to be performed wholly in New York State."

Step 1: Define the problem to be solved through the data collaboration by establishing:
- The purpose of the data collaboration;
- The scope and limitations of the data collaboration

Step 2: Define the value proposition of the data collaboration by assessing:
- The incentives for stakeholders to participate in the data collaboration;
- What operational values the data collaboration might yield

Step 3: Map the supply of data by documenting:
- The data assets, (re)sources, formats standards and technical requirements;
- Data provenance

Step 4: Document who is involved and their responsibilities by outlining:
- Parties (providers and users); standards and technical requirements;
- Rights, responsibilities, custodial duties, and access criteria

Step 5: Establish how data will be shared by evaluating:
- Operational models and sharing mechanisms;
- Resources and costs
- Permissions
- Legal and professional requirements:
- Governance and audit
- Dispute resolution and risk mitigation strategies

Step 6: Establish the boundaries of the data collaboration by deliberating on:
- Duration of agreement, frequency of updates, data retention, and termination and modification;
- Outputs

Step 7: Determine where data will be shared by establishing:
- Publication and (pre)dissemination requirements and jurisdiction implications (including indemnification);
- Enforcement procedures

CONTRACTUAL WHEEL OF DATA COLLABORATION

WHERE

WHY

WHAT

WHO

HOW

WHEN

WHEN		
Steps	**Questions to Answer**	**Example Clauses**
Step 6: Establish the boundaries of the data collaboration. This includes deliberating on: • **Duration (period of agreement), frequency of updates, data retention, and termination and modification:** Deciding the length of the agreement, the anticipated timeline for the data exchanges, provisions for if/how the data will be stored over time and under what conditions it is to be destroyed, and the circumstances under which the collaboration will end and/or when the agreement can be revised; • **Outputs:** Determining whether there is a specific product that should emerge from the effort, either throughout its lifespan or at the end of the arrangement and whether there is a specific audience in mind for it.	☐ Has each partner determined the resources that are available to sustain the initiative and whether there may be new opportunities for funding, manpower, and new systems that may be available outside the collaborative? ☐ Has the partnership collectively identified whether there are specific sensitivities around the data that necessitate a limited time frame for the effort? ☐ Have the partners come to an agreement on a target audience to determine if the insights produced have relevance for only a limited time frame or if the target issue will lose its salience after a certain date or event?	**Understanding the duration, frequency of updates, data retention and termination or modification:** • **Example #1:** "This agreement will be in force for the duration of the memorandum of understanding No. 6200005199 of 2016 and for a term of 12 months from the date of its underwriting." • **Example #2:** "FREQUENCY OF DATA EXCHANGE: Data will be exchanged as needed to meet reporting requirements as well as on an ongoing basis between X and Y staff for the entire length of the project." • **Example #3:** "5. All data transferred to TPP by RIDOH shall remain the property of RIDOH and shall be returned to RIDOH upon termination of the Agreements." **Outputs** • **Example #1:** "3. 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." • **Example #2:** "3 EXTERNAL RELEASE OF EHS DATA BY DECC 3.1DECC will provide DCLG with the detailed fuel poverty tables prior to their publication in May each year by DECC, to ensure that DCLG is satisfied with the content of these. DCLG will not share the tables externally prior to DECC publishing, nor circulate them internally beyond those requiring access on the EHS team. 3.2DCLG will be provided (upon request) with the research version of the fuel poverty datasets for internal analysis and research purposes. This data is not to be circulated for wider use beyond DCLGs analytical team."

Step 1: Define the problem to be solved through the data collaboration by establishing:
- The purpose of the data collaboration;
- The scope and limitations of the data collaboration

Step 7: Determine where data will be shared by establishing:
- Publication and (pre)dissemination requirements and jurisdiction implications (including indemnification);
- Enforcement procedures

Step 2: Define the value proposition of the data collaboration by assessing:
- The incentives for stakeholders to participate in the data collaboration;
- What operational values the data collaboration might yield

Step 6: Establish the boundaries of the data collaboration by deliberating on:
- Duration of agreement, frequency of updates, data retention, and termination and modification;
- Outputs

Step 3: Map the supply of data by documenting:
- The data assets, (re)sources, formats standards and technical requirements;
- Data provenance

CONTRACTUAL WHEEL
OF DATA COLLABORATION

WHERE

WHY

WHEN

WHAT

HOW

WHO

Step 5: Establish how data will be shared by evaluating:
- Operational models and sharing mechanisms;
- Resources and costs
- Permissions
- Legal and professional requirements:
- Governance and audit
- Dispute resolution and risk mitigation strategies

Step 4: Document who is involved and their responsibilities by outlining:
- Parties (providers and users); standards and technical requirements;
- Rights, responsibilities, custodial duties, and access criteria

17

WHERE		
Steps	**Questions to Answer**	**Example Clauses:**
Step 7: Determine where data will be shared. This requires establishing: ● **Publication and (pre)dissemination requirements and jurisdiction implications (including indemnification):** Developing standards that must be met prior to circulating the results of the collaboration and evaluating all factors that need to be considered related to the legal and regulatory context in which the collaboration will occur ● **Enforcement procedures:** Establishing the mechanisms through which parties will be compelled to abide by the provisions of the agreement	☐ Have the partnership's leads surveyed the landscape to determine which audiences would benefit most from the insights produced?. ☐ Have partners examined the legal and regulatory frameworks of the regions in which the data collaborative is active to determine if there are reporting requirements or other obligations that data collaborative participants must abide by?	**Understanding the publication and (pre)dissemination requirements and jurisdiction implications** ● **Example #1:** "If receiving agency transmits or stores confidential information via electronic means, receiving agency will utilize secure file transfer protocol or encryption in motion and at rest and other applicable security controls in accordance with the most current version of the National Institute of Standards in Technology, Special Publication 800-53, or other equally protective security controls." ● **Example #2:** "The Disclosing Party shall make reasonable efforts to ensure that Data provided to the Data Recipient is accurate, up-to-date and relevant. In the event that any information, in excess of information reasonably required in order to allow both organisations to comply with their obligations, is shared, the Data Recipient will notify the other party immediately and arrange the secure return of the information and secure destruction of any copies of that information." **Establishing enforcement procedures** ● **Example #1:** "[Partner 3] will be responsible for the monitoring and evaluation of its activities." ● **Example #2:** "9. Monitoring:The parties shall permit each other the right of a compliance review to ensure that the provisions of the Agreement are upheld and will maintain a system sufficient to allow an audit of its compliance with the disclosure and confidentiality and access to data sections above this Agreement." ● **Example #3:** "I. Flowminder shall make immediately available to [Familiar Name] all information necessary to demonstrate compliance with the obligations laid down under this Agreement and allow for and contribute to any audits, inspections or other verification exercises required by [Familiar Name] from time to time. (Article 28, para 3(h) GDPR)"

ADDITIONAL RESOURCES

STEP	ACTION	USEFUL RESOURCES
Step 1: Define the problem to be solved through the data collaboration.	**Establishing the Purpose**	The GovLab's **Problem Definition Tool** offers guidance, brainstorming exercises, and supplementary resources on problem identification and several other supporting processes (e.g. clarification cycle, system mapping). This tool can help data users critically evaluate the purpose of the data collaboration. It can be used to challenge pre-existing assumptions about the problem space and evaluate the potential impact of the collaboration for different stakeholders. The result will be a clearly defined problem statement which will be carried forward throughout the Data Sharing Agreement process.
	Documenting the scope and limitations	The **Data Value Canvases** provides a framework for data users to build upon their problem statements and evaluate how it could evolve through different elements of the project. The framework highlights enabling and disabling conditions and notes proxies for monetary value. This process can be used to assess what the scope of the collaboration could look like as it evolves.
Step 2: Define the value proposition of the data collaboration.	**Aligning on incentives for stakeholders to participate in the data collaboration**	**A User's Guide to the 9Rs Framework** provides an evidence-based model to further evaluate why an organization might open its data and the benefits it can yield, whether that be in producing Knowledge and Insights, Improving Brand Equity, or Establishing a License to Operate. This framework can help data users communicate the value of the data collaboration in the context of different organizational cultures and deeply held fears of opening up data. The result will be a defined value proposition that can be presented to different stakeholders.
	Determining what operational values the data collaboration might yield	**The Third Wave of Open Data Toolkit** provides an explanation of the different kinds of insights and how they lend themselves to certain kinds of societal and institutional values.
Step 3: Map the supply of data.	**Mapping the data assets and (re)sources and understanding formats and technical requirements**	The **RD4C Data Ecosystem Mapping Tool** can be used to identify and understand the systems generating data to achieve insights around responsible data use, gaps and redundancies in the data, and more. This tool can be used to assess the lineage of the data assets involved in the data collaboration, ensure it adheres to technical standards, and is of sufficient quality. This step is essential to determine whether the data in question is right for the particular context.

	Understanding the data provenance	The **RD4C Decision Provenance Tool** is a methodology that data suppliers or users can use to evaluate the key decision points that contributed to the data in question and document which internal and external parties influenced those decisions. It aims to help identify gaps in decision-making processes and accelerate accountable data practices. Data users and suppliers can leverage the supporting worksheets to ensure the data lineage meets the quality standards.
Step 4: Document who is involved and their responsibilities.	**Establishing rights and responsibilities**	The **RD4C 22 Questions** is an audit tool that aims to help stakeholders involved in the administration of data systems to align their practices with The GovLab and UNICEF's RD4C Principles. This tool can help those looking to set up a data collaboration establish who is responsible for each aspect of the data system and how they can ensure all actions correspond with guiding principles.
Step 5: Establish how data will be shared	**Establishing operational models and sharing mechanisms**	The **Data Collaboratives Canvas** can help deconstruct the data sharing requirements in implementing a data collaborative at an institutional level. The online resource includes examples, enablers, tools, and resources for each phase of setting up a data collaborative. This includes specific resources around operational models.
	Describing legal and professional requirements	The **Data Responsibility Journey** is an assessment tool that outlines the opportunities and risks to consider at each stage of the data lifecycle when implementing a data collaborative. It can be used to assess the type of expertise needed in each part of the data collaboration process and who should be responsible. This process can help data users engage stakeholders across disciplines and build knowledge across different domains.
	Establishing governance and audit mechanisms	The **Third Wave of Open Data Toolkit** explains the complementary processes that facilitate open data and provides a series of "primers" explaining how each element of the Third Wave of Open Data manifests. This toolkit can be used to assess the competencies required for data re-use. It provides several actions and resources that can help establish new governance frameworks and infrastructure for the data collaboration.
	Implementing dispute resolution and risk mitigation strategies	The **RD4C Opportunity and Risk Diagnostic** helps evaluate the implementation of governance frameworks and ensure underlying principles are integrated in the design of all facets of the agreement. The tool provides organizations with a way to take stock of The GovLab and UNICEF's RD4C principles and how they might be realized as an organization reviews a data project or system. This tool can help users evaluate the ethical implications of specific data collaboration decisions and ensure all practices protect the rights of data subjects
Step 6: Establish the boundaries of the data collaboration.	**Understanding the duration, frequency of updates, data retention and termination or modification**	The **Contracts for Data Collaboration Library** is a centralized online resource that collects clauses from numerous Data Sharing Agreements. Under "List of Example Agreements", the reader can view the complete list of agreements represented in the Library. This Library can help make sense of how the duration, frequency of updates, data retention, and other clauses were handled in previous data collaborations. These clauses can be repurposed for different aspects of your Data Sharing Agreement.

Step 7: Determine where data will be shared.	**Understanding the publication and (pre)dissemination requirements and jurisdiction implications**	The **State of Open Data Policy Repository** includes recent policy developments surrounding open data, data reuse, and data collaboration around the world. It looks specifically at how these policies can better enable a third wave of open data. This repository can be used to evaluate the latest jurisdictional requirements for data collaborations and ensure the dissemination requirements adhere to those standards.
	Establishing enforcement procedures	**The State of Open Data Policy Repository** compiles some recent policies surrounding open data, data reuse, and data collaboration around the world that collaborative organizers may wish to keep in mind.