

# SHARING DATA *for* SOCIAL IMPACT: GUIDEBOOK *to* ESTABLISHING RESPONSIBLE GOVERNANCE PRACTICES

*A guidebook produced by  
Natalie Evans Harris,  
program fellow with the*

**beekcenter**  
social impact + innovation

## ABOUT THE BEECK CENTER

[The Beeck Center](#) The Beeck Center works alongside experts and students to surface, accelerate, and scale promising social impact efforts that drive institutional-level change. Positioned as a neutral, academic-based player, the Beeck Center has built credibility working across sectors to generate new ideas that will provoke leaders to think and act differently.

## ABOUT DIGITAL SERVICES COLLABORATIVE

This project is part of the [Digital Service Collaborative](#), a partnership between the Beeck Center and The Rockefeller Foundation that leverages the network of professionals working on data and digital services to scale solutions for greater impact. The Digital Service Collaborative brings together members of the data and digital service community to work together solving specific problems and sharing those solutions throughout the network.

## ABOUT THE AUTHOR

[Natalie Evans Harris](#) has spent nearly twenty years advancing the public sector's strategic use of data, including a sixteen year career at the National Security Agency, and eighteen months with the Obama Administration. She co-founded and currently serves as Head of Strategic Initiatives of [BrightHive](#), a data trust platform to help organizations, networks, and communities securely and responsibly link their data to enhance their impact, empower individual and collective decision making, and increase equity of opportunity. She founded the [Community-driven Principles for Ethical Data Sharing \(CPEDS\)](#) community of practice focused on strengthening ethical practices in the data science community through crowd-sourcing a Data Science Code of Ethics. As a Senior Policy Advisor to the US Chief Technology Officer in the Obama Administration, Natalie founded the [The Data Cabinet](#) - a federal data science community of practice with over 200 active members across more than 40 federal agencies.

## ACKNOWLEDGEMENTS

I would like to thank [The Rockefeller Foundation](#) for providing the financial support that made this work possible. I would also like to thank the Beeck Center team for supporting me through this year plus journey. My deepest appreciation goes to the Student Analysts who conducted research and contributed to the creation of this Guidebook: [Nitya Biyani](#), [Moonlan Zhang](#), [Robert Roussel](#), and [Ben Lang](#). For the experts that contributed time for interviews, emails, phone calls, and reviewing the many iterations of this report. In particular, I am forever in-debt to: [Bill Yock](#), [Sara Holoubek](#), [Kathryn Pettit](#), [Bill Hoffman](#), [Reg Leichty](#), [Leah Hendey](#), [Michelle Shevin](#), [David Park](#), [Dan Wu](#), [Kristin Wolff](#), and [Kevin O'Neil](#). Finally, I must thank my colleagues at [BrightHive](#) for showing me everyday that the practices recommended in this guide are not only doable but impactful.



*This work is licensed under a [Creative Commons Attribution-ShareAlike 4.0 International License](#).*

# THE GUIDEBOOK: HOW TO USE IT

## Who is this guidebook for?

This guidebook is for policy makers, whether elected or appointed, and policy staff looking to understand how to leverage data and data sharing towards evidence-based policy making. It can be used for policy makers and organizations interested in giving agency to individuals to allow for full consent to the use of their data, along with organizations interested in ethically and responsibly sharing data. This guidebook will help actors drive impact with stakeholders.

### ACTORS

Those who take action on the data

Those who use data to drive impact

Examples: policy makers; federal, state and local staff; school districts

### STAKEHOLDERS

Those with a vested interest in data sharing

Those from whom data is being collected

Examples: parents; community members; program managers; IT staff

## What will you find in this guidebook?

This guidebook is essentially a resourced framework. It enumerates specific phases, steps, considerations, and outcomes that creators of data-sharing partnerships for social impact should consider. It includes resources throughout and additional resources for a deeper dive in the [annotated appendix](#).

## Why was this guidebook created?

The adoption of data-sharing governance practices has traditionally been limited and under-resourced, with the sustainability and effectiveness of these practices often harmed by leadership, capacity, and strategies that are not forward-thinking. To match a growing demand for innovative and data-driven approaches to achieving social impact, we need to reform data-sharing governance practices. Many resources exist to help stakeholders establish processes for data sharing. These resources often demand capacity that stakeholders lack, or limit stakeholders' flexibility. As a result, projects can be overly ambitious or lack specificity or impact.

## What Is this guidebook NOT useful for?

Written resources alone may not be enough to generate buy-in from leadership, data teams, partners, or funders. Creating thoughtful and forward-thinking data-sharing partnerships for social impact still has significant barriers that only motivated leaders and good timing can overcome. Getting leaders to invest significant resources at the start of a project may ensure responsible and sustainable operations, but this can sometimes come at the expense of quick wins, having results to show early on. Stakeholders who appreciate new approaches to creating social impact is a critical precursor to impactful data-sharing innovations.

## What are some quick tips on how to use this guidebook?

Start by reading the summary and the table of contents for an overview. If you want to create a collective from the ground up, this overview will help you identify your biggest opportunities and your most significant barriers, as well as think through who you should be working with inside and outside of your organization. If you are part of a collective—a group of three or more organizations with a shared goal for driving impact through linked data—or designing one and want to make sure you are on an effective and sustainable track to generating social impact, you might want to start from the beginning of this guidebook. Consider consulting the resources listed in the appendix, rethinking your stakeholder mix, reconfiguring your team, and/or redesigning your approach based on the insights generated by this guidebook.

# TABLE OF CONTENTS

- The Guidebook: How To Use It** ..... 1
- Executive Summary** ..... 2
- Data Sharing Today** ..... 5
- A New Model for Data-Sharing Governance.** ..... 6
- Phase One: Build the Collective** ..... 7
  - 1.1 Define the Collective .....7
  - 1.2 Establish Shared Vision. ....8
  - 1.3 Understand Data-Sharing Capacity .....10
  - 1.4 Understand the Data Being Shared. ....11
  - 1.5 Survey Security and Privacy Related to the Data. ....12
  - 1.6 Examine the Ethical Implications of Sharing the Data. ....13
  - 1.7 Draft Your Charter .....14
  - 1.8 To Be Done by the End of Phase One .....15
- Phase Two: Define the Operations.** ..... 16
  - 2.1 Determine Governance Framework Structure .....16
  - 2.2 Formalize Responsible Data-Sharing Practices .....19
  - 2.3 Examine the Sustainability of the Operations .....20
  - 2.4 Develop an Ethical Framework for Data Sharing .....20
  - 2.5 Define What Success Looks Like .....22
  - 2.6 Draft Data-Sharing Agreement .....23
  - 2.7 To Be Done by the End of Phase Two .....24
- Phase Three: Drive Impact** ..... 25
  - 3.1 Reinforce, Update, and Share Governance Best Practices .....25
  - 3.2 Reinforce, Update, and Share Best Practices in Ethics Governance. ...26
  - 3.3 Monitor and Assess Privacy and Security Approaches .....26
  - 3.4 Develop Processes to Improve Data Quality Over Time. ....27
  - 3.5 To Be Done by the End of Phase Three. ....27
- Conclusion: Community of Practice** ..... 28

# DATA SHARING TODAY

A growing number of [companies](#) and [government agencies](#) are [forming data partnerships](#) and linking data across institutional and geographic boundaries in support of the social good. They [may be doing](#) so in pursuit of philanthropy or corporate social responsibility, to unlock shared value, or to demonstrate goodwill to their customers, government, or the business community. [Surveys find](#) that consumers are willing to share their data or have their data shared when it is done in pursuit of some common good and when some [norms](#) are respected (privacy, anonymization, no resale, etc.) but are also worried that they have [lost control](#) over what data they share with private companies. In order to unlock the potential of data sharing to help solve social problems and to overcome significant [barriers to sharing](#) you need to rely on robust frameworks for governance that address major points of risk and ambiguity that prevent many actors from engaging in this kind of sharing.

This framework aims to be a resource for organizations interested in sharing data to leverage best practices in legal, technical, and cultural approaches to establishing models for positive social impact. The data-sharing governance landscape for social impact is already populated with a [wide range of terms](#). Let us start by defining some key terms:

- **Data-sharing governance:** A set of processes that manage how data are shared and ensure that data are protected, administered, used, and shared responsibly
- **Governance structure:** The people, policies, procedures, and technologies required to manage the operations of multi-party data sharing under the direction of a managing body
- **Governing board:** A group of key stakeholders who oversee the operations of a data partnership; in accordance with a charter and/or contractual agreement, they supervise how data sharing accomplishes goals and evaluate progress in the pursuit of shared vision
- **Collective:** A group of organizations with a shared goal for driving impact through linked data
- **Minimum viable coalition:** A group of stakeholders united by a charter that outlines a common set of values and motivations, sets specific goals for the partnership, and considers how various barriers to sustainability can be overcome
- **Partnership:** Two organizations with a shared goal for driving impact through linked data
- **Charter:** A document which outlines a common set of values and motivations, sets specific goals for the partnership, and considers how various barriers to sustainability can be overcome

There are plenty of exemplars of government agencies using data to drive impact, from the [Colorado Department of Education](#) creating a data governance board to the [North Carolina Early Childhood Foundation](#) driving impact through data collection and action. Despite these exciting success stories, there are still gaps between governance practices and the idealized version of data sharing. Some of the most pressing needs for reform of data-sharing governance come from the fact that it is currently:

- Short-sighted, addressing immediate needs using limited bilateral agreements without a plan for sustainability, whereas it needs to be sustainable, addressing immediate needs while supporting the potential for long-term growth.
- Isolated, often sustained by a single champion but which often fails due to changes in leadership or lack of stakeholder buy-in, when it needs to be coordinated, supporting shared decision-making for trust-managed data resources and organizational data ownership.
- Imposed on organizations that lack the capacity to participate on equal footing, when it needs to empower them by ensuring that all data-sharing members, regardless of size, capacity, or incentives to participate.
- Self-interested, where lawyers protect their organization's interests and limit liabilities at the expense of value to stakeholders, when it needs to be group-oriented, with liability balanced with creating value for communities and individuals who should benefit most from the data sharing.

## A NEW MODEL *for* DATA-SHARING GOVERNANCE

This framework defines a model to support emerging and established data-sharing collaboratives through guidance on the legal, technical, and cultural approaches. It aims to provide a holistic guide that encourages parties to consider technical, legal, as well as ethical components of data sharing. It is divided into three main phases.

### Phase 1: Build the Collective

- Understand gap filled by data sharing
- Understand data being shared
- Understand ethical implications

### Phase 2: Defining the Operations

- Define success
- Establish internal trust within the collective
- Define operations of the collective

### Phase 3: Driving Impact

- Formalize ethical principles and guidelines
- Establish feedback loop for evaluating impact metrics
- Create internal monitoring and assessment

# PHASE ONE: BUILD *the* COLLECTIVE

The first phase is all about getting everyone on board. You need to clearly identify the problem or policy issue data sharing can address, discuss which stakeholders should be involved, identify the barriers to sharing, and establish a theory of change that explains how impact might be achieved. Bringing stakeholders together can help leadership understand overlap and divergence in motivations, capacity, values, and goals, and ensure that all parties are operating under the same expectations. Through consultation, dialogue, and understanding, stakeholders can rally around a shared vision that can be enshrined in a charter that outlines a common set of values and motivations, sets specific goals for the partnership, and considers how various barriers to sustainability can be overcome.

## **Collective**

A group of organizations with a shared goal for driving impact through linked data

## **Partnership**

Two organizations with a shared goal for driving impact through linked data

OR

## **Minimum Viable Coalition**

A group of stakeholders united by a charter that outlines a set of values and motivations, sets specific goals, and considers how various barriers to sustainability can be overcome

OR

## 1.1 Define the Collective

### *What problem are you solving?*

Finding problems is often harder than finding solutions. You cannot expect to have an impact by simply giving an interesting dataset to a data scientist: the policy question should inform your choice of what data might be appropriate and how you should form the collective. You must understand who your stakeholders are and how to best engage with them to solve your problem. This means understanding if you will be diagnosing data of yesterday, today, or tomorrow. Each situation warrants a different approach for yourself and stakeholders.

Ongoing dialogue between stakeholders is key to matching a problem with capacity to reach a solution that delivers impact. Successful projects in this area often succeed thanks to leadership that values data sharing and leverages existing networks to get others on board. Matching a policy problem to a data source usually comes directly from practitioners, although [some platforms](#) and [resources](#) are emerging that attempt this matching

### *What impact are you pursuing and what is your theory of change?*

Identifying the social value proposition and outlining a [theory of change](#) are key steps for any project. Having a clear idea of how you can achieve impact informs what data and metrics you might need, who your stakeholders might be, and how to approach the problem holistically to achieve [collective impact](#). Collective impact is about monitoring the interaction of interventions and initiatives over time and evaluating them in a dynamic and adaptive fashion that focuses on different elements at different stages. Funders and organizations in the social space are increasingly seeking to fund projects that have a demonstrated impact, but doing so [must be done thoughtfully](#) to [avoid wasting resources](#).

### *Who are the key stakeholders?*

Based on your understanding of the problem and the data that could help achieve impact, identify a set of organizational and individual stakeholders. You may be an expert in a subject but you cannot be an expert on your users. Failing to involve the [end-users' perspectives](#) in the development, implementation, and review of the project runs the risk of creating a solution that does not serve their real needs and fails to deliver impact. As an [example](#), a city could have a million dollar investment in educational improvements and data sharing, but without an appropriate system to get buy-in and communication from stakeholders will fail to have the support to actually drive the impact.

### *The stakeholder landscape includes:*

- Leadership or data stewards in the data-providing organization
- The demand-side organization
- Third-party data providers with complementary data
- Third-party intermediaries who facilitate sharing and data use
- Funders
- Government agencies
- User advocates
- Users themselves

## 1.2 Establish Shared Vision

### *What are the motivations and goals of data-sharing stakeholders?*

You should design your approach based on a clearly articulated narrative that flows from stakeholders' motivations (organizational values and mission that explain why they want to share the data) and goals (concrete outcomes that seek a solution to the problem). The goals and motivations of stakeholders including their involvement, privacy rules, and desired impact must be kept separate from the implementation plan. This is because the implementation plan must be flexible to make compliance easier for all parties and follow best practices. On the other hand, the stakeholders will be more or less constant and should therefore be separate from the ever-evolving implementation plan. It is important to keep in mind that whatever data you are using is to benefit your stakeholders. [According to GovLab](#), there are five primary motivations for sharing data:

1. Situational awareness and response
2. Public service design and delivery
3. Knowledge creation and transfer
4. Prediction and forecasting
5. Impact assessment and evaluation

Each of the five motivations presents different barriers and opportunities and while there is no one-size-fits-all, it is essential for different stakeholders to be clear about their goals and expectations. While finding an overlap in motivations is likely when building a collective, finding shared goals might be more difficult to identify. The disconnect might be a matter of scope (one actor's goals may be more ambitious than another's) or disagreements over how to conceptualize the theory of change, but agreeing on goals will be essential to define the desired impact that, together with an articulation of motivations, make up your shared vision.



### *What are the barriers to sharing data and can they be overcome?*

Entering into a data-sharing agreement requires a commitment of resources and incurs an opportunity cost. The benefits of entering into the agreement are clearly articulated and framed in terms of the organization's values and motivations. A consultation with stakeholder leaders should aim to tease out their main apprehensions about sharing data and identify barriers that exist. Sharing past examples of successes and failures could help identify ways that different barriers might be addressed. In their January 2019 article "Data Driven Social Partnerships: Exploring an Emergent Trend in Search for of Research Challenges and Questions," Sussha, Grönlund, and Van Tulder offered a comprehensive [meta-review of data sharing](#) in social partnerships. They found that the most commonly cited barriers to sharing are privacy issues, conflicting or lack of legal provisions, difficulty in accessing or discovering data, lack of insight into incentives, difficulty getting data providers to participate, and resource constraints. For the purposes of this guidebook, we will highlight four primary types of barriers, with examples:

#### *Legal/Regulatory*

**Examples:** legislation out of date or inconsistent; organizational data-sharing policies ambiguous; standards not set or inconsistent; data ownership & copyright; privacy concerns

#### *Organizational/Motivational*

**Examples:** poorly designed or aligned incentive structure; value unclear; lack of coordination of roles & resources; difficulty in collaboration; attracting data providers

#### *Social/Ethical*

**Examples:** difficulty measuring impact or value; customer data ownership debate; digital divide/digital invisibility; place within the political debate over sharing; equity concerns; difficulty de-biasing data; public perception; translating data insights into effective interventions

#### *Technical/Data-Related*

**Examples:** privacy issues; security issues; ethical or bias concerns; data quality issues; data poorly matches solution; lack of technical capacity; lack of data standards or metadata; lack of control over data; language or terminology barriers

Beyond these examples, you can look to case studies covering [legal](#), [organizational](#), [social](#), and [technical](#) barriers that all show the complications that can arise through different barrier types. Even more resources on each type of barrier is available in the resources appendix.

## 1.3 Understand Data-Sharing Capacity

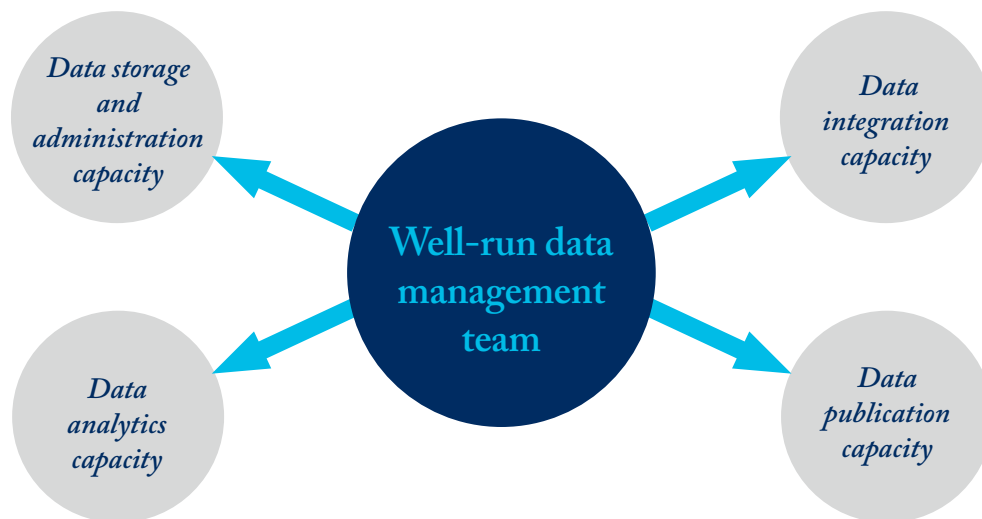
From a [technical perspective](#), it is important to consider the kind of data with which the collective will be working. Many organizations are only using their data and their data teams to answer organizational questions and thus never have invested in capacity to properly share data. This often results in a poor technical infrastructure, lack of expertise, insufficient will or resources, or poor metadata that makes tracking data difficult. Thus, before the operations of the collective move forward, all parties should understand their data assets, as well as their capacity and limitations to effectively sharing data. Through cross-stakeholder consultations, data technicians can understand and agree upon an approach to share specific data for a delimited purpose, whose extent is the minimum needed, and whose level of aggregation, encryption, and anonymization is appropriate for the type of data and reason for sharing. Data sharing may not be all that useful unless it is paired with “interpretive resources” like methods, models, and inside information to improve its usefulness. In this sense, data sharing is not just about the data: in effective partnerships, you share interpretations. Effective sharing relies not just on technical but analytical capacity. Data management capacity is essential, but you need to have the talent in your organization, or train people within it, to draw actionable insights from the data being shared.

### *Have you taken stock of your data-sharing capacity?*

- Who is the point person or team within your organization to teach you about your data and technical infrastructure? It is likely a data management team, a group that establishes and ensures strong data governance practices.
- If nobody has that role, does someone fill this role informally? Depending on existing capacity and your needs, [there are many ways](#) to build a data management team that range from project-specific and function-specific to systems-specific. Once you have identified a relevant group of people, you should be able to understand how your capacity does or does not satisfy your specific project needs.

### *Have you identified technical barriers to data sharing?*

A well-run [data management team](#) will have elements of data storage and administration capacity, data integration capacity, data analytics capacity, and data publication capacity. Approach your data management team, take stock of infrastructure and human capital in each of these four areas, communicate the policy problem and data-sharing approach to members, and ask them to identify needs, barriers, opportunities, and resources as they relate to fulfilling this particular project. Note how the requirements differ from those identified by the leadership, and convey these differences to leaders.

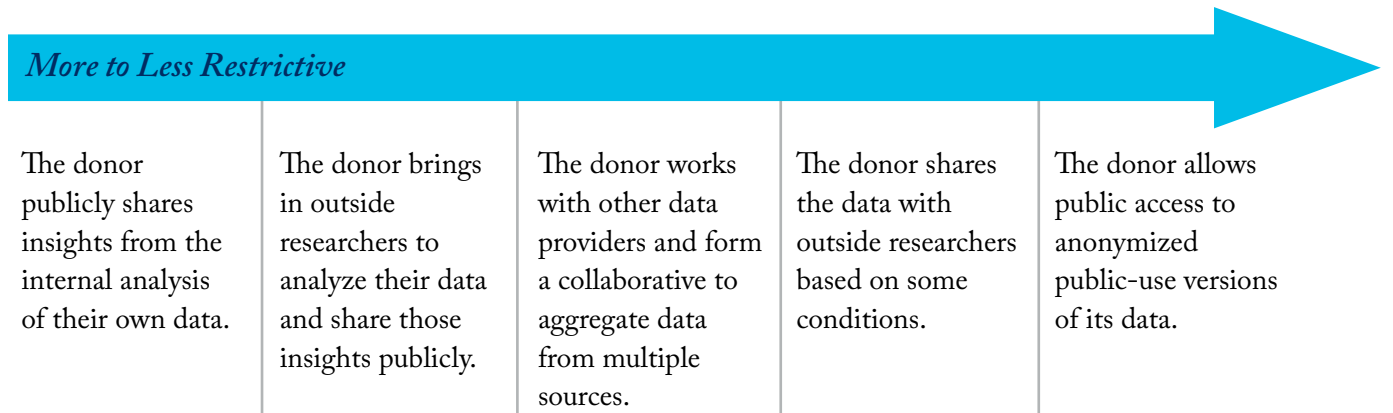


*Have data literacy programs/processes been created?*

If there is an identified lack of [data literacy](#) among key actors in the collective, you can help those actors gain a working understanding of data formats and infrastructures. Some may be tempted to dismiss the need to learn about the technical aspects of data management, but it will help bridge the gap between leaders and data managers, ensure that they are speaking a common language, and help create buy-in. While organizations could rely on the expertise of another member of the collective, they could also encourage their own members to gain data literacy by leveraging [many](#) of the online [resources](#) and [academic programs](#) available.

*Who will receive the shared data? In what form will the data be shared?*

There are multiple forms that sharing can take, from sharing access and insights to sharing the data itself. A particularly useful framework that can help navigate and address different barriers is one put out by the [Urban Institute](#), which offers a continuum from more to less restrictive access:



## 1.4 Understand the Data Being Shared

*What is the quality of the data?*

There is no one gauge of quality of data. Issues with respect to accuracy, completeness, availability, documentation, standardization, or bias all affect data quality according to what the data are being used for, and they affect the validity of its use in different ways. A well-run management team will operate according to a [data management plan](#) as it defines, manages, and tracks an organization’s data. If there is such a team, ask for an assessment of data quality with respect to the basic elements listed above. Without a well-resourced team, data may become siloed, quality of the data will suffer, and issues with the data may remain hidden. If there is neither a data management team nor a point person who can offer insight on quality, then outside teams can work with point-people, leaders, and lawyers of the data-owning organization and get a sample of the data. But internal teams will always be best positioned to provide insights about issues related to quality due to their familiarity with programs, systems, and culture and their proximity to relevant colleagues.

### *Have the appropriate data standards been identified?*

In an effort to extract and convey insights from data while protecting individual-level privacy, nonprofits have published massive amounts of aggregated indicators, measures, and averages. While these coarse datasets are often easily digestible by non-technical audiences and may be fine for answering internal organizational questions, data scientists will need to understand the following to evaluate the data quality:

- Are the data in a standardized format, [interoperable](#) and easy to merge?
- Are there multiple sources that can be joined to validate the data?
- How well have the data been documented (provenance, processing history, metadata, etc.)?
- What data masking/anonymization standards are followed to protect individual identity but still provide analysis on personal demographics and events?
- Are these data aggregated? If so, why was that done? Is it reversible?

## 1.5 Survey Data-Related Security and Privacy

Ensuring that privacy and security concerns and possible solutions are identified early will ease uncertainty that can prevent stakeholders from taking part due to lack of buy-in from leadership, legal, or data teams. Considering the reputational, legal, and technical characteristics of the data, ask your teams the following questions:

### *How sensitive are the data?*

Certain kinds of data warrant extra protection because of their value to nefarious actors (e.g., Social Security numbers, family information, political affiliation). In particular, a strict set of federal and state laws govern the use of data that [are protected by certain laws](#) (e.g., [medical data](#), [children's data](#), education data, financial data). Even if the data in question do not fit into one of these categories, be careful to treat seemingly “non-sensitive” data with extreme care. There is a growing chorus of privacy practitioners who are warning of the [mosaic effect](#), which occurs when of seemingly non-sensitive data is recombined to allow for a sensitive set of data to emerge.

### *How is encryption being managed through the collection, transit, and access of data?*

You [must distinguish](#) between encryption of the data as they exist on the servers of either the data-holders or data-receiver (at rest) and when data are being transferred (in transit). Depending on your infrastructure, your data might be encrypted, depending on whether they are transferred over the internet or a local network. Files that are sent encrypted through a method known as [Secure File Transfer Protocol](#) (SFTP), also arrive encrypted but the sender might not have the skills required to encrypt the data or to effectively coordinate their decryption with teams at receiving organizations. Take careful stock of risk and capacity when deciding how to approach encryption. When sharing occurred in the case studies (available in the [annotated appendix](#)) we reviewed, data were sent to a central repository using SFTP, or brought to the repository's servers on a physical medium such as a CD or hard drive, for upload. The data were then immediately encrypted and protected with a password given only to those needed to process the data and make them ready for analysis.

### *What are the likely feelings of the relevant communities towards privacy?*

Even if you are using state-of-the-art security protocols that protect data from breaches, you should be sure that the subjects of the data you are using would not be opposed to your use of their information. It is imperative to maximize the agency they have over their data. As we've seen with cases such as [InBloom](#), public

perspective is likely to pose a serious risk to the longevity of the collective, it should be considered early. Even if a backlash poses no legal threat, it could harm the reputation of and the trusting relationship between stakeholders. Surveying relevant populations to gauge their openness to the project can help minimize the risk of a backlash, generate positive marketing and communications, and get more leadership buy-in both within the organization and at similar organizations in support of scaling up the project.

### *What forms of anonymization should be employed to protect personal data?*

Figuring out the appropriate degree of anonymization is tricky because there is no telling what the capacity to re-identify the data will be in the future. If you play it too safe and anonymize to a significant degree then the data may no longer be useful. Considerations of the sensitivity of the data and the risk of re-identification can be addressed through addendums and memorandums of understanding attached to the data-sharing agreement.

### *What might the recovery process look like?*

Find out where the data are being stored and if there is a backup strategy. In this area, redundancy can be considered an asset rather than a sign of poor data management. In some cases, data-sharing partnerships never permanently modify a dataset, even if they are sure that they are modifying it in ways that improve their accuracy.<sup>1</sup> This can act as an important safeguard that can help recover from data loss, accidental manipulation, and loss of documentation surrounding processing.

## 1.6 Examine the Ethical Implications of Sharing the Data

Phase One is also where you should begin considering the ethical implications of data sharing. Unintended consequences occur frequently because people often use big data regardless of its quality, legality, and/or understanding of the original purposes for collection. Some points to consider at this stage include: who is the target of the data-driven project, who might be negatively affected through the sharing or the project, and how can these affected communities be engaged at an early stage? Possible steps to take here include creating educational initiatives to get practitioners thinking about the ethical implications of data use, frameworks that help guide ethical use of data, and ethical review boards and community advisory boards. Include user advocates and user perspectives in this process to help guide the collective in navigating ethical pitfalls surrounding data sharing. It is important to keep in mind the rights of the users at hand. [The Oxfam Responsible Data Program Policy](#) provides a perfect list of rights to do so.

### *What is the impacted population? What ethical considerations does the nature of the population or its data pose?*

Despite its potential to enable inclusive growth, data used in the social sector often reflect bias or discrimination against vulnerable populations and can therefore, even unintentionally, [result in worsening inequities](#).

---

<sup>1</sup> For example, South Carolina's multi-agency Integrated Data System (IDS) kept the original dataset and performed analysis on a separate "cleaned" one to play it safe. See Erika M. Kitzmiller, *IDS CASE STUDY: South Carolina* (Philadelphia, PA: Actionable Intelligence for Social Policy, March 2014), [https://www.aisp.upenn.edu/wp-content/uploads/2015/08/SouthCarolina\\_CaseStudy.pdf](https://www.aisp.upenn.edu/wp-content/uploads/2015/08/SouthCarolina_CaseStudy.pdf)

### *What is the urgency of the problem weighed against the potential harm of the solution?*

For data-sharing projects, the main tension at play is preventing misuse while avoiding missed opportunities for using data. It is difficult to know what is appropriate behavior when terms like “fair,” “transparent,” or “consent” have ambiguous meanings and are shaped by a profit motive. The meaning of these terms will remain ambiguous under poor data governance, profit-seeking on behalf of leadership, and lack of consultation of end-users. The [weak regulatory framework](#) around corporate data stewardship in the U.S. has some talking of “[ethics washing](#),” whereby self-regulation is considered to absolve companies of ethical due diligence. The argument goes, in essence, “it’s okay because it’s legal.” While some cities and states have taken concrete action to balance data-driven innovation with individual trust and protection of privacy, in most parts of the country it is up to data scientists to practice responsible [data ethics](#).

### *How can the community be engaged from the very early stages?*

Despite the widespread understanding among leaders that community must be involved to ensure that impact is achieved, evaluation activities still tend to serve the needs of funding organizations over the communities that the projects serve. Figure out if any stakeholders already have strong ties to the community of interest. If you believe that the community’s interests are not being well represented in project development, one way forward is to set up a [community advisory board](#).

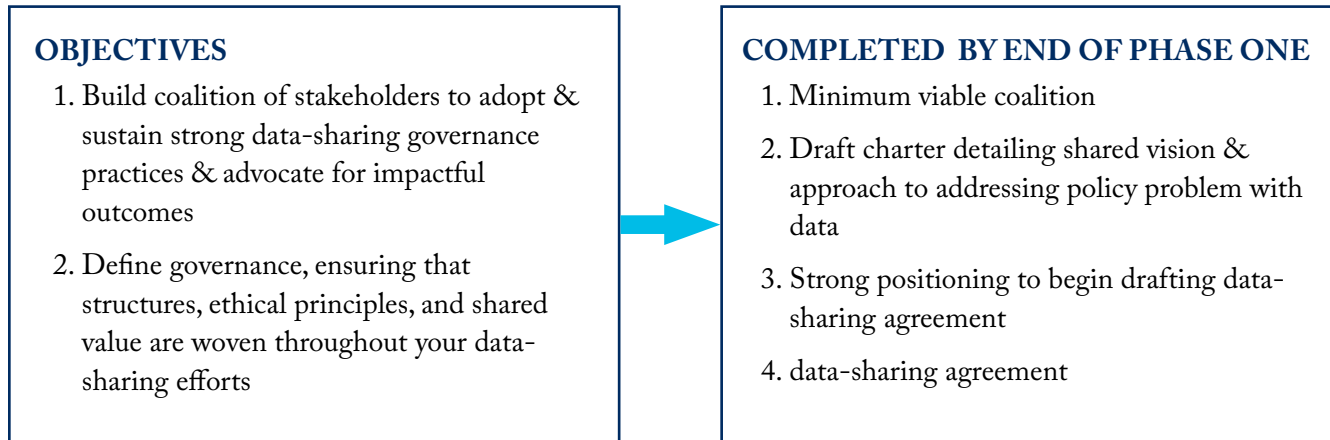
## 1.7 Draft Your Charter

Through the identification of your shared vision, potential barriers, and their possible solutions, draft a charter that provides a foundation for the collective to begin determining its operations. The [Medicaid governance charter](#) provides a good example of how to frame your own charter. Key elements to address include:

- The purpose of your collective
- The context and background surrounding the collective and identification of impacted populations and stakeholders, as well as their values and motivations; acknowledgement of the most salient legal, security, privacy, organizational, and ethical considerations
- The scope of the project and the goals of the collective
- The members of the collective, their roles, and points of expertise
- The authority, if one exists, or the mechanism to confer authority on specific actors to ensure effective operations
- The operational plan, such as decision-making mechanisms, membership change procedures, or rule-making processes
- The mechanism to assess performance
- A timeline and milestones to guide performance assessments and timeline for the collective

## 1.8 To Be Done by the End of Phase One

Phase One will likely be the most difficult and resource-intensive phase as its focus is on building a coalition of stakeholders in order to adopt and sustain strong data-sharing governance practices and advocate for impactful outcomes. It also prepares the collective for defining the governance structure, ensuring that the structures, ethical principles, and shared values are woven throughout your data-sharing efforts. At the conclusion of Phase One, your data-sharing effort should have (1) a minimum viable coalition, (2) a draft charter detailing the shared vision and approach to addressing a policy problem with data, and (3) good positioning to begin drafting a data-sharing agreement in Phase Two.



## PHASE TWO: DEFINE *the* OPERATIONS

In Phase Two, you will create a governance framework that determines roles and responsibilities, accountability, credit, decision-making processes, ethical and security safeguards, and evaluation mechanisms for incorporating feedback when you move forward into Phase Three. In this phase, the collective evolves from a general conception to a specifically defined set of operations with a minimum viable coalition. This includes detailing concrete roles and responsibilities for each party, to ensure continued trust and accountability between parties, and to outline how collective decisions can be made. Processes for changing the governance framework should be outlined. One of the main tensions in the data-sharing governance world [identified by the Data Stewards Network](#) is that between the need for experimentation, customizability, and iterative improvement on one hand, and the comforting certainty of following prescriptive approaches on the other. A well-crafted agreement should create governance-modification rules that address this. By the end of the phase, you should be in a good position to answer questions like those posed in [this data maturity assessment](#). When Phase Two ends, your collective's governance framework will be viable and should begin operations.

### 2.1 Determine Your Governance Framework Structure

*What models for data sharing exist today that support this work?*

[GovLab](#), a research center at NYU that focuses on open and efficient data uses in government, has released a taxonomy of models that can spur and frame data-sharing agreements that it calls “[data collaboratives](#).” This [resource](#) provides a long list of examples of collaborative databases. The [main types](#) it identifies are:

#### From Most to Least Common

1. Corporate data pooling: important data holders like companies or governments collaborate to create databases
2. Research partnerships: private companies share data with research organizations
3. Trusted intermediaries: private companies share data with certain trusted partners
4. Intelligence products: companies share data, usually aggregated, to offer insight into a market, a demographic, or some other trend
5. APIs: give developers access to data for analytics & model testing purposes
6. Prizes & challenges: private companies allow certain qualified researchers access to data to develop innovative ways to use it to a certain end



### *Which model best fits your needs?*

Depending on your goals, the sensitivity or legal restrictions of the data, your organization's mandate, or the resources available, sharing data can take one or multiple of many forms. Review [case studies](#) and consult with stakeholders to see which type of sharing might be most appropriate. The following two examples taken directly from the Data Collaboratives Explorer provide a good understanding of how to model based on needs.

#### *Accelerating Medicines Partnership, Data Pooling*

AMP partners focus on four diseases—Alzheimer's disease, type 2 diabetes mellitus, rheumatoid arthritis, and systemic lupus erythematosus--and share genomic and molecular data to accelerate research into disease treatment.

All AMP partners have agreed to make the AMP data and analyzes publicly accessible to the biomedical community through an online portal. AMP governs its projects through steering committees for each of the three disease areas, with representation from all partners. The steering committees are managed by the Foundation of the National Institutes of Health (FNIH) under the direction of an AMP Executive Committee comprised of representatives from the NIH, participating industry leaders, the FDA, and non-profit organizations.

The AMP initiative combines public-private expertise and pooled data to reduce the time and cost of developing biomarkers for disease treatment. The project attempts to overcome fragmentation in the pharmaceutical industry and improve innovation in drug therapy. By combining data the AMP portal can find new drug targets and reduce wasteful repetition of testing found when companies work in silos.

#### *Capital One Hackathons, API, Prizes and Challenges*

Capital One created an API for its hackathon called Nessie which gives programmers access "to a multitude of real public-facing data - such as ATM and bank branch locations - along with mock customer account data." Participants can use "HTTP requests to set up peer-to-peer transactions, simulate a weekly paycheck, or even schedule bills for customers." This is all structured in a way that resembles how things are run at Capital One.

Hackathon participants can access the data through the Nessie API, which is accessible through GitHub.

The purpose of the hackathons is to encourage co-creation and developer engagement in order to build better customer experiences and address economic challenges with Capital One's APIs.

*How will the public and affected communities be informed and engaged when determining the data-sharing governance process?*

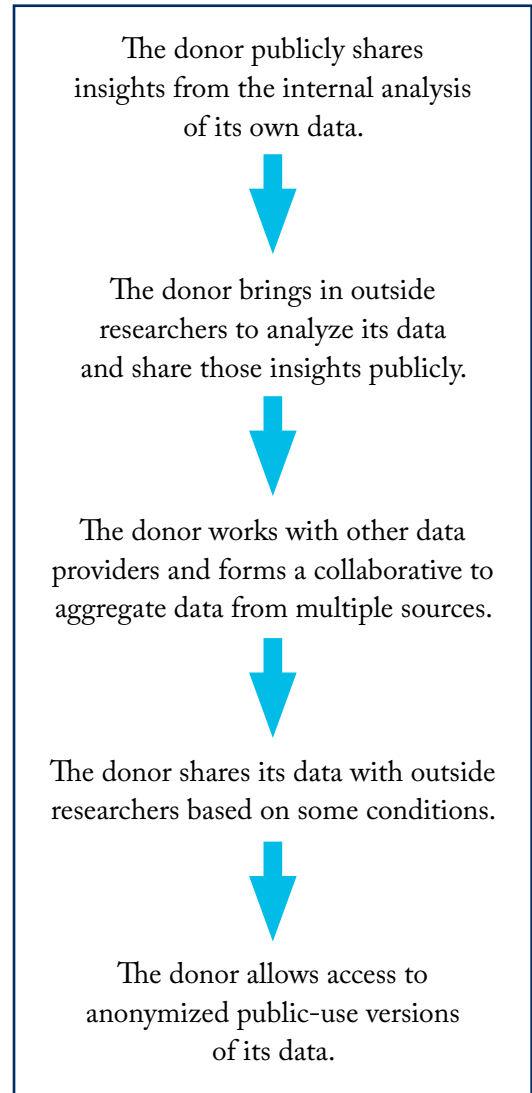
The affected community should at all times be considered as an important stakeholder, as exemplified by much of the work done by the [Future of Privacy Forum](#). Always keep in mind the user-centered-design mantra: “build with, not for.” Design how to receive public feedback through open forums and community outreach. Develop brochures and other educational materials and distribute through different media, social, and organizational websites.

*Who will receive the shared data? In what form will the data be shared?*

There are multiple forms that sharing can take, from sharing access to sharing insights to sharing the data itself. In section 1.3, we showed an important framework put out by the [Urban Institute, which offers](#) a continuum from more to less restrictive access to data that is also relevant for section 2.1.

A review of [many case studies by the Actionable Intelligence for Social Policy \(AISP\)](#) found that government agencies all followed more or less the same procedure for sharing data: first the data are encrypted and sent to a central repository to be analyzed for obvious red flags related to quality that could indicate invalid elements or errors. If any of these red flags are found, the concern is relayed to the original data-holder. Multi-stakeholder working groups may be convened to address these concerns and provide a context for understanding the data. A data dictionary is developed or updated to help staff members understand the data. The data are then cleaned, verified, and linked together. One can look to the AISP case study of Allegheny county as an example:

Allegheny County’s Department of Human Services (DHS) launched its data warehouse in 1999 with initial start-up funds from the Human Services Integration Fund (HSIF)/. Since its development, the DHS data warehouse has served as a central repository that brings together human service and other client data to support a wide range of administrative, decision-making, and policy activities within and external to DHS. Over time, the department expanded efforts and added data sources from other agencies such as the Department of Public Welfare and the Pittsburgh Public Schools. Currently, the DHS data warehouse connects data from DHS programs and a number of external sources including the 10 local public school systems, the courts and jail, and the housing authorities of both Allegheny County and the City of Pittsburgh.



## 2.2 Formalize Responsible Data-Sharing Practices

Building a framework for data-sharing governance essentially boils down to four main elements, according to [interviews conducted by AISP](#): securing and maintaining legal agreements; establishing governance processes; establishing data management and analytics; and addressing the economic and organizational barriers to sustainability. Decide how to structure your data-sharing framework by reviewing templates and standardized data-sharing agreement language, realizing that there is no best practice that applies to all situations and that you should seek the practices that flow from the case studies closest to your own situation. Think through how to customize the data-sharing agreement to craft an appropriate governance framework. Repeat partners may find it onerous to craft agreements from scratch in later ventures and may opt for creating a template.<sup>2</sup> Los Angeles County's creations of an IDS for the delivery services to its homeless populations demonstrates important lessons:

### *Have you established the roles and responsibilities of each party?*

Using standardized contracting processes and not just standardized contracts represents a major opportunity to determine the actual form of the governance structure. Rather than creating a contract or sharing agreement from scratch, these processes can save resource-constrained organizations significant time and effort. Standard processes can also help overcome reticence on behalf of leadership over legal and economic barriers. This [example](#) from the 13th International Conference on Electronic Government provides important use cases for roles and responsibilities.

### *Have you solicited and adapted to feedback from stakeholders throughout the defining process?*

Effective data-sharing governance requires ongoing collaboration that is facilitated by the initial discussion of expectations, documented delineation of responsibilities, and point people responsible for ensuring collaboration. Sufficient resources and political will within the leadership, enabled by buy-in and an understanding of the benefits of sharing, can help guide important considerations surrounding the structure of the governance.

### *How are data used externally beyond the collective?*

Data science solutions in organizations tend to focus on answering organization-specific problems, but, since many organizations in a particular sector are focusing on the same problem, sticking to these uses of data can fail to deliver collective impact. Realizing that your data and your insights become more valuable when they are shared can spark a culture that actively seeks opportunities to share data, share insights, share expertise, or report on the successes or failures of novel approaches to help others repurpose data solutions and improve collective impact. Thus, within the limits of your security and privacy considerations, be generous with sharing outside of the collective.

---

<sup>2</sup> For example, when Los Angeles County created an IDS for the delivery of services to its homeless population, it abandoned project-specific MOUs because they were too cumbersome in favor of a template that outlines data-use practices and procedures for each agency with an attached form that describes specific data-use policies for each project, the details of the project, and the associated list of data sources for agency approval. The standardization of the MOUs expedited the MOU process significantly. See Erika M. Kitzmiller, *IDS CASE STUDY: Los Angeles County* (Philadelphia, PA: Actionable Intelligence for Social Policy, January 2014), [http://www.aisp.upenn.edu/wp-content/uploads/2015/08/LosAngelesCounty\\_CaseStudy.pdf](http://www.aisp.upenn.edu/wp-content/uploads/2015/08/LosAngelesCounty_CaseStudy.pdf)

## 2.3 Examine the Organizational Sustainability

Some of the reasons that data-sharing collectives have traditionally failed to be sustainable is a lack of flexibility, weak goal alignment between stakeholders, an underdeveloped theory of change that fails to map to measurable goals and impact metrics, and a poor sense of the resources organizations have to spare. Many in the civic technology space are still figuring out what the new architecture for data sharing, collaboration, and use looks like; this is why the conversation around sustainability is so important. In-depth coordination in Phase One around these issues should place the collective in a situation conducive to realistic discussion surrounding sustainability.

### *How is the collective funded?*

Responsibility for providing resources for the project should flow from the discussion of the resources, barriers, and capacity that occurred in Phase One. If some organizations in the collective do not have funds to dedicate to this project, perhaps they have non-monetary resources such as expertise, networks, or technological capacity that can be used.

### *For what period of time will this collective exist?*

An important question to ask is “how will you know your job is done?” This can be answered by getting stakeholders to clearly articulate their goals and motivations, by consulting with subject-matter experts, and by taking stock of capacity and resources. Will the collective continue as long as this end-goal has not been met or will you end at a certain date regardless of progress? Your end-goal should guide, at least in part, your choice of metrics and your choice of metrics should track cleanly to different stages of your theory of change.

### *Have rules around modifying the stakeholder mix of the collective been drafted?*

To enable iterative improvement in Phase Three, set up a mechanism for organizations to enter or leave the collective along with adjusting leadership rights and decision-making rights. If organizations leave because they fail to see the value in the collective, they feel they are not being heard, or they feel their goals are no longer being served, there is perhaps a failure in Phase One that needs to be addressed through the feedback loop that is Phase Three. The legal and technical procedures and implications of leaving or entering the collective should be made clear in order to ease uncertainty that might harm the collective’s ability to attract the right stakeholders.

- How does the collective determine who to partner with?
- How can the collective incorporate new members?
- How can existing members leave the collective?
- What technical and legal systems need to be in place for membership changes?

## 2.4 Develop an Ethical Framework for Data Sharing

### *How will your data governance ensure that ethical concerns are considered, heard, and addressed throughout the project?*

Whereas data ethics used to be a discussion around security and privacy, there is a [growing understanding](#) that ethical data stewardship asks us to find new ways to use data to advance the common good while being thoughtful about consequences and transparent about intentions. [Good leadership](#) and a culture that rewards careful considerations of data ethics will always be more effective than oaths or checklists but these can be useful to motivate a discussion. This is important for policy makers as well as organizations to effectively represent their data users.

### *How do you get data practitioners to think about data ethics?*

Data ethics is becoming incorporated into data science certificate and degree curricula and it is now common to find courses in [data ethics](#) in major universities. For data practitioners who entered the labor force before these trends, there are [plenty of training resources available](#) that present data ethics frameworks. Go from “can you build this?” to “should you build this?” Build an organizational culture that sacrifices rushing products out the door for discussions about fairness. Investment in ethics in a time of low public trust and rising consumer concern can be [good for the bottom line](#).

Using just one of about [20 data ethics frameworks](#), data scientists can seek [this training from datapractices.org’s courseware](#). Next, data scientists can start with the FORTS Framework (shown in the graphic below) created by the Global Data Ethics Project to create a code of ethics that emphasizes the following principles:

#### *Fairness:*

Make a dedicated effort to understand, mitigate, and communicate the presence of bias in both data practice & consumption.

#### *Openness:*

Practice humility & openness, since transparent practices, community engagement, and responsible communications are an integral part of a data ethics practice.

#### *Reliability:*

Ensure that every effort is made to glean a complete understanding of what is contained within data, where they came from, and how they were created.

#### *Trust:*

Work to build public confidence in data practitioners and make every effort to use data & algorithms in ways that maximize the informed participation of people around the world.

#### *Social benefit:*

Place people before data and be responsible for maximizing social benefit & minimizing harm. Consider the impact of your work on human communities, other living beings, ecosystems, and the world at large.

Other popular frameworks include:

- [Association for Computing Machinery Code of Ethics and Professional Conduct](#): An annotated code of ethics for computing professionals that outlines general principles, enumerates responsibilities, and offers advice for leadership.
- [Ten Simple Rules for Responsible Big Data Sharing](#): An annotated 10-point framework of key data ethics principles.
- [American Statistical Association’s Ethical Guidelines for Statistical Practice](#): A detailed list of ethical principles aimed at statistical professionals centered around promoting integrity and highlighting responsibilities.

- Data Science Oath of National Academies of Science, Engineering, and Medicine: An actual oath that data practitioners can take, situated alongside the Hippocratic oath.
- [Manifesto for Data Practices of data.world](#): A simple 12-point list of ethical principles centered around four values: inclusion, experimentation, accountability, and impact.

### *How do you get data practitioners to act on principles of data ethics?*

Instead of being ends in and of themselves, frameworks should be used as launching points to motivate and frame discussion about ethics and thoughtful self-governance.

While off-the-shelf oaths and principles of data ethics can serve as a great starting point, a [recent booklet](#) by Mike Loukides, Hilary Mason, DJ Patil highlights some important limitations:

1. These oaths and principles are one-off commitments and there is no guarantee that you will continue considering them and evaluating your adherence over the long term.
2. Principles and terms are ambiguous and good data are more about execution than theory.
3. “Taking the oath” can give cover to organizations which interpret the terms loosely to justify less-than-ethical work.
4. Oaths fail to connect principles to practice.

These authors suggest that checklists are different than oaths in that they connect principles to practice, like the UK government’s Data Ethics Framework and Data Ethics Workbook that ask open-ended questions to probe compliance with certain identified principles.

Some government agencies use an institutional review board (IRB) process to review projects, but this approach has been criticized for being too slow and inflexible at a recent convening of the [Data Stewards Network](#). That convening suggested an alternative approach that could fulfill ethical review purposes that could consist of one-off “ethical councils” that bring together thought leaders from academia, business, the public sector, and civil service to provide counsel over the ethics of a project. However, this approach should be temporary and it cannot replace the effectiveness of internal data ethics capacity that periodically evaluates new and ongoing considerations within a data management team.

The norms and practices surrounding data sharing are constantly evolving, so the ethical considerations of data sharing should be an ongoing discussion. It should also highlight the importance of flexibility and iterative improvements in the data-sharing agreement drafting and updating process. Data ethics principles do not need to be relegated to a voluntary-only basis and can often be codified as an addendum to a data-sharing agreement paired with checklists that avoid dictating vague principles in favor of verifiable action items.

## 2.5 Define What Success Looks Like

You must establish how to define success. What impact metrics will you examine? How do you see your theory of change being realized? Continue to engage stakeholders in order to ensure that your impact is in accordance to your previously established ethical principles.

### *What impact metrics are appropriate?*

Impact metrics are data that help you assess progress towards achieving a predetermined impact. [Deloitte's review](#) of the future of effective impact evaluation includes three key elements:

*Purpose (why):* Choose impact metrics that feed into decision-making mechanisms of grantor & grantee. Avoid the “streetlight effect” of using the most available data at the expense of more effective metrics. Start with determining a purpose, then find an approach, and only then find indicators that match.

*Perspective (who):* Design participation to empower end-users & promote a diverse, equitable, and inclusive process to determine what is needed, what success looks like, and what impact is happening. When you choose a metric, you choose whose perspective you prioritize; doing so can either reinforce or address power inequalities.



*Alignment with other actors (what):* Develop collective knowledge and share insights & learning from successes & failures of other organizations in the same space.

## 2.6 Draft Data-Sharing Agreement

The core function of a data-sharing agreement is to formalize the roles and responsibilities, decision-making mechanisms, and accountability processes that support the actions outlined in the charter. This might include answering questions such as:

- What are the data to be used; how will they be shared and used?
- Which stakeholders are involved in what part of the data collection, processing, and analysis?
- What security and ethical safeguards have been put in place?
- How are collective decisions made and what process do these decisions go through?
- When are confidentiality agreements required and how are they drafted?
- How are ethical and security [reviews and audits](#) conducted?
- How are external requests for data or insights including review, approval, and pricing addressed?
- Who contributes what resources to the project?

- Who owns the intellectual property of the data and of the resulting publications and who decides whether to publish or share them externally?
- Who gets credit for what part of the collective?
- How are potentially ambiguous terms that might interfere with mutual understanding of the contract defined, especially if the partners are from different sectors?
- How are changes that members want addressed?

## 2.7 To Be Done by the End of Phase Two

By the end of Phase Two, members will have defined the governance activities tied to the charter and to overcoming the barriers identified in Phase One, signed a data-sharing agreement that formalizes and commits to that governance structure, and determined how that agreement will be implemented by a governing board. The board would put into practice the governance framework that defines the roles and responsibilities of each stakeholder, the funding and decision-making mechanism, and the process of entering and leaving the collective. You will have formalized a mechanism for evaluation and iterative improvement using impact metrics that agree with your theory of change. Ethical principles, connected to specific actions, are to be enumerated and baked into the implementation and review process.

### BY ALL PARTIES

1. Defined governance activities tied to charter & to overcoming barriers identified in Phase One
2. Signed data-sharing agreement that formalizes & commits to governance structure
3. Determined how agreement will be implemented by a governing board

### BY THE BOARD

1. Have put into place governance framework that defines roles & responsibilities of each stakeholder
2. Have put into place funding & decision-making mechanisms
3. Have put into place processes for entering & leaving the collective

### BY YOU

1. Formalized mechanism for evaluation & improvement using impact metrics that agree with your theory of change
2. Enumerated ethical principles, connected to specific actions, baked into the implementation & review process



## PHASE THREE: DRIVE COLLECTIVE IMPACT

In this final phase, the collective has been established and the operations have begun. Using the feedback loop that was integrated into the governance model developed in Phase Two, [continue](#) to [revisit](#) and [revise](#) the operational, ethical, and legal aspects of the collective. Members should continue to adhere to best practices and standards previously established and document opportunities and challenges as they come up, not just to support their own improvement but to further best practices to help the entire sector.

### 3.1 Reinforce, Update, and Share Governance Best Practices

From an operational perspective, the data collection process can be improved as the collective develops. This might include collecting better metadata and documentation, refining data structures, and improving security standards. The framework also needs to have the flexibility to change in order to minimize human errors based on insights generated by the technicians and stewards who are managing the data. This iterative process not only improves the specific data-sharing project but should improve data-sharing capacity for all parties over time. Throughout execution of the agreement, ask yourself the following questions in support of this goal:

- How can you ensure that parties continue to adhere to previously established standards and practices?
- How can you change the data-sharing agreement based on the lessons you learn from evaluating progress towards goals on each step of your theory of change?
- How do you decide on new research projects?
- How can iterative improvements best be documented and publicized to stay transparent, help other organizations learn from mistakes and failures, and enrich the literature to promote data-sharing for social impact elsewhere?
- How do you minimize and assess risk as you continue working?

The ethical aspects of data sharing should continue to be addressed in this phase. It is important to establish some kind of feedback loop with the community to ensure that the earlier ethical and impact goals of the collective continue to be met. Create an internal system that prevents and reports misuse of data.

**Encourage Dialogue.** Both between and within the stakeholder organizations and the affected communities, link goals to the impact metrics and situate them within the theory of change. Specific steps you can take towards this goal include:

- Communicate insights to the community and seek feedback.
- Create opportunities and incentives for stakeholders to voice concerns and share experiences with each other with an eye towards governance improvement.
- Seek feedback from the affected communities.
- Implement a review and adjust your process based on feedback to continually update and improve the structure of the data-sharing agreement and the governance structure that implements it.
- Actively promote mutual awareness and teach community how to best protect and use its data (taken from the [NCVHS roundtable](#))

## 3.2 Reinforce, Update, and Share Best Practices in Ethics

In Phase One, you began thinking about the ethical implications of your use of shared data and what capacity you had for thinking about data ethics. In Phase Two, you created the actual mechanisms for continuously incorporating data ethics into your actions. Phase Three is when you implement those mechanisms and encourage feedback about ongoing and new ethical concerns. Beyond thinking about the issues originally identified, ask yourself if any of the following elements have changed:

- Who are the affected communities?
- Are there new externalities, and have previously identified ones been properly addressed?
- Has technology changed in a way that increases the potential for data misuse?
- Can fail-safes be implemented to prevent misuse of data?
- Is conducting an equity audit appropriate?

## 3.3 Monitor and Assess Privacy and Security Approaches

Phase Three is when you sustain established governance practices and seek feedback about ongoing and new privacy and security concerns. This can take many forms, like a community feedback loop, and must be combined with an evolving approach on the current technological environment surrounding privacy and security. You will grow from your initial thoughts and mechanisms in Phases One and Two to a continually improving privacy and security approach that changes over time. The [USAID framework for assessing risk](#) is helpful in considering how to use data skillfully. Beyond thinking about the issues originally identified, ask yourself if any of the following elements have changed and identify what those changes might mean for privacy and security:

- Has the structure of the collective changed?
- Has the technological environment changed?
- Has the regulatory environment changed?
- Has the organizational culture of any stakeholder organization changed?
- Has society changed in ways that affect the impacted communities?
- Are there system processes that can be put in place to minimize or resolve human errors?

### 3.4 Develop Processes to Improve Data Quality Over Time

During Phase Two, consultation with data management teams and subject-matter experts generated insight into the quality and format opportunities and issues of the data. Like all other parts of the collective, what is considered appropriate formatting and quality may change over time. If the data have to be cleaned, reshaped, labelled, or otherwise transformed in order to become usable for purposes of the project, the data-holding organization should consider changing the way the data are collected to help avoid having to repeat this process in future iterations of the collective and in future data-sharing agreements. In some cases, the data source itself may be sub-optimal and effort should be put into finding a source better suited to some sort of pre- or post-comparison or benchmarking. Some of the key questions you should ask are:

- What metadata can be collected to make the data more useful?
- How can the documentation process be improved and how can insight about potential bias be baked in?
- How can capacity be built over time, whether technical infrastructure or data management expertise?
- How has the situation or our understanding changed [in a way that changes](#) our idea of effective impact metrics? Can it be changed to improve benchmarking and comparisons with other efforts?

### 3.5 To Be Done by the End of Phase Three

Phase Three is essentially split into two elements: making iterative improvements to the governance framework and sharing best practices. When making iterative improvements, you are seeking to gather feedback and incorporate lessons drawn from it into the data-sharing governance, and the agreement that formalizes it, in Phase Two. When sharing best practices, you are seeking to model your actions after those of other successful endeavors. You will then be able to share what works for an organization like yours going forward. It is always important to keep in mind the agency of the community you are working with. Only through effective community engagement can you continue to make these improvements and use best practices.

# CONCLUSION: COMMUNITY *of* PRACTICE

This entire framework builds an environment where stakeholders and actors can both focus on acting for good in society. The goal of all three phases—building the collective, defining the operations, and driving the impact—is to use data sharing for social good. By using this building framework that emphasizes 1) overcoming potential data-sharing barriers, 2) emphasizing ethical operations, and 3) stressing community engagement, everyone involved in the process has a foundation of understanding of responsible data sharing. This will allow for a shared understanding of who is impacted by the data sharing and how the linked data can best support improved social service delivery.

## Phase 1: Build the Collective

- Understand gap filled by data sharing
- Understand data being shared
- Understand ethical implications

## Phase 2: Defining the Operations

- Define success
- Establish internal trust within the collective
- Define operations of the collective

## Phase 3: Driving Impact

- Formalize ethical principles and guidelines
- Establish feedback loop for evaluating impact metrics
- Create internal monitoring and assessment

Today, data-sharing governance may be short-sighted, isolated, imposed from the outside, and even self interested, but the more people and organizations begin to incorporate the details set forth in this guidebook like minimum viable coalitions, data-sharing agreements, community feedback loops, and iterative frameworks for improvement, the more responsibly they will move forward with data sharing. Simply put, data collaboratives can have a stronger force for impact and good in society.



*GEORGETOWN UNIVERSITY*

# EXECUTIVE SUMMARY

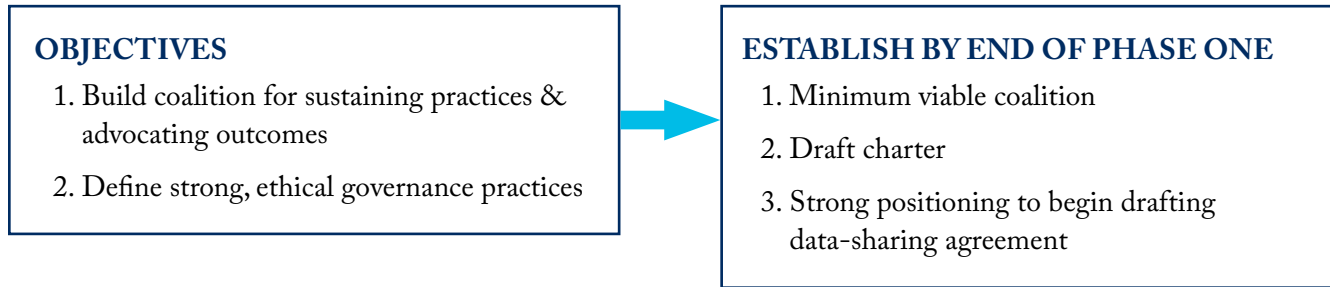
Many government agencies have well-established use cases for sharing data in pursuit of improved social service delivery in areas such as [K-12 education](#), public transportation, and [healthcare](#). Florida's Broward County, for example, instituted a community feedback loop on data sharing to strengthen family incomes. [Los Angeles County](#) created an integrated data system (IDS) for the delivery of services to its homeless population with a template that outlines data-use practices and procedures for each agency with a form that describes specific data-use policies for each project, the details of the project, and the associated list of data sources for agency approval. On a larger scale, the National Committee on Vital and Health Statistics held a roundtable about self improvement and building trust in the community to drive impact.

But cross-sector data-sharing collaboratives have been a more recent phenomenon, which private companies, governments, and nonprofits are increasingly forming them in support of the social good. While we've seen many success stories in deploying data-sharing collaboratives, many government agencies are also experiencing the pains of data-sharing governance practices that are limited and under-resourced, resulting in practices that can be unsustainable, not effective, and not forward-thinking.

Organizational rules for data sharing typically advance risk mitigation strategies focused on restricting the availability of data to external actors. In an environment where data are only used for making funding decisions or to narrowly evaluate programs, this model can work well. In pursuit of innovation or improved social service delivery, this model is less encouraging. As the amount of data and methods for collecting it increase, so have opportunities for drawing insights about society. Bringing together diverse data sources is crucial to ensuring that insights promote equitable growth. And as promising as data sharing is for improving societal outcomes, the analysis of integrated data (especially through predictive analytics) can easily replicate inequities learned from patterns of past service delivery. Contextualizing data analysis with methods used by the social sciences and ongoing community engagement is crucial to ensuring data analytics do not replicate or worsen inequitable outcomes.

There is a fundamental need to reform data-sharing governance practices to ensure their effectiveness and sustainability. We advance here a guidebook, each step of which is equipped with resources and case studies, to aid stakeholders and actors as they navigate the legal, technical, organizational, and societal challenges to creating good governance practices for responsible and impactful data-sharing collaboratives. This framework aims to provide a holistic guide that encourages data-sharing parties to consider technical, legal, as well as ethical components of sharing. It is divided into three main phases:

**Build the collective:** get everyone on board. Start with the policy problem. Identify stakeholders. Take stock of capacity, motivations, barriers, and potential data solutions. Demonstrate value and reduce uncertainty to generate buy-in. Establish a minimum viable coalition and enshrine your shared vision in a charter.



**Define the operations:** get everyone in line. Create the governance framework tied to the charter. Design a feedback loop and integrate it into the governance framework. Formalize those two elements into a data-sharing agreement. Launch the operations of the minimum viable coalition.

**BY ALL PARTIES**

- 1. Define governance activities tied to the charter and to overcoming barriers identified in Phase One
- 2. Sign data-sharing agreement to formalize governance structure
- 3. Determine how agreement will be implemented by governing board

**BY THE BOARD**

- 1. Put into practice governance framework that defines the roles & responsibilities of each stakeholder
- 2. Put into place funding & decision-making mechanisms
- 3. Put into place process for entering & leaving collective

**BY YOU**

- 1. Formalize mechanism for evaluation & improvement with impact metrics that agree with your theory of change
- 2. Enumerate ethical principles, connected to specific actions, baked into implementation & review process

**Drive impact:** get everyone to improve and share. Re-evaluate assumptions, approach, and metrics. Survey impacted communities and stakeholders. Use feedback loops to enact iterative improvements to the governance structure. Repeat this process until feedback becomes minimal. Scale up. We recognize that many different actors will be involved in this process and that each actor may face unique challenges, goals, motivations, and opportunities. This guidebook is for people looking to understand how they can leverage data and data sharing towards evidence-based policy making. Moreover, it can be used by policy makers and organizations interested in giving agency to individuals over their data along with organizations interested in ethically and responsibly sharing data.

*Two Main Elements of Driving Impact*

**Element one: Creating iterative improvement**

**Goal:** Gather feedback & incorporate lessons into the data-sharing governance and the agreement that formalizes it, formed in Phase Two

**Element two: Sharing best practices**

**Goal:** Model actions after those of other successful endeavors & share what works for an organization like yours with others going forward