Sharing Data for Social Impact: Guidebook to Establishing Responsible Governance Practices

This report, from Beeck Center Fellow Natalie Evans Harris, lays out a three-phased approach to adopting sustainable data sharing governance practices including resources, case studies, and best practices.

Unlocking the potential of data to help solve social problems requires a robust framework for data governance and data sharing. A framework can address major points of risk and ambiguity that prevent many actors from engaging in meaningful data sharing, and better data sharing and data governance will in turn lead to improved social service delivery.

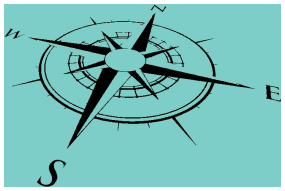
Contents

Vision....4
Mission....4
Values....4
Phase One. The Collective....5
1.1. Definition....5
1.1.1. Problems....5
1.1.2. Change & Impact....5
1.1.3. Stakeholders....6
1.2. Vision....6
1.2.1. Motivations & Goals....6
1.2.2. Barriers....7
1.2.2.1. Legal/Regulatory....7
1.2.2.2. Organizational/Motivational....7
1.2.2.3. Social/Ethical....7
1.2.2.4. Technical/Data-Related....7
1.3. Data Sharing....8
1.3.1. Capacities....8
1.3.2. Technical Barriers....8
1.3.3. Data Literacy....9
1.3.4. Recipients & Formats....9
1.4. Understanding....9
1.4.1. Data Quality....9
1.4.2. Data Standards....10
1.5. Security & Privacy....10
1.5.1. Sensitivity....10
1.5.2. Encryption....10
1.5.3. Privacy....11
1.5.4. Anonymization....11
1.5.5. Recovery....11
1.6. Ethics....11
1.6.1. Impacted Populations....12
1.6.2. Urgency & Harms....12
1.6.3. Engagement....12
1.7. Charter....12
Phase Two. Operations....13
2.1. Governance....13
2.1.1. Needs....14
2.1.2. Information & Engagement....14
2.1.3. Recipients & Formats....14
2.2. Data Sharing....15
2.2.1. Roles & Responsibilities....15

DEMONSTRATION ONLY

2.2.2. Feedback....15
2.2.3. Usage....15
2.3. Sustainability....16
2.3.1. Funding....16
2.3.2. Persistence....16
2.3.3. Stakeholders....16
2.4. Ethics....17
2.4.1. Procedures....17
2.4.2. Consideration....17
2.4.3. Action....18
2.5. Success....18
2.5.1. Purpose....19
2.5.2. Perspective....19
2.5.3. Alignment....19
2.6. Agreement....19
Phase Three. Impact....20
3.1. Best Practices....20
3.1.1. Insights & Feedback....20
3.1.2. Opportunities & Incentives....20
3.1.3. Feedback....21
3.1.4. Review & Adjustment....21
3.1.5. Awareness & Teaching....21
3.2. Ethics....21
3.2.1. Affected Communities....21
3.2.2. Externalities....21
3.2.3. Technological Change....21
3.2.4. Fail-Safes....22
3.2.5. Equity Audits....22
3.3. Privacy & Security....22
3.3.1. Structure....22
3.3.2. Technology....22
3.3.3. Regulations....22
3.3.4. Organizational Culture....22
3.3.5. Social Change....23
3.3.6. Human Errors....23
3.4. Data Quality....23
3.4.1. Metadata....23
3.4.2. Documentation & Bias....23
3.4.3. Capacities....23
3.4.4. Metrics....24
Administrative Information....24

DEMONSTRATION ONLY



Beeck Center for Social Impact & Innovation (BCSII)

Description:

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.

Stakeholder(s):

Digital Service 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.

Natalie Evans Harris :
Beeck Center Fellow, 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.

The Rockefeller Foundation :
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.

Student Analysts :
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.

Nitya Biyani

Moonlan Zhang

Robert Roussel

Ben Lang

Experts :
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.

Bill Yock

Sara Holoubek

Kathryn Pettit

Bill Hoffman

Reg Leichty

Leah Hendey

Michelle Shevin

David Park

Dan Wu

Kristin Wolff

Kevin O'Neil

BrightHive :
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.

Stakholders :
Those with a vested interest in data sharing. Those from whom data is being collected. Examples: parents; community members; program managers; IT staff

— continued next page

DEMONSTRATION ONLY

Stakeholders (continued)

Policy Makers :
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

Policy Staff

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

Vision

Sustainable data sharing

Mission

To lay out an approach to the adoption of sustainable data sharing governance practices

Values

Data

Sharing

Sustainability

Governance

DEMONSTRATION ONLY

Phase One. The Collective

Build the Collective

Stakeholder(s)

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

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

Coalitions :
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

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.

1.1. Definition

Define the Collective

1.1.1. Problems

Identify the problem to be solved.

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 of 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.

1.1.2. Change & Impact

Specify the impact to be achieved as well as the theory of change.

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.

DEMONSTRATION ONLY

1.1.3. Stakeholders

Define the key stakeholders.

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.

Stakeholder(s):

Data Sharing Stakeholders :
The stakeholder landscape includes:

Data-Providing Organization :
Leadership or data stewards in the data-providing organization

Demand-Side Organization :
The demand-side organization

Data Providers :
Third-party data providers with complementary data

Data Intermediaries :
Third-party intermediaries who facilitate sharing and data use

Funders

Government Agencies

User Advocates

Data Users :
Users themselves

1.2. Vision

Establish Shared Vision

1.2.1. Motivations & Goals

Determine the motivations and goals of the data-sharing stakeholders.

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-fitsall, 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.

DEMONSTRATION ONLY

1.2.2. Barriers

Identify the barriers to sharing data and the means to overcome them.

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," Susha, 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:

1.2.2.1. Legal/Regulatory

Identify legal/regulatory barriers.

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

1.2.2.2. Organizational/Motivational

Identify organizational/motivational barriers.

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

1.2.2.3. Social/Ethical

Identify social/ethical barriers.

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

1.2.2.4. Technical/Data-Related

Identify technical/data-related barriers.

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

DEMONSTRATION ONLY

1.3. Data Sharing

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.

Stakeholder(s):

Data Management Teams :
Well-run data management team:

- *Data storage and administration capacity*
- *Data integration capacity*
- *Data analytics capacity*
- *Data publication capacity*

1.3.1. Capacities

Take stock of data-sharing capacity.

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.

1.3.2. Technical Barriers

Identify technical barriers to data sharing.

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.

DEMONSTRATION ONLY

1.3.3. Data Literacy

Create data literacy programs/processes.

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.

1.3.4. Recipients & Formats

Determine who will receive the shared data and in what form.

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. More to Less Restrictive:

- The donor publicly shares insights from the internal analysis of their own data.
- The donor brings in outside researchers to analyze their data and share those insights publicly.
- The donor works with other data providers and form a collaborative to aggregate data from multiple sources.
- The donor shares the data with outside researchers based on some conditions.
- The donor allows public access to anonymized public-use versions of its data.

1.4. Understanding

Understand the Data Being Shared

1.4.1. Data Quality

Assess the quality of the data.

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.

DEMONSTRATION ONLY

1.4.2. Data Standards

Identify the appropriate data standards.

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. Security & Privacy

Survey Security and Privacy Related to the Data

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:

1.5.1. Sensitivity

Determine sensitivity of the data.

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.

1.5.2. Encryption

Determine how encryption will be managed through the collection, transit, and access of data.

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.

DEMONSTRATION ONLY

1.5.3. Privacy

Assess the feelings of the relevant communities towards privacy.

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.

1.5.4. Anonymization

Determine how anonymization should be employed to protect personal data.

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.

1.5.5. Recovery

Determine the recovery process.

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.1 This can act as an important safeguard that can help recover from data loss, accidental manipulation, and loss of documentation surrounding processing.

1.6. Ethics

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.

DEMONSTRATION ONLY

1.6.1. Impacted Populations

Address the ethical considerations associated with impacted population.

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.

1.6.2. Urgency & Harms

Assess the urgency of the problem against the potential harm of the solution.

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.

1.6.3. Engagement

Determine how to engage the community from the very early stages.

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. Charter

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 rulemaking processes • The mechanism to assess performance • A timeline and milestones to guide performance assessments and timeline for the collective

DEMONSTRATION ONLY

Phase Two. Operations

Define the Operations

Stakeholder(s)

Parties :
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

Boards :
BY THE BOARD 1. Have put into place governance framework that defines roles & responsibilities of each stakeholder 2. Have put into place funding & decisionmaking mechanisms 3. Have put into place processes for entering & leaving the collective

Individuals :
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

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... 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.

2.1. Governance

Determine 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

DEMONSTRATION ONLY

2.1.1. Needs

Determine which model best fits your needs.

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.

Stakeholder(s):

Data Collaboratives

2.1.2. Information & Engagement

Determine how the public and affected communities will be informed and engaged.

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.

Stakeholder(s):

The Public **Affected Communities**

2.1.3. Recipients & Formats

Determine who will receive the shared data and in what form.

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.

Stakeholder(s):

Actionable Intelligence for Social Policy :

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 :

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, decisionmaking, 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.

City of Pittsburgh

DEMONSTRATION ONLY

2.2. Data Sharing

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.2 Los Angeles County's creations of an IDS for the delivery services to its homeless populations demonstrates important lessons:

2.2.1. Roles & Responsibilities

Establish the roles and responsibilities of each party.

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.

2.2.2. Feedback

Solicit and adapt to feedback from stakeholders

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.

2.2.3. Usage

Determine how data used externally beyond the collective.

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.

DEMONSTRATION ONLY

2.3. Sustainability

Examine the Sustainability of the Operations

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.

2.3.1. Funding

Determine how the collective will be funded.

How is the collective funded? Responsibility for providing resources for the project should flow from the discussion of the resources, barriers, and capacity that occured 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.

2.3.2. Persistence

Determine how long the collective will exist.

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.

2.3.3. Stakeholders

Draft rules governing the modification of the stakeholder mix of the collective.

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?

DEMONSTRATION ONLY

2.4. Ethics

Develop an Ethical Framework for Data Sharing

Stakeholder(s):

Global Data Ethics Project :
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.*

Association for Computing Machinery :
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 :
Ten Simple Rules for Responsible Big Data Sharing: An annotated 10-point framework of key data ethics principles.

American Statistical Association :
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

National Academies of Science, Engineering, and Medicine :
Data Science Oath of National Academies of Science, Engineering, and Medicine: An actual oath that data practitioners can take, situated alongside the Hippocratic oath.

data.world :
Manifesto for Data Practices of data.world: A simple 12-point list of ethical principles centered around four values: inclusion, experimentation, accountability, and impact.

2.4.1. Procedures

Ensure that ethical concerns are considered, heard, and addressed.

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.

2.4.2. Consideration

Encourage data practitioners to think about data ethics.

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

DEMONSTRATION ONLY

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.

Stakeholder(s):

Data Practitioners

2.4.3. Action

Encourage data practitioners to act on principles of data ethics.

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.

Stakeholder(s):

Data Practitioners

Mike Loukides :

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.

Data Stewards Network :

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. Success

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:

DEMONSTRATIVE ONLY

2.5.1. Purpose

Choose impact metrics that feed into decision-making mechanisms of grantor & grantee.

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.

2.5.2. Perspective

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.

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.

2.5.3. Alignment

Develop collective knowledge and share insights & learning from successes & failures of other organizations in the same space.

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

2.6. Agreement

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?

DEMONSTRATION ONLY

Phase Three. Impact

Drive 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... 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.

3.1. Best Practices

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:

3.1.1. Insights & Feedback

Communicate insights to the community and seek feedback.

3.1.2. Opportunities & Incentives

Create opportunities and incentives for stakeholders to voice concerns and share experiences with each other with an eye towards governance improvement.

DEMONSTRATION ONLY

3.1.3. Feedback

Seek feedback from the affected communities.

3.1.4. Review & Adjustment

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.

3.1.5. Awareness & Teaching

Actively promote mutual awareness and teach community how to best protect and use its data.

(taken from the NCVHS roundtable)

3.2. Ethics

Reinforce, Update, and Share Best Practices in Ethics Governance

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:

3.2.1. Affected Communities

Identify the affected communities.

Who are the affected communities?

3.2.2. Externalities

Determine whether externalities have been properly addressed.

Are there new externalities, and have previously identified ones been properly addressed?

3.2.3. Technological Change

Assess how technology may have changed in ways that increase the potential for data misuse.

Has technology changed in a way that increases the potential for data misuse?

DEMONSTRATION ONLY

3.2.4. Fail-Safes

Determine if fail-safes be implemented to prevent misuse of data.

Can fail-safes be implemented to prevent misuse of data?

3.2.5. Equity Audits

Determine whether to conduct an equity audit.

Is conducting an equity audit appropriate?

3.3. Privacy & Security

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:

3.3.1. Structure

Consider whether the structure of the collective has changed.

Has the structure of the collective changed?

3.3.2. Technology

Consider changes in the technological environment.

Has the technological environment changed?

3.3.3. Regulations

Consider changes in the regulatory environment.

Has the regulatory environment changed?

3.3.4. Organizational Culture

Consider whether the organizational culture of any stakeholder organization changed.

Has the organizational culture of any stakeholder organization changed?

DEMONSTRATION ONLY

3.3.5. Social Change

Consider whether society changed in ways that affect the impacted communities.

Has society changed in ways that affect the impacted communities?

3.3.6. Human Errors

Consider whether there system processes that can be put in place to minimize or resolve human errors.

Are there system processes that can be put in place to minimize or resolve human errors?

3.4. Data Quality

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:

3.4.1. Metadata

Consider what metadata can be collected to make the data more useful.

What metadata can be collected to make the data more useful?

3.4.2. Documentation & Bias

Consider how the documentation process can be improved and insight about potential bias can be taken into account.

How can the documentation process be improved and how can insight about potential bias be baked in?

3.4.3. Capacities

Consider how capacity can be built.

How can capacity be built over time, whether technical infrastructure or data management expertise?

DEMONSTRATION ONLY

3.4.4. Metrics

Consider changes related to impact metrics.

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?

Administrative Information

Start Date:
End Date:

Publication Date: 2020-01-31
Source: https://beeckcenter.georgetown.edu/report/sharing-data-for-social-impact-guidebook-to-establishing-responsible-governance-practices/

Submitter:
Given Name: **Surname:**
Email: **Phone:**

DEMONSTRATION ONLY