OPEN
ENVIRONMENTAL
DATA PROJECT

COMMUNITY DATA PLAYBOOK

GOVERNANCE STRATEGIES FOR ENVIRONMENTAL DATA STEWARDS

EMELIA WILLIAMS · KATIE HOEBERLING · CATHY RICHARDS

TABLE OF CONTENTS

ACKNOWLEDGMENTS 3

INTRODUCTION 4

SECTION ONE: COMMUNITY CASE STUDIES 8

- SELF-HELP ENTERPRISES 10
- CHARRS 22
- COACT 30

SECTION TWO: A PLAYBOOK TO ADVANCE ENVIRONMENTAL DATA GOVERNANCE 40

SECTION THREE: ASSESSING THE COMMUNITY DATA HUBS MODEL 60

CONCLUSION 65

ACKNOWLEDGMENTS

Open Environmental Data Project Team

We want to acknowledge the work of the OEDP team, alumni, and partners broadly, in creating this body of work and Playbook. This work is the result of careful deliberation and teamwork—all of which cannot be conveyed within an Acknowledgments page. With gratitude, we recognize the OEDP team members below who supported the development of this Playbook through thought leadership and conceptualization, workshop coordination and facilitation, research and writing, revision and editing, and administrative support.

Michelle Cheripka, *Senior Communications and Content Manager*
Shannon Dosemagen, *Senior Fellow*
Brittany Janis, *Executive Director*
Katie Hoeberling, *Director of Policy Initiatives*
Cathy Richards, *Civic Science Fellow and Data Inclusion Specialist*
Lyra Elizabeth Tyson, *Community Data Hubs Advisor*
Emelia Williams, *Research and Policy Manager*
Kate Wing, *Community Data Hubs Advisor*
Megan Zimroth, *Operations Manager*

Advisory Group

We want to acknowledge the significant contributions of our Advisory Group in shaping the strategic outlook of this project and playbook. This team provided thought leadership and strategic guidance, sharing foundational literature, and offering substantial feedback in review. Participants in the Advisory Group include:

Tim Davies, *Connected By Data*
Persephone Lewis, *Whose Knowledge Foundation*
Matías Milia, *University of Notre Dame*
Luis Felipe Rosado Murillo, *University of Notre Dame*
Kate Wing, *Intertidal Agency*

Community Partners

We want to acknowledge the openness, trust, and thoughtfulness that our community partners offered our team during our collaboration. Community partners at CHARRS, CoAct, and Self-Help Enterprises not only participated in the meetings and workshops with a great deal of care, but also supported the playbook review process with fact checking, revision, and multimedia contributions. Participants from each community partner are listed below:

Gwen Smith, *CHARRS*
Na'Taki Osbourne Jelks, *CHARRS and West Atlanta Watershed Alliance*
Guillermina Actis, *CoAct*
Valeria Arza, *CoAct*
Leticia Castro Martinez, *CoAct*
Giselle Munno Dithurbide, *CoAct*
Jeremías Fabiano, *CoAct*
Emily McCague, *Self-Help Enterprises*
Tami McVay, *Self-Help Enterprises*
Andrew Lockman, *Self-Help Enterprises*
David Visser, *Self-Help Enterprises*
Melissa Looney, *Self-Help Enterprises*
Julie Cabanas, *Self-Help Enterprises*
Kent Stahl, *Self-Help Enterprises*
Christine Austin, *Self-Help Enterprises*

The creation of this playbook was generously supported by the Omidyar Network and the Gordon and Betty Moore Foundation.

OMIDYAR NETWORK

GORDON AND BETTY MOORE FOUNDATION

GRAPHIC DESIGN AND ILLUSTRATIONS BY SUZE MYERS ↗
SATELLITE IMAGES VIA CALTOPO.COM ↗

INTRODUCTION

In response to climate and environmental injustices, communities around the world are collecting and using environmental data to strengthen awareness, engagement, advocacy, or research. There are myriad examples of community organizations utilizing environmental data in generative and innovative ways, spanning the gamut of environmental challenges. In New Orleans, coastal communities have flown balloons ↗ with cameras to photograph the Deepwater Horizon oil spill. In upstate New York, community scientists partnered with university researchers ↗ to hold the Tonawanda Coke Plant accountable for decades-long industrial pollution in the air and soil. In Houston, environmental justice activists created a way ↗ for members of the public to report common signs of pollution, such as flares, black smoke, or acrid chemical smells, and routinely flag these incidents to local authorities.

Yet in doing this work, communities often face barriers in using data to its fullest extent, including:

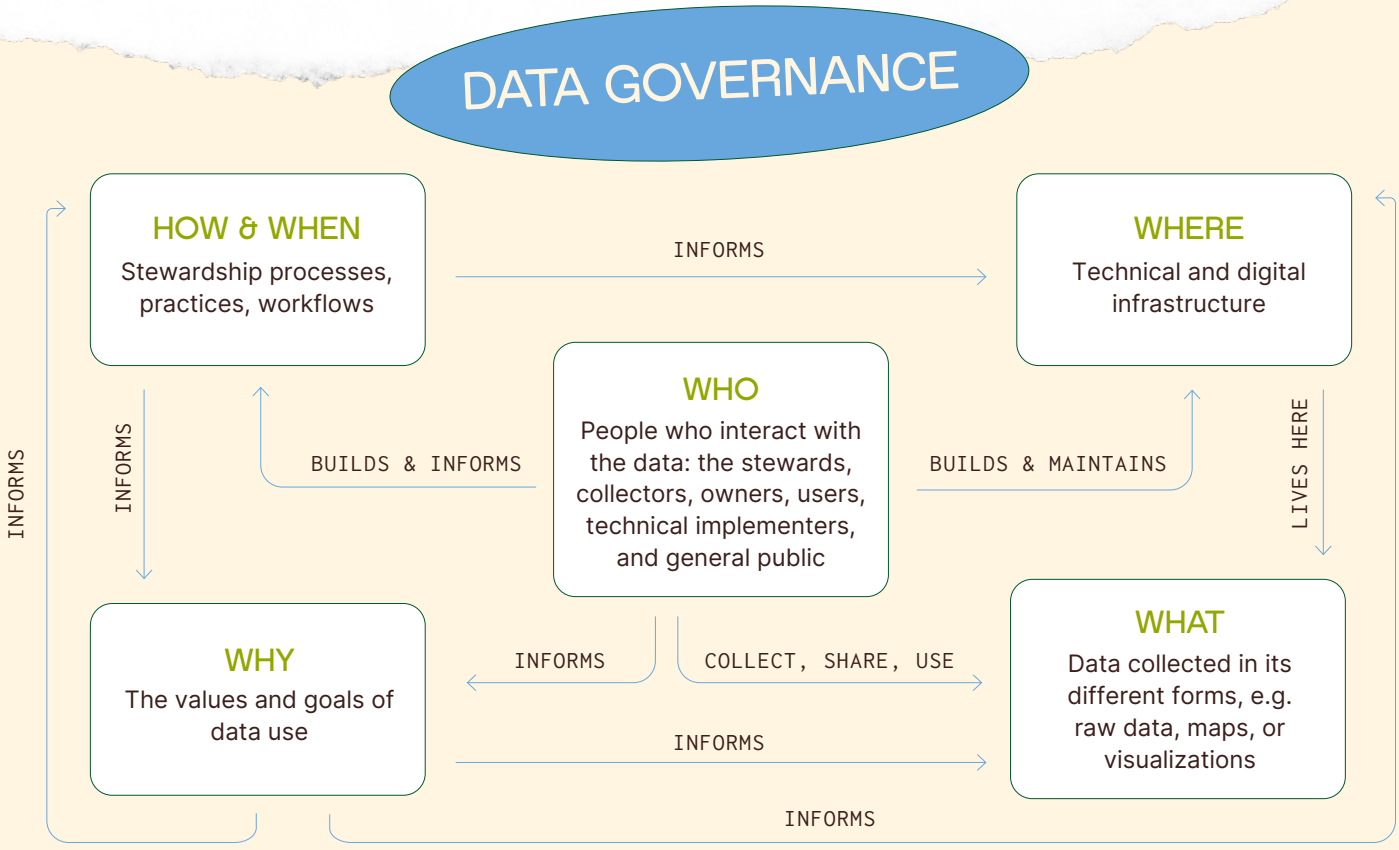
- Technical solutions that are cost-prohibitive or unsuited to social context and need
- Heightened risk of the misuse of sensitive information and lack of legal protection
- Complex local, regional, and federal laws and policies within dense legal landscapes

To unlock the potential benefits of community-generated environmental data, better data governance approaches are needed. Data governance is the who, what, where, how, when, and why of using, storing, and sharing data. Each of these aspects inform and interact with the others in differing ways, as visualized in the figure below. When appropriately designed, data governance knits together the social and technical infrastructure so as to maximize data's value and ensure its security.

We also acknowledge that data governance can seem like a difficult or nebulous concept. In conversations we've had, we often observe an assumption that you need an understanding of data science to employ data governance practices, but this isn't the case. With this playbook—and through accessible exploration of the governance methods that can make data collection, sharing, and use more attainable for environmental data stewards of all levels of expertise—we aim to reduce the barrier to entry of having these conversations and doing this critical work.

While data governance approaches across sectors continue to evolve, there is a dearth of policies and protections that are specific to community-generated environmental data. Direct replications of data governance cannot be neatly inserted into many of these contexts, since priorities and risks differ across regions, contexts, sectors, and populations. Environmental data can require specific protections:

- If certain information about a community's air or water quality becomes public, home values may plummet or insurance rates may skyrocket.
- When environmental data collection includes personally identifiable information, environmental and climate activists, and especially immigrants or people of color living in frontline communities, run the risk of higher levels of surveillance from government actors, police, and corporations.



- Researchers and universities may be interested in utilizing data collected from communities in ways that do not serve communities' data use values or purposes.

There is a tension between communities' needs to be represented through data in order to inform and advocate for themselves on the one hand, and the need to protect sensitive information on the other. This tension, along with context-specific needs, can be addressed with an orientation to data governance that can identify solutions along a spectrum of openness, rather than pushing communities to be fully open or closed with their information. This spectrum allows for environmental data governance systems to be adaptable to shifting political and environmental landscapes, and to new opportunities and evolving risks. Appropriately designed environmental data governance can support communities as they navigate these shifts.

OUR PROCESS

At Open Environmental Data Project (OEDP), we wanted to examine these complexities, ground truth our working assumptions about environmental data governance, and learn from and support communities generating environmental data. In our examination, we asked:

- What are communities' challenges?
- What are their needs to ensure accessibility and usability of their data?
- What trends exist across the landscape, and what solutions can be adapted and scaled?

To follow these lines of inquiry, we focused on two activities: 1. conducting a landscape analysis of environmental data governance alongside an advisory group of topical experts, and 2. collaborating with three community partners working with environmental data to examine their governance challenges and co-design solutions.

ADVISORY GROUP SESSIONS

JUNE-SEPTEMBER 2023

OEDP assembled a team of researchers and practitioners with broad-ranging expertise in data governance and infrastructures, community knowledge, environmental justice, and open source tools and practices. Through bi-weekly meetings, we examined data governance scenarios in various environmental and social contexts, discussing topics including: community buy-in, participatory design, data protection, building trust, and project sustainability. These sessions also supported the strategic design of our Community Data Hubs Co-Design Workshops.[1]

COMMUNITY DATA HUBS CO-DESIGN WORKSHOPS

DECEMBER 2023 -JULY 2024

OEDP launched an open call for participation in a workshopping space for community partners to work through environmental data governance questions and challenges. We partnered with three communities that were collecting environmental data, each situated in a distinct social and environmental context, and conducted workshops to identify and co-design governance solutions. Emelia Williams of OEDP worked alongside Kate Wing from Intertidal Agency ↗ to design and facilitate planning meetings, workshops, resources, and debriefs with each community partner. You can find full documentation of this process here ↗, and in-depth accounts of our community partnerships in SECTION 1 of this Playbook.

1. The members of the Advisory Group are listed on our Acknowledgements page.

HOW TO USE THIS PLAYBOOK

The Playbook is designed for anyone who is collecting and using environmental data at a community level, anyone who hopes to collect or use environmental data, and all of the individuals and organizations who act as intermediaries, interacting with those collecting and using this data. We designed this playbook to speak to different kinds of readers, encompassing many identities and affiliations, some overlapping, as they all have a role to play in data stewardship and governance. Those identities and affiliations include, but are not limited to:

- Community or citizen science initiatives
- Community-based organizations focused on addressing environmental, health, or other social issues
- Environmental and climate justice activists
- Organizations and individuals in the open source community
- Academics, researchers, or those affiliated with research institutions
- Government agencies that have a vested interest in data and evidence

Readers interested in explorations of organizations working through data governance and stewardship challenges and building processes and tools, see SECTION 1: COMMUNITY CASE STUDIES. This section includes a discussion of the environmental justice (EJ) concerns in each area, the data each partner collects, and their current governance and management systems. It also includes our workshop process and major takeaways from each case.

Readers interested in data governance and stewardship templates, see our APPENDIX. Throughout our workshop process, we derived landscape-level takeaways and created resources that can be adapted and used by others collecting and using community environmental data. These resources can be found in the Appendix.

Readers interested in relational and technical strategies to employ as a data steward or as an intermediary supporter, see SECTION 2: A PLAYBOOK TO ADVANCE ENVIRONMENTAL DATA GOVERNANCE. In Section 2, we synthesize takeaways from our partnerships and our broader research, and link them to specific "plays," or modular strategies and opportunities that data stewards can use to employ better data governance. These plays demonstrate how networked data streams can preserve important local values while enabling timely, relevant, and trustworthy decision-making—both by and with community members. Additionally, we provide support plays for intermediary organizations to work with community data stewards as they build social and technical infrastructure for their environmental data.

Readers interested in a deeper dive of this project's conceptual underpinning and what we've learned, see SECTION 3: ASSESSING THE COMMUNITY DATA HUBS MODEL. In this section, we discuss the assumptions that were affirmed or invalidated from our original Community Data Hubs (CDH) model, and provide opportunities to explore the efficacy of this model in the future. Ultimately, our hope with this playbook is that we provide ample space for learning and exploration of the current and future landscape of environmental data governance.

SECTION ONE

COMMUNITY CASE STUDIES

The following three case studies delve into the collaborations we had with our three community partners: Self-Help Enterprises in California's Central Valley, Community Health Aligning Revitalization Resilience and Sustainability (CHARRS) in Atlanta, Georgia, and CoAct in Buenos Aires, Argentina. Our role was to facilitate conversations on their data governance challenges, and to co-design resources that address those challenges. We approached these conversations as learning and sharing opportunities between facilitators and participants, and planned each workshop with elements of design thinking[2] and active listening. Each of these partnerships represent different types of environmental data, governance frameworks, capacity and funding resources, and contexts, both geographically and socially. See the next page for a breakdown of each case study and their varying contexts.

2. Design thinking is a methodology for problem solving creatively. It's often process-based and centers around understanding users, refining problems, and challenging assumptions. Read more here. ↗

MORE ON THE CASE STUDIES

	SELF-HELP ENTERPRISES	CHARRS	COACT
GEOGRAPHIC SCOPE	San Joaquin Valley, California	Atlanta, Georgia	Buenos Aires, Argentina
TYPE OF DATA COLLECTED	Water level and water quality data from private wells	Air quality data	Observational data about water quality, biodiversity, and redevelopment
TYPE OF ORGANIZATION	Medium-sized non-profit	Small non-profit	Collaboration between university researchers and legal non-profit
CHALLENGES / GOALS	• Building trust with communities • Standardizing co-ownership with participants • Sharing data responsibly with government, researchers, and the public	• Opening up access safely and effectively • Sharing data responsibly with government, researchers, and the public • Interoperability and data analysis	• Communicating data governance with non-academic partners and the public • Reach agreement between diverse stakeholders • Bolster their data governance framework to withstand changing political or social conditions

SELF-HELP ENTERPRISES
Visalia, California

CHARRS
Atlanta, Georgia

COACT
Riachuelo de la Matanza, Argentina

SELF-HELP ENTERPRISES

Self-Help Enterprises (SHE ↗) is a community development organization located in Visalia, California, whose mission is to work together with low-income families in the Central Valley to build and sustain healthy homes and communities. We worked specifically with their Emergency Services team ↗, who collect and use data sourced from domestic well assessments at private properties. They use these data to identify issues related to water contamination and water access and determine who might be in need of resources such as interim water access, well replacement, or filtration systems. While this is SHE's primary data usage, they also use the data to understand more about the groundwater supply and contamination levels in order to predict how households will be affected by agriculture and natural disasters, such as severe drought. Water supplied by most domestic wells is not regulated by government agencies in California, so the data being collected is critical for current and future planning.

GROUNDWATER CONTAMINATION AND DEPLETION IN THE SAN JOAQUIN VALLEY

SHE's service area spans across nine counties that are located in the San Joaquin Valley, one of the most productive agricultural areas in the world. This level of agricultural productivity has played a role in both depleting and contaminating the groundwater supply. Chronic overpumping of groundwater has resulted in dry wells and threatens long-term reserves.[3] Excess manure and fertilizer runoff adversely affect the quality of the groundwater that remains, causing algae blooms and contaminating well systems with nitrates, arsenic, coliform bacteria, pesticides and disinfectant byproducts.[4]

Nitrates are the most prevalent chemical found; "one in every 10 water samples collected from 20,000 wells in the Tulare Lake Basin and the Salinas Valley exceeded the drinking water standard for nitrate in 2012."[5] Exposure to nitrates over long periods of time can lead to dangerous health conditions, including serious reproductive health issues and many types of cancers.[6]

CLICK OR SCAN THE CODE HERE TO HEAR MORE ABOUT ENVIRONMENTAL ISSUES IN THE SAN JOAQUIIN VALLEY

3. *Water Use in California's Agriculture*. Public Policy Institute of California. Access here ↗.
4. *Nitrate Contamination in San Joaquin Valley, California. National Partnership for Women & Families.* Access here ↗. *Water Use in California's Agriculture. Public Policy Institute of California.* Access here ↗.
5. *After the deluge: Floods may taint more drinking water in California*. Cal Matters. Access here ↗.
6. *Water & Health in the Valley: Nitrate Contamination of Drinking Water and the Health of San Joaquin Valley Residents*. Community Water Center. Access here ↗.

CHALLENGES

Self-Help Enterprises described several challenges surrounding the use, sharing, and general governance of these data. They feel that there is a disconnect between government and private homeowners: homeowners don't want their water usage being monitored by the government, yet the data that is being collected to understand the environmental conditions of communities can inform policy and funding decisions. SHE acts as a middleman, with the goal of using this data to provide services to community members and share data in a protected and controlled manner with relevant government agencies. By utilizing safe data sharing practices, SHE's goal is to build trust between their organization and the communities they serve, while supporting government agencies in building trust with constituents themselves.

As a part of funding agreements with government agencies, SHE has to share certain categories of domestic well data with various entities, including government agencies and partner organizations, while maintaining the privacy and confidentiality of domestic well users. Funding agencies also require synthesized and sometimes, raw data sharing as part of the reporting requirements outlined in funding agreements. The funder uses the data both to monitor performance and to understand environmental conditions in the community and inform policy and funding decisions. SHE typically shares aggregated data with partners such as community-based organizations (CBOs) and local government agencies in the context of collaborative working groups and task forces to support shared goals of public health and safety (this is not necessarily a requirement, but an expectation given their role in the domestic well space). SHE co-owns this data with the property owners, and they want to build out policies that standardize what co-ownership looks like in principle and practice.

SELF-HELP ENTERPRISES COLLECTS...

SHE collects, uses, and manages many different types of data, including:

Water quality data from wells during site assessments

Standing water data (depth to groundwater) data and **well depth** from wells during site assessments, measured at a specific point in time

Real time **groundwater levels** post-new well drilling

Isotope testing to track the levels and movement of contamination

The team at SHE is committed to "doing good" with the co-owned data; they understand that once data leaves their hands, there is almost no way of knowing how that data is going to be used. With this understanding, they do want to share data with the broader community and region in a manner that is representative of the issues and supports advancing solutions, while protecting the individuals and communities that they work with from undue risk. Examples of risks to low-income families and communities include: depreciated home and neighborhood values, tenant displacement, reduced affordable housing, stock condemnation, and code violations. SHE has team members with extensive expertise in creating data products that de-identify or obfuscate the data, and the team wants to create heat maps that buffer individual data values into aggregated representations. They want to both act responsibly with their co-owned data and also contribute to larger scientific and research efforts. SHE has contributed to water quality and water level data collection for research projects and local planning projects on behalf of universities and local water districts. Regarding training, SHE's water quality staff undergo training through the Water Quality Association ↗ for credentialing. Because of the uniqueness of their programs, much of the standards and protocols for our data collection and program implementation have been developed in-house.

CLICK OR SCAN THE CODE HERE TO HEAR MORE ABOUT THE PROJECT CHALLENGES.

CURRENT GOVERNANCE AND MANAGEMENT

The SHE team is large—over sixty staff members are involved with managing data in some way. Currently, the data is not publicly available, but program participants can see the data collected from their own site assessments. SHE is planning to create aggregated data products in the future, but currently, there aren't any available. There are several organizations that they share data with, including the California State Water Resources Control Board, as a condition of funding. They also share with the California Department of Water Resources, to report when a domestic well goes dry, or when it is repaired or replaced, as a part of the Dry Well Reporting System ↗. In the past, they have shared data with universities conducting research on drought in the San Joaquin Valley.

GOALS

SHE isolated an overarching objective for our collaboration: explore data governance topics related to the policies and procedures of building a water quality data access and storage system. They wanted to accomplish three main goals within that objective:

1) strengthen agreements with participants about data sharing,
2) create language about "How We Share Data" and SHE's core data practices, and
3) leave the workshop with a clear understanding and policy direction on data use and desired data products.

The team chose these goals with some long-term outcomes in mind, including stewarding a dataset that can be shared statewide and managed responsibly in order to prepare for disasters and protect the most vulnerable populations. The team viewed this work as a preparedness effort and a community risk prevention effort. Ultimately, they also wanted to be able to save community members from risky situations by recommending areas where it's not sustainable to build houses.

IN THE WORKSHOP

We designed the workshop to cover three modules: 1) creating foundational documents to support organizational culture for data sharing, 2) a data sharing agreement building exercise, and 3) supporting data products. We had a half-day in-person workshop to run through these modules, in which twelve members of the SHE team participated.

MODULE #1

CREATING FOUNDATIONAL DOCUMENTS TO SUPPORT ORGANIZATIONAL CULTURE FOR DATA SHARING

The first module reflected SHE's need to create some organizational language related to its values and practices on how it shares data. This language could be referenced in the future when entering into partnership with a new data sharing party, collecting a new type of data, or creating a new tool or policy. We facilitated a conversation that brainstormed and surfaced the team's values on data usage and sharing, with the goal of setting up a "How We Share Data" document for internal and external use. We pulled together organizational values from their mission and the planning session, and asked participants to add other aspects from their established workflows and organizational working culture. See FIGURES 1 and 2 for the notes we documented from the session.

The team generated a robust list of values that are derived from their practices and that actively inform their approach to their work. Some of the values reflected their overall ethos and understanding of what their community experiences, with participants underscoring the maxim that water is essential to life, and that they try to get to root causes to address multiple issues, rather than just treat the symptoms of a problem. These values reflected the team's positionality as a trusted liaison between communities, and government agencies, and how they navigated these relationships with deliberation and care.

Building on those values, we posed two additional questions for their How We Share Data document: 1) How does data sharing support your mission and program efforts? and 2) How do you plan to use the data? Opening this up to discussion, a response to the first question was clear: data sharing supports connecting participants to resources (i.e., interim water access, and technical and funding assistance), which supports the protection of water, which is essential to life. In discussing the second question, the team delineated how they would use data based on a spectrum of data types, from Personal Identifiable Information (PII) and contact information, to individual and property level data, to aggregated forms of data.[7] See FIGURE 3 for this spectrum. This allowed the team to understand and articulate current and future data uses. We used this conversation to build a How We Share Data document template (RESOURCE 1 in the Appendix) that SHE can adapt and share with future external partners.

TAKEAWAY #1

Data values statements are vital to create a basis for decision-making and team cohesion.

Values are the foundation for collaboration, and data governance relies on collaboration, both internally and externally. Our work with SHE once again demonstrated this, and creating a data values statement, like a "Why and How We Share Data"

FIGURE 1

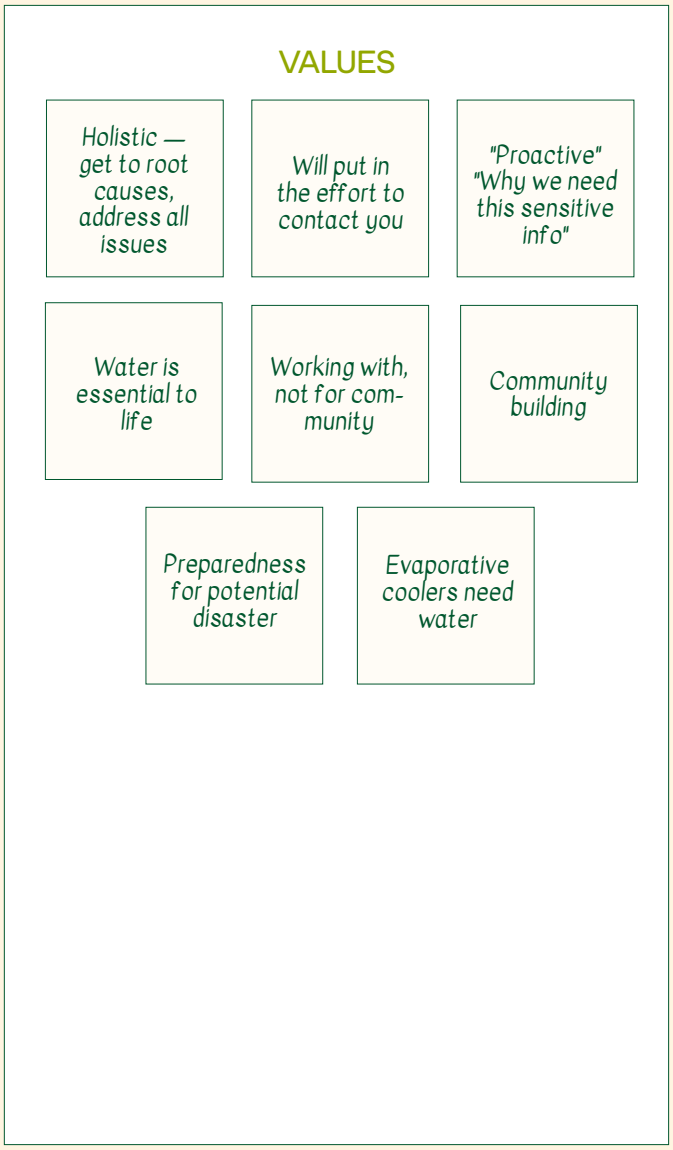


FIGURE 2

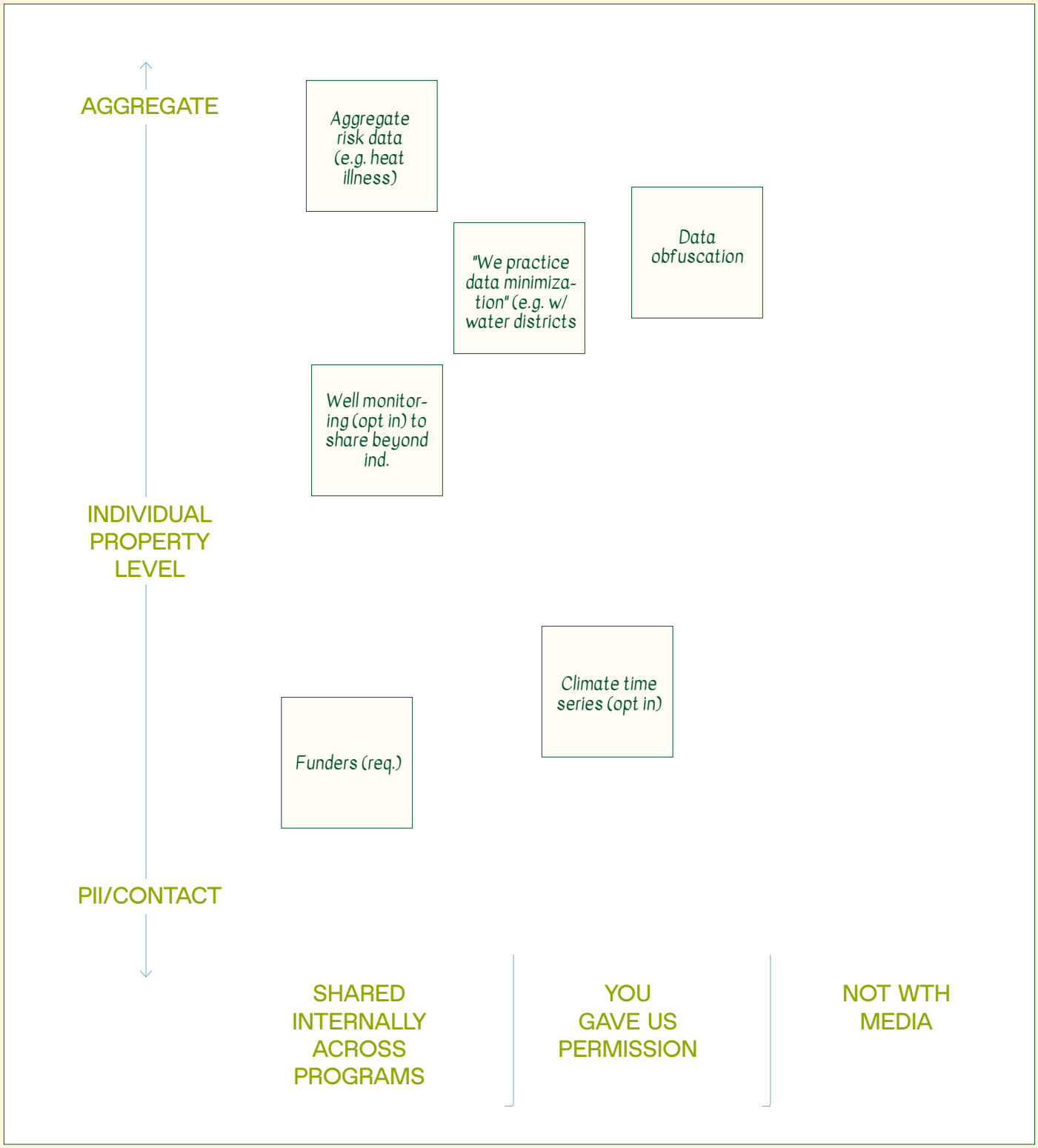

AGGREGATE
INDIVIDUAL PROPERTY LEVEL
PII/CONTACT
Aggregate risk data (e.g. heat illness)
"We practice data minimiza-tion" (e.g. w/ water districts
Data obfuscation
Well monitor-ing (opt in) to share beyond ind.
Climate time series (opt in)
Funders (req.)
SHARED INTERNALLY ACROSS PROGRAMS
YOU GAVE US PERMISSION
NOT WTH MEDIA

FIGURE 3

document is one way to establish a formal recognition of these values. Just as an organization's mission and vision guide their purpose and align the team's actions, a set of data use values can provide a strategic direction for an organization that relies on using and sharing data to fulfill its goals. Internally, it creates a baseline for how to create data workflows, products, and infrastructures; externally, it sets expectations of others and serves as a guide to organizational commitments.

A data values statement can serve as an explicit call to bring ethics into organizational practices and programs that may otherwise not see the connection between their activities, data, and goals. The team at SHE recognizes the value of their data, both to execute their mission and serve their program participants; it also wants to send a signal to bad actors who seek to use this important information to serve their own purposes at the expense of communities. Data values statements, when instrumentalized by legal and technical protections, can draw a line that demarcates who can and will benefit from this data, what will be done with it, and how it will be managed throughout the data lifecycle.

MODULE #2

BUILDING A DATA CO-OWNERSHIP AGREEMENT

In the second module, we focused on building out a data co-ownership agreement template to bolster current workflows at SHE. The Emergency Services team's participation agreement does not currently include any questions related to data sharing. As such, the goal of this session was to draft 1) the structure of a data co-ownership document for participants in the well site monitoring program, and 2) create a list of considerations for data sharing for another program of their choosing.

We opened the discussion by asking team members about the agreements they'd like to make with participants as co-owners of their site data, and we posed questions to spark ideation. These questions are listed in FIGURE 4, and include queries about access, condition of deletion, sharing, and usage. This conversation highlighted what the team valued most in collaborations with program participants, including consent and transparency. For example, there are certain reporting requirements imposed by agencies and funders that the team feels participants need to know up front.

We then analyzed data sharing considerations for SHE's Tank and Hauled Water Program ↗. This program temporarily supplies water to participants whose wells are dry. We mapped the general flow of data and information shared between participants, delivery vendors, funders, and SHE. We diagrammed a detailed data ecosystem, and, after the workshop, formalized the ecosystem graphically, including possibilities for how and where a customer portal could be used to support two-way data exchange. We also listed questions and considerations for each stakeholder, including discussion topics for contracts, data sharing, and funding agreements. Notes from this work are documented in FIGURES 5 and 6. This data ecosystem and accompanying information can be found as RESOURCE 3 in the Appendix.

TAKEAWAY #2

A strong connection between social, legal, and technical components is key to cohesive data governance.

Oftentimes, environmental data are collected, used, and managed without accounting for the entire data lifecycle and the corresponding technical, legal,

WHAT ARE THE AGREEMENTS YOU'D LIKE TO MAKE WITH PARTICIPANTS AS CO-OWNERS OF THEIR SITE DATA?

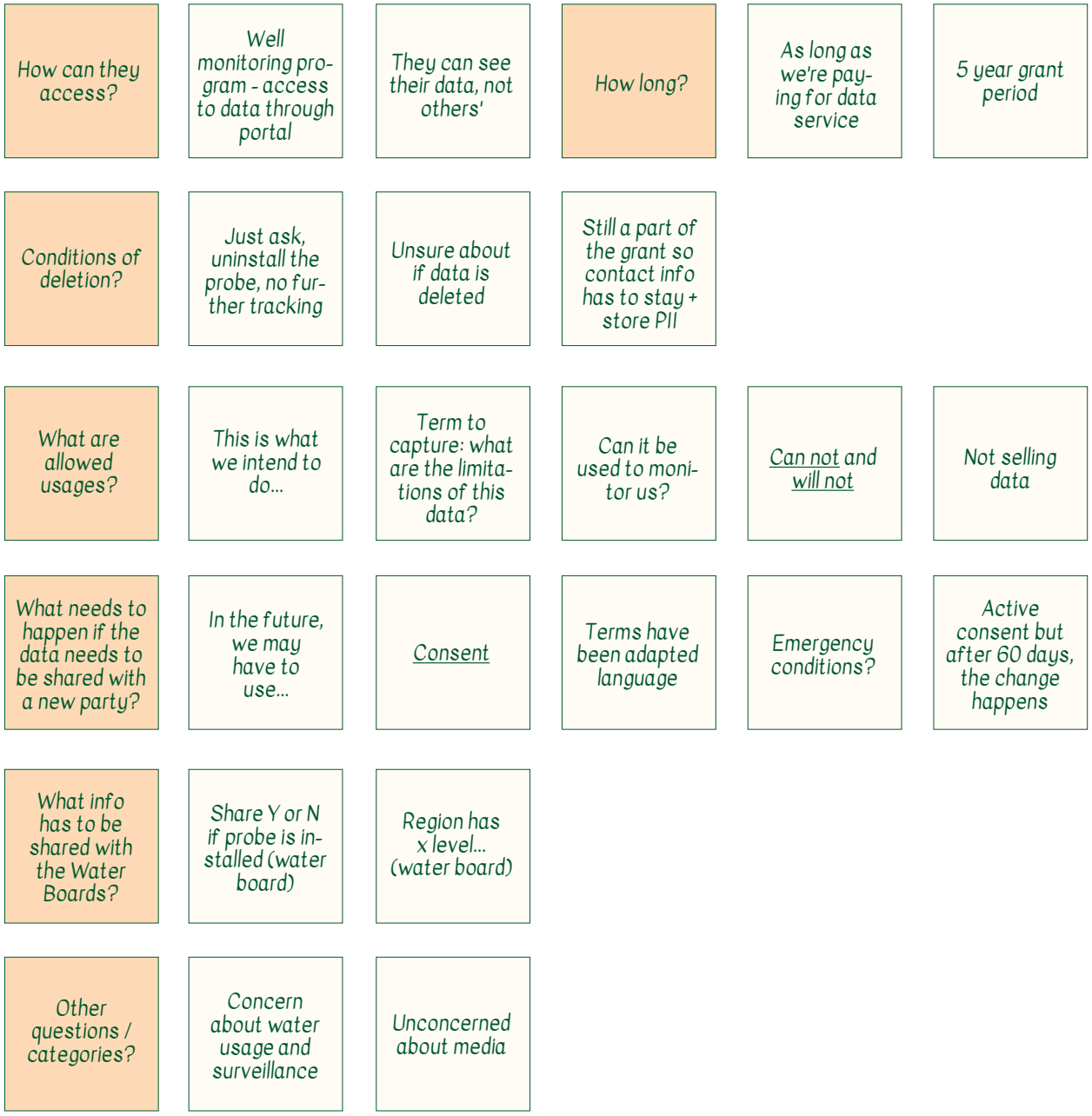


FIGURE 4

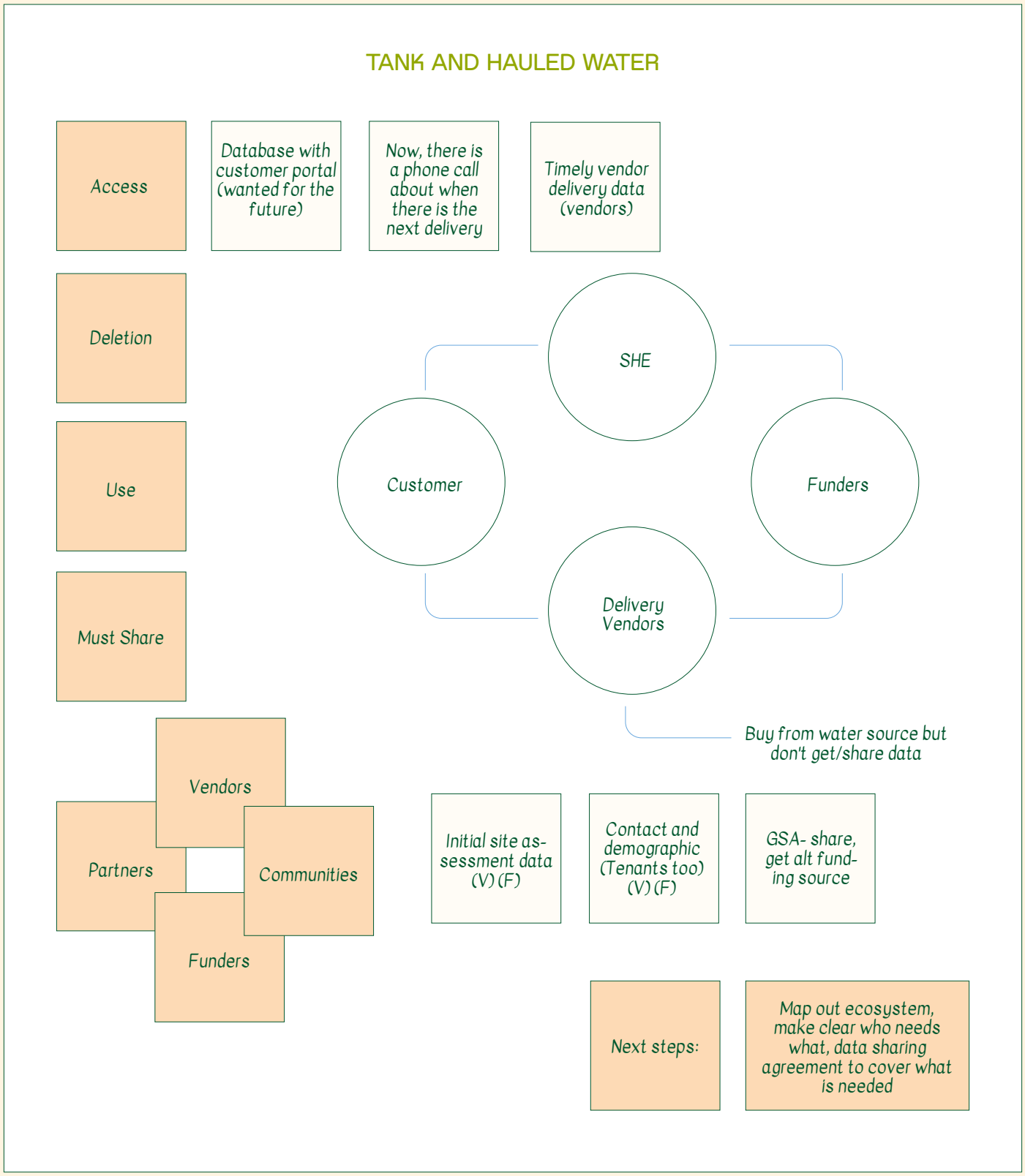

TANK AND HAULED WATER
Access
Database with customer portal (wanted for the future)
Now, there is a phone call about when there is the next delivery
Timely vendor delivery data (vendors)
Deletion
Use
Must Share
SHE
Customer
Funders
Delivery Vendors
Buy from water source but don't get/share data
Vendors
Partners
Communities
Funders
Initial site assessment data (V) (F)
Contact and demographic (Tenants too) (V) (F)
GSA- share, get alt funding source
Next steps:
Map out ecosystem, make clear who needs what, data sharing agreement to cover what is needed

FIGURE 5

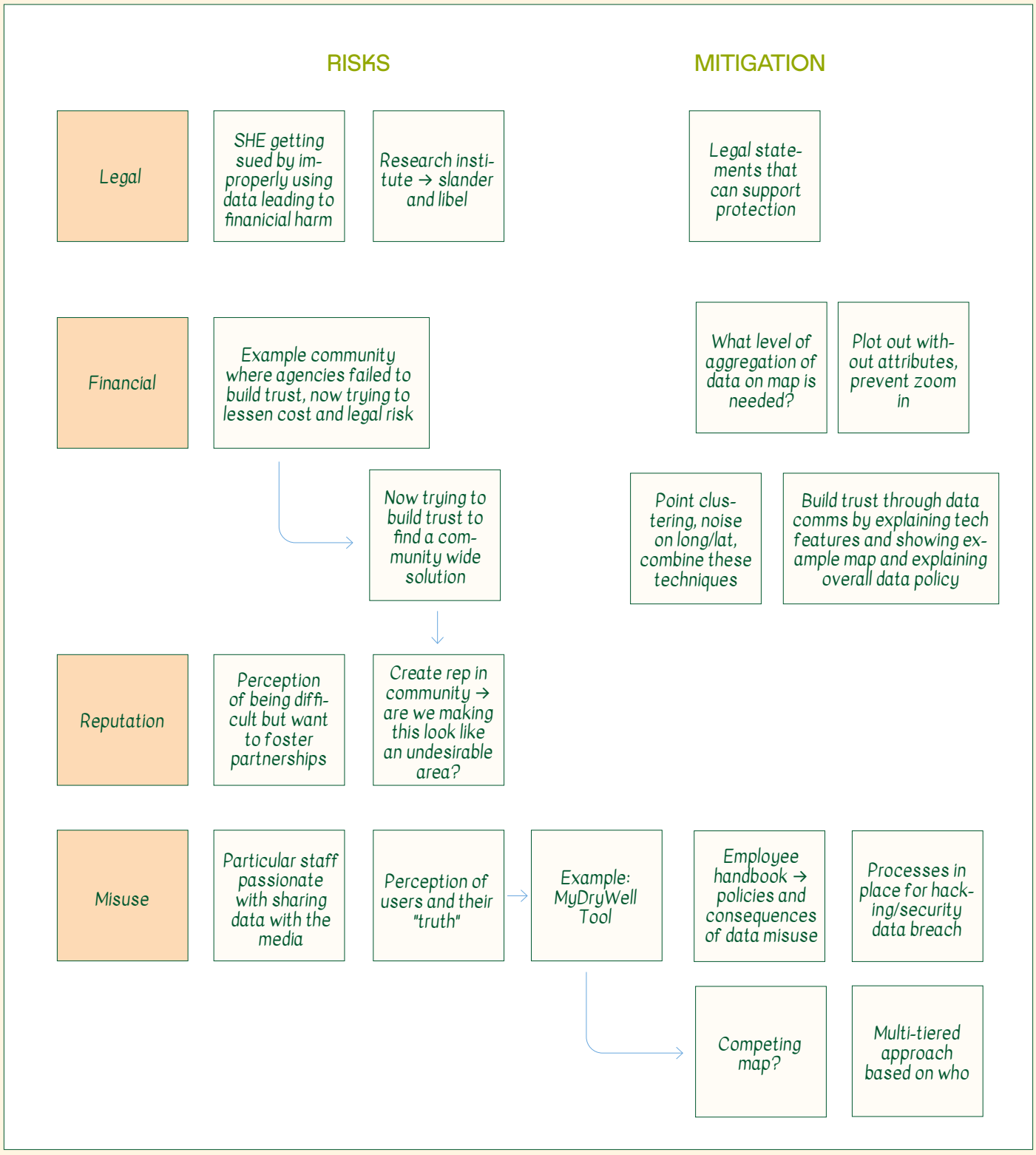

RISKS
MITIGATION
Legal
SHE getting sued by improperly using data leading to finanicial harm
Research institute → slander and libel
Legal statements that can support protection
Financial
Example community where agencies failed to build trust, now trying to lessen cost and legal risk
What level of aggregation of data on map is needed?
Plot out without attributes, prevent zoom in
Now trying to build trust to find a community wide solution
Point clustering, noise on long/lat, combine these techniques
Build trust through data comms by explaining tech features and showing example map and explaining overall data policy
Reputation
Perception of being difficult but want to foster partnerships
Create rep in community → are we making this look like an undesirable area?
Misuse
Particular staff passionate with sharing data with the media
Perception of users and their "truth"
Example: MyDryWell Tool
Employee handbook → policies and consequences of data misuse
Processes in place for hacking/security data breach
Competing map?
Multi-tiered approach based on who

FIGURE 6

and social infrastructures. Organizations may have a stellar technical database, but the management policies and legal infrastructures could be lacking, or vice versa. In the case of SHE, the team has set up a robust technical infrastructure where data is safely held, but has yet to build out a larger data governance framework for legal mechanisms for sharing (e.g., data sharing or ownership agreements). They have certain policies for understanding if participants want their data to be shared with media, but lack other policies detailing explicit permissions and uses of the data.

This presents an opportunity to build out the formal data governance framework for their organization using the resources we co-designed in the workshop. SHE could further codify their current workflows, detail internal permissions and roles in data management, and solidify sharing processes. The social, legal, and technical components could be designed to reinforce one another and systematically serve the ultimate goals of the data collection and use.

MODULE #3

SUPPORTING DATA PRODUCTS

The third and final module focused on data products and risk. SHE is in the process of creating a data repository for data from site well assessments. They are currently working to anonymize and de-identify the data, but want to create products that aggregate the data, such as a heat map that could represent the data and account for risk.

The team identified several data products that they'd like to create, including dashboards of water delivery schedules and programmatic milestones for program participants, ArcGIS maps with water access and water quality layers, and heatmaps that give information on dry wells at a neighborhood-level. These kinds of products would be beneficial to participants within the community and, potentially, other communities within the broader Central Valley.

One product in particular would feed the well information into a database and generate customized, property-level reports on water quality and quantity, as well as environmental risk. Different types of risk can be associated with this data product, including legal, financial, reputational, and misinterpretation risks. Based on identified risks, we brainstormed potential mitigation measures through technical design or other approaches. Our aim was not to scare the team by presenting all potential risks, but to build the muscle of recognizing risk and factoring in mitigation methods as data products are created, not afterwards.

The risk-related themes that emerged were largely related to reputation and misuse: Self-Help Enterprises wants to uphold its long-held reputation as a trusted partner that works with communities to find solutions on issues, spanning from water to housing. Their work relies on that trust and those relationships. Misuse was also an area of heightened potential for risk, especially related to press coverage and research partnerships; these stakeholders don't have the same social bonds with the communities, and aren't accountable to them in the ways Self-Help Enterprises is. Several mitigation strategies were identified, ranging from legal statements and contracts to technical workarounds that obfuscate data and internal policy changes that address data misuse and security breaches.

TAKEAWAY #3

Risk exists, but deliberately designed data governance can create pathways for opening up data in responsible ways.

There is a common phrase in data governance that one can create a governance framework where data is "as open as possible and as closed as necessary."[8] At OEDP, we see openness as a spectrum, and that the binary of open/closed data does not respect the nuance of many environmental data uses and contexts. SHE recognizes the need to protect their program participants while also using this data to advance both immediate services as well as strategic and holistic solutions for the Central Valley that address the root causes of water scarcity and contamination.

RESOURCES CREATED, NEXT STEPS, AND OUTCOMES

After the workshop, we created a set of templates to support SHE's data governance work moving forward. The templates included a data values statement (i.e., a document describing why and how SHE shares data) and a data co-ownership document to share with participants. We also shared a workbook featuring a diagram of the hauled water program and data questions and considerations for each stakeholder, such as discussion topics for contracts and data sharing and funding agreements. These resources are included in the APPENDIX (Resources 1-3) for others to adapt and reuse.

With each community partner, we held check-ins at the three and six month marks after our official collaboration ended. These check-ins were designed to report on progress and discuss emergent questions. SHE reported that they will be integrating their data values statement into their organization's five-year strategic planning in 2025, and they are interested in developing a broader strategy for data governance across the organization during the upcoming strategic planning process, as well. They also reported some emerging ideas regarding integrating climate change data into their work as affordable housing developers.

8. *Data Collaboratives with the Use of Decentralised Learning.* 2023 ACM Conference on Fairness, Accountability, and Transparency. Access here ↗.

CHARRS

COMMUNITY HEALTH ALIGNING REVITALIZATION, RESILIENCE AND SUSTAINABILITY

Located in Atlanta, Georgia, Community Health Aligning Revitalization Resilience & Sustainability (CHARRS ↗) is an organization that examines the social determinants of health and their impact on African American and other underserved communities in order to implement solutions to the inequities and injustices associated with them. CHAARS is currently collecting hyper-local air quality data using reference and handheld monitors. The organization is developing and involved with multiple projects, including PROJECT REMOVE ↗ and AQEarth ↗, across which they seek to expand data accessibility to EJ nonprofits and community members. CHARRS invited Dr. Na'Taki Osbourne-Jelks from the West Atlanta Watershed Alliance (WAWA ↗) to participate in our collaboration. Dr. Osbourne-Jelks is "working to improve the quality of life within the West Atlanta Watershed by protecting, preserving and restoring [the] community's natural resources."[9] WAWA represents African American neighborhoods in West Atlanta that are most inundated with environmental stressors and least represented at environmental decision-making tables.

9. West Atlanta Watershed Alliance. Access here ↗.

ENVIRONMENTAL JUSTICE IN NORTH AND SOUTHWEST ATLANTA

CHARRS and WAWA work largely in neighborhoods in West Atlanta, made up mostly of African Americans who have lived in this area due to histories of racial and economic segregation—in the 1950s and 60s, redlining and the placement of Interstate 20 solidified segregation, and thus, unequal opportunity and adverse living conditions.[10] Communities in this area have been battling environmental injustices since as early as 1900 when activists opposed the actions of a furnace operator who was dumping waste into Proctor Creek.[11] Today, these communities are still actively fighting for cleaner water, air, and land. A 2012 report from GreenLaw determined that a section of West Atlanta ranked highest among other pollution hotspots in Metro Atlanta, largely because it has historically contained the largest warehousing and transportation building concentration east of the Mississippi River.[12] The Chattahoochee River, the lifeblood of the city and region, runs less than a mile from many of the resulting pollution sources. Both inactive and active industrial sites across West Atlanta contribute to the release of toxic chemicals into the air and waterways, and leaching into the soil.

Another threat to Atlanta waterways comes from outdated and problem-prone sewage infrastructure. In a decades-long ordeal, overflow events occur after heavy rainfall when the volume of water flowing through the pipes (carrying sewage, rainwater drainage, and industrial wastewater) exceeds the wastewater treatment plant capacity. Some excess water must be diverted to nearby waterways or out of manholes into city

CONTINUED ON NEXT PAGE

10. A History of Environmental Justice in Georgia. Science for Georgia. Access here ↗.
11. Ibid.
12. Ibid.

CHALLENGES

CHARRS has several challenges surrounding the usage, sharing, and general governance of its data. The data CHARRS collects are safely secured, stored in CSV and XLS files, but currently closed to the public, and the datasets lack deeper analysis and context. CHARRS wants to be able to provide access to aspects of the data while ensuring that the data doesn't identify specific people or locations. CHARRS has strong relationships with local universities and other EJ organizations in the region, so it wants to safely share data with relevant collaborators, but lacks the data sharing and use agreements that might ease this sharing. During our collaboration, CHARRS had one full-time employee, thus had limited capacity to develop these resources.

The interoperability of its data (or lack thereof) poses another challenge for CHARRS. CHARRS has quantitative data from the AQEarth project and are interested in finding ways to tell a more representative narrative of pollution and its impacts. CHARRS and WAWA also noted that they want to use the data from government sources like the Center for Disease Control's (CDC) PLACES population health data ↗ and the EPA's Criteria Pollutant data ↗ to perform localized analysis alongside their own collected data on air and water quality. However, it is difficult to analyze these datasets together alongside CHARRS's collected data because of the difficulty matching up the geographic scopes, or differences in units of measurement. CHARRS's longer-term goal is to build a platform where data can be uploaded, managed, and shared among different EJ organizations, the public in Atlanta, and, eventually, all of Georgia.

CONTINUED FROM PREVIOUS PAGE

streets, contaminating them with untreated or partially treated sewage.[13] Climate change exacerbates this issue with more frequent, unpredictable, and extreme rainfall events. Activists have been organizing to address this kind of contamination since the 1990s. Today, organizations including CHARRS and WAWA are working to engage residents from these communities on these issues while also supplying them with the data and tools to elevate their voices and seek self-determined outcomes for environmental, policy, and systems change.[14]

CHARRS COLLECTS...

CHARRS collects, uses, and manages many different types of data, including:

Indoor **radon gas levels** in residences (as part of Project REMOVE)

Air quality data, including hyperlocal air monitoring data of 6 criteria pollutants (CO, PM2.5, PM1, O3, NO, NO2) (as part of AQEarth)

Black carbon data (in collaboration with Dr. Christina Fuller from the University of Georgia)

Survey data from West Atlanta residents who attended meetings (as part of Project REMOVE and AQEarth)

Data about people who bring items to repair and what is being repaired (alongside the **Atlanta Repair Cafe**)

13. *Wastewater: Chronic Sewage Spills by City of Atlanta. Chattahoochee Waterkeeper.* Access here ↗.
14. *Atlanta: Reflections on Beholding, Protecting, and Dismantling.* Villa Albertine. Access here ↗.

CURRENT GOVERNANCE AND MANAGEMENT

As mentioned above, the data from each of CHARRS's collaborations (AQEarth and Project REMOVE) is currently not public. There are slightly different governance and management structures for each project, since each includes different collaborators. For example, with Project REMOVE, which collects radon data, the metadata comes to CHARRS ready to use and is publicly shared. This data is stored by Georgia State University and is under Dr. Dajun Dai's control. AQEarth datasets are stored on Montrose Environmental's servers, on their Sensible Environmental Data Platform ↗, and are not currently open to the public. Decisions about this data are made collaboratively by the respective project teams, and CHARRS is interested in understanding what shared data ownership and preservation could look like over time.

GOALS

With CHARRS, we collectively delineated three central targets to address in the workshop:

1) Establish elements of a future data governance framework for data stored in a central repository,
2) Build data governance skills to support intentional sharing, and
3) Develop a list of data science and management questions to inform how CHARRS hires researchers and writes grants.

These modules largely focused on the data that CHARRS currently has, but also provided space to workshop potential future governance for a platform that is collaboratively managed by (or at least sourced from and used by) the broader community of EJ-focused organizations in Georgia. CHARRS recognized that it would be useful to understand how to approach questions of data governance and stewardship before a technical tool was developed.

IN THE WORKSHOP

We designed the workshop to cover three modules: 1) future sharing and data frameworks, 2) a data sharing agreement building exercise, and 3) data science and management question prioritization and preparation. We had a three-hour virtual workshop to run through these modules with Gwen Smith of CHARRS and Na'Taki Osbourne Jelks of WAWA.

MODULE #1

FUTURE SHARING

The first module identified some possible design features of a central repository that would enable CHARRS to share and co-manage data with other regional EJ organizations. Through facilitated brainstorming exercises, we identified the potential types of data, stakeholders, roles and permissions, and technical features that could support its preferred sharing methods and governance approaches.

The potential data types included data both currently managed by CHARRS and WAWA, as well as more expansive types of climate, environmental, and health data, including brownfield data,[15] heat island data, and information about local emergency and risk planning. They are interested in understanding environmental and health issues holistically, and these additional datasets would support further analysis and programming. See a full list of data types in FIGURE 7.

Potential stakeholders included state EJ groups, students enrolled in primary and secondary education, Atlanta city councilmembers, community

15. Brownfields are "underutilized properties where reuse is hindered by the actual or suspected presence of pollution." See more here ↗.

residents, and state elected officials. Stakeholder roles and permissions would depend on 1) what they would want to use data for, 2) what data they would contribute, and 3) what level of data literacy and capacity for data management they had. Our conversations around these questions made explicit some opportunities and limitations for certain groups, and also built an understanding of what types of data could be shared—and how—within this system, i.e., who would get access to all of the data, a limited portion of the data, or data products.

Through the module, we made note of particular design features that would not only accommodate the data types, stakeholders, and roles and responsibilities identified, but also respond to existing challenges in current data needs and management capacities. The following table lays out both these challenges and how specific design considerations could be implemented to address them.

CHALLENGE	SUGGESTED DESIGN FEATURE FOR FUTURE SYSTEM
State and public data is difficult to access and understand; it can be especially difficult to "match" this data with CHARRS's data.	If collaborators want to add data, contributors must agree to use specific data formats to facilitate data matching and integration.
It is currently difficult to analyze both qualitative and quantitative data side by side.	A future data system must have the ability to hold multiple data types, including photos.
Users would need different levels of access to protect community members' privacy.	Sensitive data would require access codes, and some data that includes PII would not be available to anyone outside of the project team. Some users would only be able to access data products and not raw data. A licensing scheme could also be established to allow for specific types of access or use.
There is data categorized by county and census tracts, but less so for census block groups.[16] CHARRS and WAWA need finer-grained data than what usually exists, and WAWA works on a watershed level that crosses over jurisdictional lines.	A future data system must be location-centric, and include mapping abilities with different layers that allow for analysis at the neighborhood scale.

16. The US Census Bureau reports data in differing units, including counties and census tracts. Census tracts are "statistical subdivisions of a county that aim to have roughly 4,000 inhabitants," and block groups are a subdivision of a census tract, usually containing between 250 and 550 housing units. See more on these distinctions here ↗.

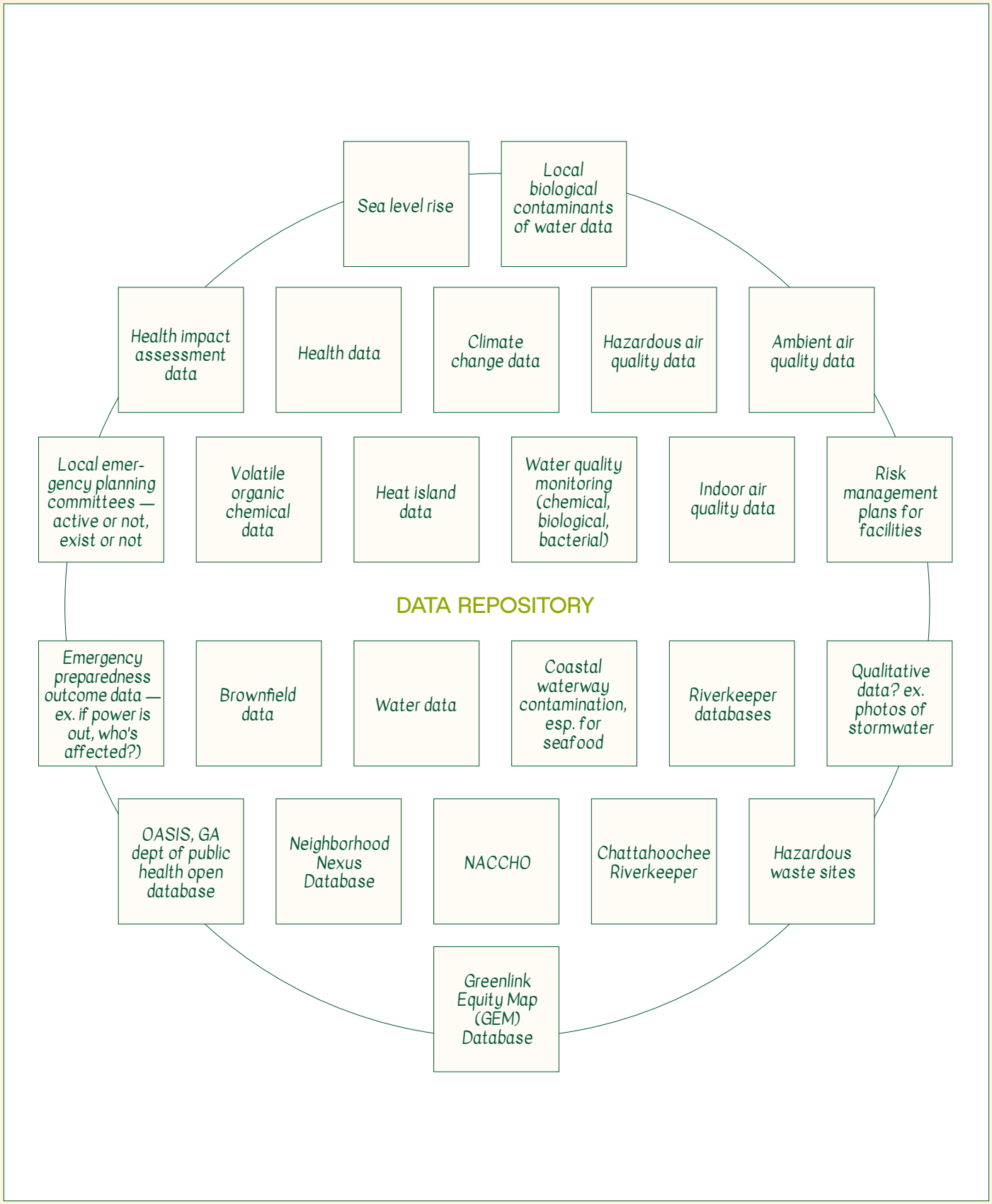

Sea level rise
Local biological contaminants of water data
Health impact assessment data
Health data
Climate change data
Hazardous air quality data
Ambient air quality data
Local emergency planning committees — active or not, exist or not
Volatile organic chemical data
Heat island data
Water quality monitoring (chemical, biological, bacterial)
Indoor air quality data
Risk management plans for facilities
DATA REPOSITORY
Emergency preparedness outcome data — ex. if power is out, who's affected?)
Brownfield data
Water data
Coastal waterway contamination, esp. for seafood
Riverkeeper databases
Qualitative data? ex. photos of stormwater
OASIS, GA dept of public health open database
Neighborhood Nexus Database
NACCHO
Chattahoochee Riverkeeper
Hazardous waste sites
Greenlink Equity Map (GEM) Database

FIGURE 7

TAKEAWAY #4

The technical design of a data system can, and should, reflect data users' needs to unlock the value of the data.

CHARRS and WAWA have worked hard to collect hyperlocal data that is critical to answering environmental and health questions in their community, yet their current digital tools don't allow them to use the data in ways they would like. Data governance is a continuous balance between the roles that people play, the actions they take, the processes they create and participate in, and the technology that supports the data throughout its lifecycle. Technical design features can either hamper effective use or enable greater control and collaboration. While it is hard to prescribe a pre-made system of governance that will cater to each community's social and technical needs, there are processes and resources that can support a community's understanding and utilization of data governance practices and frameworks. As a data steward, facilitating conversations to understand user priorities, and identifying how those priorities translate into technical design features for digital infrastructure, can address some of these challenges. Other strategies are available and outlined in SECTION TWO of this playbook.

TAKEAWAY #5

Data ownership, in practice, has material requirements; it requires somewhere to house the data and someone to maintain it.

This takeaway is essentially a direct quote from one of the participants in the first module, and its insight is clear-cut: data ownership and community control require resources and capacity to work properly.

The costs associated with building and maintaining databases or other digital infrastructures, creating data analysis tools, and enforcing data standards are often out of reach for small, community-based organizations, and require consistent, high-level fundraising to maintain. It sounds simple to "share data," but there are high costs associated with the technical tools and human resources needed to meaningfully govern data.

MODULE #2

BUILDING A DATA SHARING AGREEMENT

At the time of the workshop, CHARRS wanted to know how to responsibly and effectively share their collected data with community members, researchers, and state governments, but did not have infrastructure set up to support this. We ran through a scenario with a potential data sharing partner: a professor at Emory University who wanted to use reference monitor air quality data. We asked the following questions to develop a data sharing agreement:

- What are the main considerations for sharing with this recipient?
- Who are the parties to the agreement?
- What specific data items will be shared?
- Are there restrictions of use? What are they?
- What is the time period of access?
- How will CHARRS maintain access to the data or end access to the data?
- How will the recipient communicate how the data is being used during the sharing period?

These considerations are meant to both protect and acknowledge the work that CHARRS and others have done to collect and steward this data. CHARRS recognizes the potential for data extractivism, where

researchers or commercial entities can "extract resources like knowledge, wisdom and stories in the form of data from communities."[17] This is especially prevalent in communities facing environmental injustice, and data sharing agreements can be a strategy to bolster relational accountability, and to avoid data sharing partnerships that are exploitative and bring little positive impact to the community whose data is being utilized.

The main considerations that stemmed from this discussion were that CHARRS wanted to make sure that they received authorship credit in resulting publications and articulated how they would like to be attributed within those publications. In cases when a researcher planned to publish something using this data, CHARRS also wanted to include parameters allowing them to review drafts prior to publication to make sure sensitive data wouldn't be released. Along these lines, CHARRS was also concerned with how the data might be used. To account for this last point, we suggested that they develop a set of pre-approved purposes or use Creative Commons noncommercial licenses ↗ so that the data couldn't be used for commercial purposes.

Throughout this exercise, we emphasized that a data sharing agreement is like any other contract—the terms are up to the parties—and so we encouraged CHARRS and WAWA to be explicit about what they wanted in the agreement. For the restrictions of use, CHARRS suggested that potential users should take a data ethics training and register as a user within a future platform. This would enable CHARRS to more easily track which users have downloaded specific datasets, and when those downloads happened. CHARRS also stipulated that partnering organizations notify CHARRS if the main point of contact changed, and provide an emergency contact or access option. This consideration is critical, as project teams representing research institutions can shift based on personnel changes or funding.

17. *Beyond Extractivism in Research with Communities and Movements.* The Commons, Social Change Library. Access here ↗.

TAKEAWAY #6

Data sharing agreements are contracts. They don't build relationships, but rather maintain existing relationships.

Data sharing agreements are essential tools that community environmental data holders can utilize to 1) share their data safely and 2) get what they need from the interaction while providing what a collaborator is requesting. Still, they are simply one aspect of a larger relationship or partnership. There are often power asymmetries between two data-sharing parties—such as community-based organizations and university-affiliated researchers—that can prompt specific needs and restrictions from both parties, as with any contract. Before creating this contract, there is usually a process of communicating these needs, restrictions, and shared goals. This communication is an essential piece of relationship building, especially for community environmental data users and holders, who are often faced with large power asymmetries and are often situated in histories of data misuse and extractivism.

SYNTHESIS AND CLOSING

As we closed the workshop, CHARRS had many technical data science management questions that we couldn't answer in this session alone. We generated a list of these questions (see FIGURE 8), many of which other communities in this collaboration had asked, as well. Some notable themes that came up include how to create appropriately scoped data analysis and visualization tools, what kind of agreements or policies can ensure data ownership, and

how to build digital infrastructures and governance systems that can adapt to changing conditions and secure data in perpetuity.

RESOURCES CREATED, NEXT STEPS, AND OUTCOMES

After the workshop, we created a synthesis of next steps for CHARRS to consider and a worksheet of questions and considerations around data sharing. This worksheet included considerations about data owners' roles and responsibilities, a table with potential scenarios around who could use what and how they could use it, and other conditions CHARRS could place in a data use agreement or attach to data and data products. This resource is included in the Appendix (RESOURCE 4) for others to adapt and reuse.

After the official collaboration culminated, CHARRS, in partnership with OEDP, submitted an application for the EPA's Community Change grant program ↗, and integrated data governance workshops with the broader community into the application. With each community partner, we held check-ins at the three and six month marks after our official collaboration ended. In our check-in meetings, Gwen Smith reported that she had used the data sharing guidance derived from our workshops to support decision making when approached by a researcher who was interested in applying for a grant regarding local air quality monitoring.

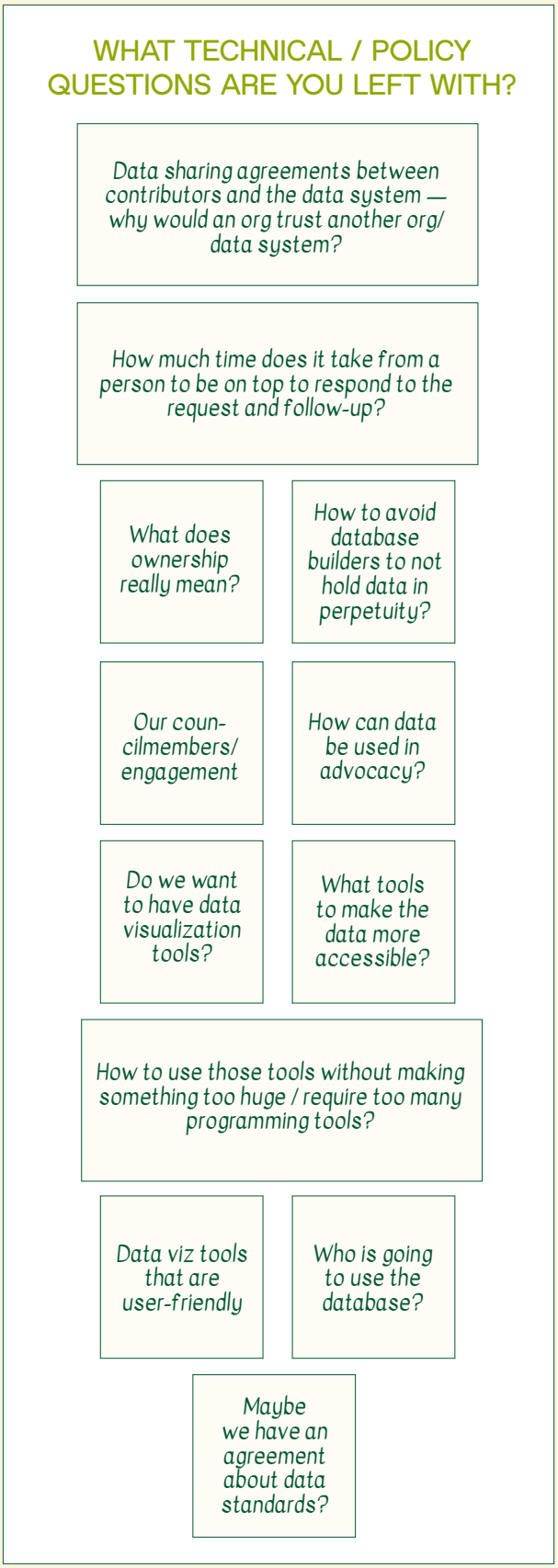


FIGURE 8

COACT

CoAct ↗ is a project team that includes El Centro de Investigaciones para la Transformación, or the Research Centre for Transformation, at the National University of San Martin (CENIT) and Fundación Ambiente Y Recursos Naturales (FARN), or the Environment and Natural Resources Foundation ↗, in English. Together, they created and now actively maintain the social citizen science tool ¿Qué pasa, Riachuelo? ↗ (QPR). This tool collects and displays data on water quality, as well as the uses and perceptions of natural areas in the Matanza-Riachuelo basin. CoAct uses participatory research and citizen social science practices[18] to integrate these stakeholders into the project at different stages, whether during design or result interpretation. CENIT and FARN aim to promote synergies between scientists and academic researchers working in the basin, communities and organizations interested in environmental education, and decision makers in the region. The research group investigates three key themes: natural areas, relocation and redevelopment, and water quality.

CONTAMINATION IN THE MATANZA-RIACHUELO BASIN

CoAct's collaborative mapping platform covers the Matanza-Riachuelo Basin, a 40-mile long river and its surrounding areas, where six million people, approximately 15% of the Argentinian population, live.[20] This river and its surroundings are heavily contaminated by multiple and diverse sources, including industrial waste, sewage effluents, open garbage dumps, and organic run-off from meat packing plants. Sewage effluents are especially common since 50% of the population in the basin is not connected to the sewage system. People near the river suffer from health issues, including severe skin rashes, vomiting, diarrhea, and headaches, as well as lead poisoning and high cancer rates.[21, 22]

The contamination became so severe that in 2008, Argentina's Supreme Court declared the river an imminent threat to those living near the water and ordered the rehabilitation and sanitation of the entire basin. A governmental authority called Matanza-Riachuelo Basin Authority (ACUMAR) was established by law in 2006 to lead the environmental and social recomposition of the river basin; this included river clean up, in addition to creating housing projects and improving the population's access to potable water, sewage services, and health care.[23] While there have been some improvements—namely installing new sewage systems, clearing garbage from the river,

CONTINUED ON NEXT PAGE

18. CoAct defines citizen social science in this way: "Citizen social science combines equal collaboration between citizen groups (co-researchers) that are sharing a social concern and academic researchers. Such an approach enables [us] to address pressing social issues from the bottom up, embedded in their social contexts, with robust research methods." Access more on their approaches and practices here ↗.
19. *Water quality and citizen social science in Buenos Aires.* CoAct Citizen Social Science. Access here ↗.
20. Ibid.
21. *Plagued with Health Issues, Residents Near Dirty River in Argentina Have Few Options.* Global Press Journal. Access here ↗.
22. *Los residentes pobres de Buenos Aires conviven con uno de los ríos más contaminados del mundo.* [*Poor residents of Buenos Aires live with one of the most polluted rivers in the world.*] Univision. Access here ↗.
23. *Life Along the Banks of One of Latin America's Most Polluted Waterways. Earth Island Journal.* Access here ↗.

CHALLENGES

CENIT and FARN described several challenges surrounding the usage, sharing, and general governance of these data during the CoAct project. The team wanted to translate the technicalities of data governance to non-academic stakeholders in the project, including the citizen scientists using the QPR tool and local and regional government decision makers. They were interested in understanding how to create the simplest way of translating the governance framework so that the public could understand and get involved. In this vein, the team was interested in exploring how to reach agreements between stakeholders with different interests or values. Considering that public policy in Argentina has been subject to drastic switches in orientation, CENIT and FARN were also interested in creating data governance structures that would not only have strong decision-making frameworks, but also be flexible enough to withstand changing conditions.[24]

The challenges that CENIT and FARN wanted to explore hit upon the idea that many data governance questions are rooted in the relational aspects of data governance. These relational aspects—the "people" side of a technical tool—include decision making structures, permissions and roles, collaborative processes, sharing and usage, and documentation.

CURRENT GOVERNANCE AND MANAGEMENT

Currently, CoAct's data is managed by FARN and stored on Amazon servers. Data is submitted by two different kinds of users: registered and non-registered. When non-registered users

24. *Ante la decisión de la Corte, se necesita coraje institucional para el Riachuelo.* [*In light of the Court's decision, institutional courage is needed for the Riachuelo.*] El Diario AR. Access here ↗.

CONTINUED FROM PREVIOUS PAGE

and shutting down abusive factories[25]—ACUMAR still struggles with appropriately allocating funds and government capacity to address some of the root causes and create holistic solutions on a basin-wide level.[26]

For example, the public information system is often considered insufficient in relaying project-level updates to the public, although it has published reports about policy implementation and regularly made datasets public in the past. Additionally, biologists have pointed out that some of ACUMAR's superficial changes, like shoreline parks and public spaces, distract from the larger ecological crisis basin-wide.[26] ACUMAR is understandably overburdened regarding capacity: the authority has to address and manage inter-jurisdictional struggles amongst three levels of government - 14 local governments, 2 state governments, and the federal level authorities. Currently, many are concerned about continued funding considering Argentinian president Javier Milei's approach to cutting what he considers "non-essential spending."[27] Moreover, at the end of 2024, the Supreme Court abandoned the supervision of the sanitation plan.[28]

25. *Riachuelo: Cleaning Up Argentina's Most Polluted River.* We Build Value Magazine. Access here ↗.
26. *Is There Still Hope for Latin America's Most Toxic River?* Pulitzer Center. Access here ↗.
27. Ibid.
28. Ibid.
29. Ante la decisión de la Corte, se necesita coraje institucional para el Riachuelo. [In light of the Court's decision, institutional courage is needed for the Riachuelo.] El Diario AR. Access here ↗.

submit data, they fill out a structured form with closed-ended questions (i.e., single-choice, multiple-choice, or dichotomous), which does not have to be reviewed by FARN. Registered users submit data using a form that includes open-ended questions and the possibility to share documents, links, and images, in addition to structured questions. These open fields are reviewed by FARN (with potential collaboration from the CoAct coordinating team) to check that the content is aligned with the platform's policy for conditions of use. When reports reach a threshold amount, they are downloaded through the platform's dashboard, curated by the CoAct team to ensure that files are accessible, and uploaded to Zenodo with a CC-BY-SA license ↗.[30] CoAct's Zenodo ↗ also hosts the metadata and more detailed information on the metadata, tutorials on how to upload open and sensitive data, and the datasets themselves are available for download.

As for social governance, there are many groups currently involved in the project, including CENIT, FARN, and community library workshop organizers participants (such as neighbors and school teachers). They also have occasional alliances with training institutions for school professors and teachers, local boards of historical studies, and local networks of environmental activism groups. As the coordinating institutions of the CoAct project, CENIT and FARN currently make the decisions, but they hope to extend decision making power to the

COACT COLLECTS...

The CoAct project collects observational water quality data and data about natural areas' uses and perceptions in order to investigate three key themes: natural areas, resettlement and redevelopment, and water quality.

Data about **natural areas**: Wetlands and green spaces are threatened by contamination, industrialization, and urbanization. The QPR platform enables the sharing of information about the biodiversity of those spaces and the threats they face.

Data about **relocation and redevelopment**: Relocation and reurbanization processes are a part of ACUMAR plans, and with QPR, communities can monitor different aspects of the process and share their experiences of participation in government programs.[32]

Data about **water quality**: QPR facilitates community participation in generating information about observable variables on water quality that complements existing public data. This supports sanitation policy, and allows people to "observe the river, assess its biocultural importance, and monitor variables associated with water quality to support decision-making."[33]

30. CC-BY-SA is a license created by Creative Commons, enables reusers to distribute, remix, adapt, and build upon the material in any medium or format, so long as attribution is given to the creator. More on this license can be found here ↗.
31. FARN has many areas of action and study, which are organized in different departments, such as environmental policy, climate policy, biodiversity, research, legal affairs, legal advice clinic, press, and communication. The department of legal affairs is responsible for the CoAct project, but there have also been interactions between the project and the biodiversity department, as well as with the press and communications department.
32. Qué Pasa Riachuelo. Access here ↗.
33. Ibid.

biodiversity department[31] of FARN and the library workshop coordinators. Decisions are made during bimonthly meetings with the CENIT team and representatives from FARN, using Slack to track issues and GitHub to document advances in the interim.

GOALS

CoAct's main objective was to involve more people in the project and design pathways to increase engagement, whether it be basic awareness, participation in curation, or even stewardship of the data. There were three main goals within this objective:

1) Create a plain language data governance framework that could be shared with community actors,
2) Develop guidance for community actors to support their involvement in curation and stewardship of the data, and
3) Understand and apply methods to increase the diversity of participants.

These goals fed into longer term outcomes, which included increasing the number and diversity of people who are invested in the project during this time of environmental crisis. CoAct wanted to work toward a model where more community actors, including community libraries, schools, and environmental activists, could play a larger role in collecting and managing the data on the platform.

IN THE WORKSHOP

Our workshop covered two modules: 1) mapping CoACT's current data governance framework, and 2) opening up data governance engagement pathways. Five participants from the CENIT and FARN teams participated in our three-hour online workshop.

MODULE #1

MAPPING COACT'S CURRENT DATA GOVERNANCE FRAMEWORK

The first module reflected CoAct's desire to 1) engage with community actors through data governance, and 2) understand in which aspects of governance they might be interested in participating (and how). We focused on how members of CoAct currently understand their governance, who is involved, and which aspects of data governance were a priority for the project. We started this conversation by asking what data governance meant to their project team. Notable responses from the team coalesced around common themes:

- Who makes decisions about data and how
- The importance of community involvement
- Data use rules and the process to set those rules
- Data management systems and how they apply rules determined by the group
- Forms and processes for adding and managing data

It was evident that a data governance framework for CoAct should explicitly outline decision-making and rule-making processes, the maintenance activities of data management, and how the technical infrastructure could support social and data governance. Most members in attendance had a pivotal role in setting up these rules or systems, but lacked a bird's-eye view piece of documentation that could show how the relational and technical pieces come together. This conversation is detailed further in FIGURE 9.

Building on this session, we led a conversation around who was significantly involved in the project, who was involved occasionally, who CoAct might want to bring into the work in the future, and who

was considered a community actor. This session surfaced details on the project's partnerships, and which groups might be prioritized for engagement. An important specification arose during the conversation: CoAct should determine who needed to know about data governance, and who needed to participate in data governance, as this would inform CoAct's plans to engage them. The CoAct team identified teachers, environmental leaders at the provincial level, and community libraries as the primary actors for participation in governance; meanwhile, educational inspectors or administrators and students might not participate in the governance but should be kept informed of processes. For a comprehensive view of all current and future actors, see FIGURE 10. We asked five questions for each priority actor group:

- What are their goals within the project?
- What is their current level of governance awareness?
- What is their capacity for engagement?
- How would their participation benefit them?
- Are there any risks to their participation?

The answers to these questions painted a fuller picture of each actor to ultimately inform the design of effective engagement strategies. Each of these aspects would influence who, how, and why someone might want to be involved. Furthermore, designing one singular approach to engagement might appeal to one of these actors, but miss the mark for others; it would run the risk of both giving someone too little information or overwhelming them with too much; or offer benefits that don't align with an actor's desires, or involve them in a process that feels risky, either politically or legally. This would be critical in scenarios that are politically fraught, like the current situation in the Matanza-Riachuelo river basin. To see the brainstormed responses for teachers, as an example, see FIGURE 11.

TAKEAWAY #7

Because data governance can be understood and employed in many different ways, it is essential to define what your data governance framework looks like within your project team.

As facilitators, we hesitated to present a conclusive definition to our partners because data governance, and correspondingly, data governance frameworks,[34] can look differently to different communities and depends on the context and priorities of a project and team. Still, within each community or project, a commonly held understanding can support putting a data governance framework into practice. Data governance is often an emergent topic for communities working with data, who have been utilizing data governance practices or processes, but not yet using this terminology. Defining data governance with a project team can be a simple conversation among those involved. They can begin by asking what each person's definition or idea of data governance is, and then identifying the commonalities and differences, and piecing together a definition describing the team's interpretation of the term. This can facilitate the further alignment of priority data governance practices and processes, constituting the beginning of a framework.

34. We will use the term "data governance framework" in this instance to indicate the set of rules, processes, and practices that a particular organization or set of stewards utilize regarding the relational and technical aspects of their data governance.

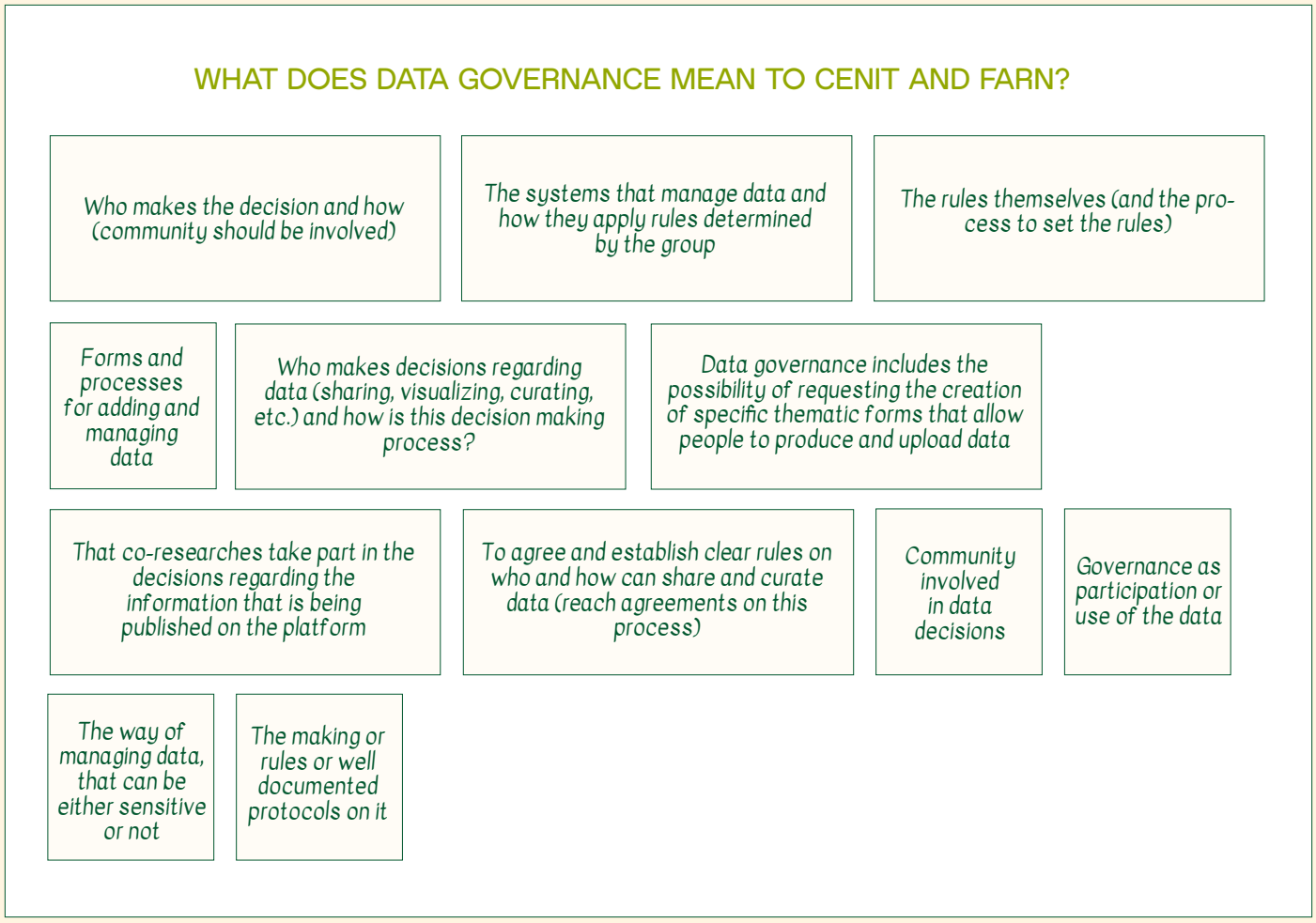

WHAT DOES DATA GOVERNANCE MEAN TO CENIT AND FARN?
Who makes the decision and how (community should be involved)
The systems that manage data and how they apply rules determined by the group
The rules themselves (and the process to set the rules)
Forms and processes for adding and managing data
Who makes decisions regarding data (sharing, visualizing, curating, etc.) and how is this decision making process?
Data governance includes the possibility of requesting the creation of specific thematic forms that allow people to produce and upload data
That co-researches take part in the decisions regarding the information that is being published on the platform
To agree and establish clear rules on who and how can share and curate data (reach agreements on this process)
Community involved in data decisions
Governance as participation or use of the data
The way of managing data, that can be either sensitive or not
The making or rules or well documented protocols on it

FIGURE 9

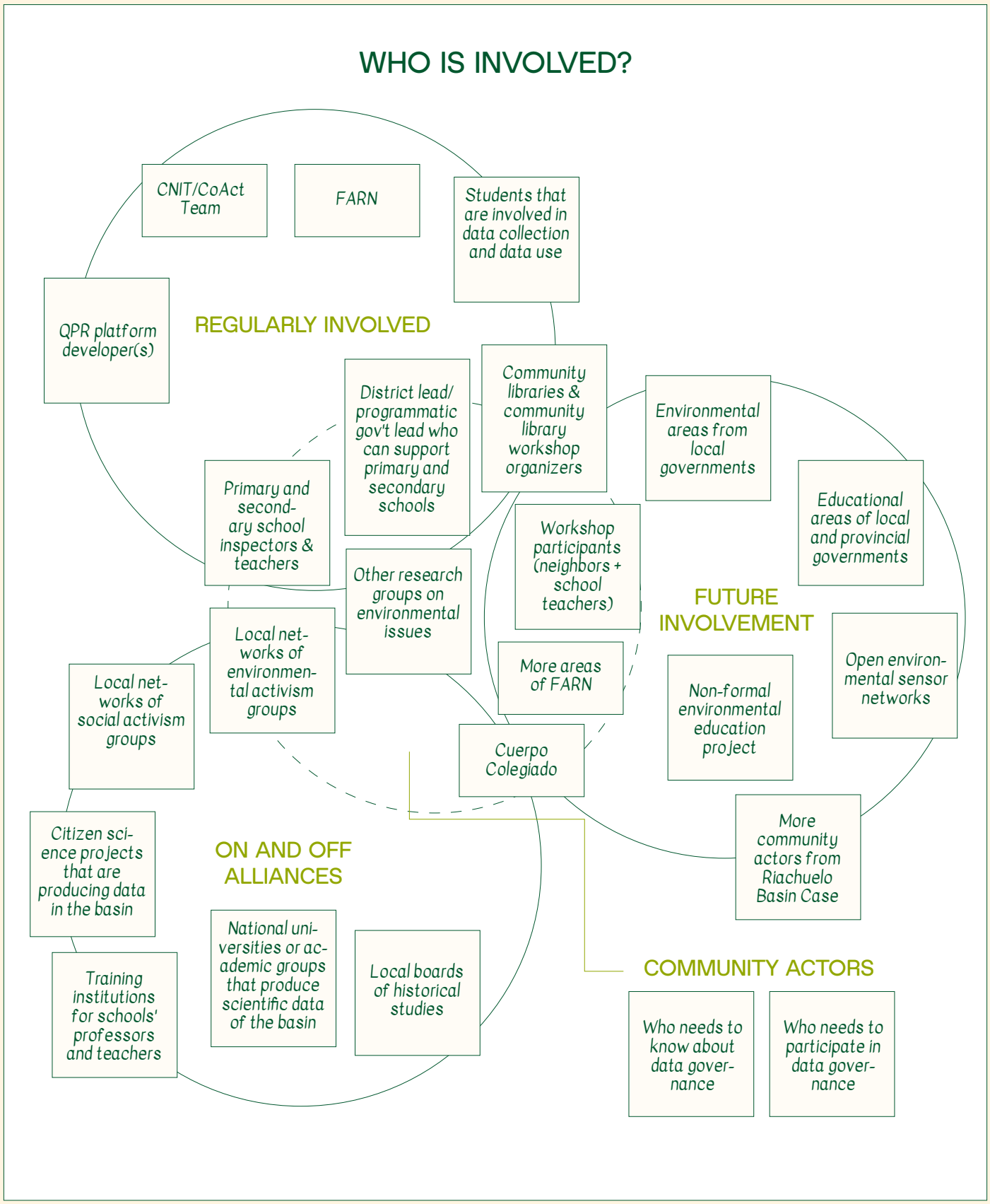

WHO IS INVOLVED?
CNIT/CoAct Team
FARN
Students that are involved in data collection and data use
QPR platform developer(s)
REGULARLY INVOLVED
District lead/ programmatic gov't lead who can support primary and secondary schools
Community libraries & community library workshop organizers
Environmental areas from local governments
Primary and second-ary school inspectors & teachers
Educational areas of local and provincial governments
Workshop participants (neighbors + school teachers)
Other research groups on environmental issues
FUTURE INVOLVEMENT
Local net-works of environmen-tal activism groups
Local net-works of social activism groups
More areas of FARN
Non-formal environmental education project
Open environ-mental sensor networks
Cuerpo Colegiado
More community actors from Riachuelo Basin Case
Citizen sci-ence projects that are producing data in the basin
ON AND OFF ALLIANCES
National uni-versities or ac-ademic groups that produce scientific data of the basin
Local boards of historical studies
Training institutions for schools' professors and teachers
COMMUNITY ACTORS
Who needs to know about data gover-nance
Who needs to participate in data gover-nance

FIGURE 10

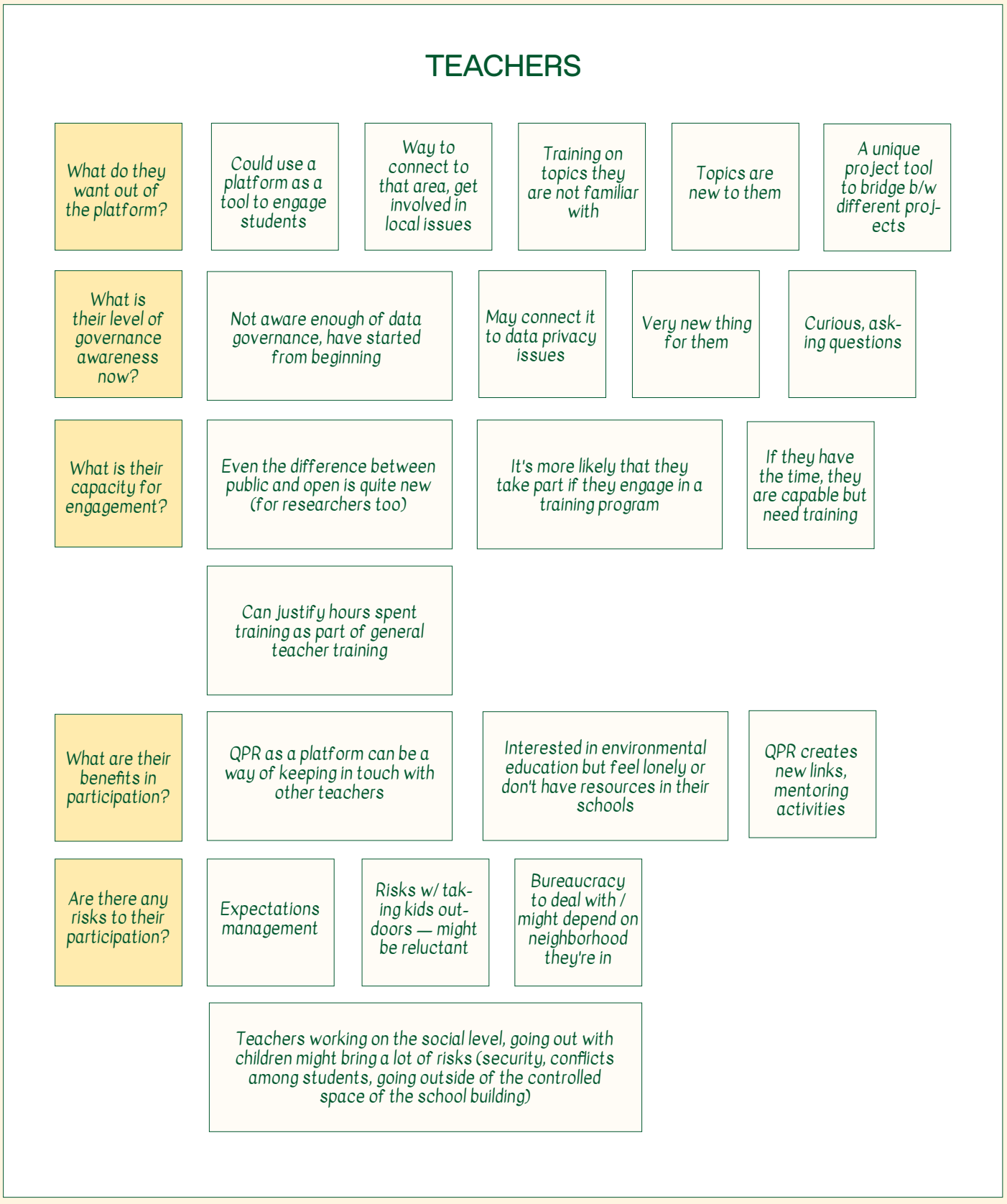

TEACHERS
What do they want out of the platform?
Could use a platform as a tool to engage students
Way to connect to that area, get involved in local issues
Training on topics they are not familiar with
Topics are new to them
A unique project tool to bridge b/w different projects
What is their level of governance awareness now?
Not aware enough of data governance, have started from beginning
May connect it to data privacy issues
Very new thing for them
Curious, asking questions
What is their capacity for engagement?
Even the difference between public and open is quite new (for researchers too)
It's more likely that they take part if they engage in a training program
If they have the time, they are capable but need training
Can justify hours spent training as part of general teacher training
What are their benefits in participation?
QPR as a platform can be a way of keeping in touch with other teachers
Interested in environmental education but feel lonely or don't have resources in their schools
QPR creates new links, mentoring activities
Are there any risks to their participation?
Expectations management
Risks w/ taking kids outdoors — might be reluctant
Bureaucracy to deal with / might depend on neighborhood they're in
Teachers working on the social level, going out with children might bring a lot of risks (security, conflicts among students, going outside of the controlled space of the school building)

FIGURE 11

TAKEAWAY #8

Documentation underlies effective collaborative data governance and institutional knowledge.

Documentation is critical to managing environmental data projects where multiple actors are involved in different aspects of the project. Like data documentation, data governance documentation can "ensure that your data will be understood and interpreted by any user."[35] This documentation can pertain to project-specific definitions; decision-making processes; frameworks related to roles, responsibilities, and permissions; privacy and risk considerations, legal mechanisms connected to the data (e.g., licenses, data sharing agreements, etc.), and necessary context about the project's provenance or history. Different documentation components can be public or private depending on the sensitivity of the information.

A shared definition of data governance and thorough data governance documentation become especially significant when it comes to communicating with new collaborators about your data governance practices. It can be difficult to involve many different actors in governance, but documentation can help determine entry points for actors with different capacities and interests. It can also be useful internally as a method for building institutional knowledge, especially when projects face turbulent political environments or if funding lapses.

35. Data Documentation. The University of Arizona Data Cooperative. Access here ↗.

MODULE #2

OPENING UP DATA GOVERNANCE ENGAGEMENT PATHWAYS

The second module explored how to best engage with the priority community actors identified in the first module, focusing on their ideal roles, and how CoAct could communicate about data governance with them and collaborate on aspects of QPR's data governance. For this session, we focused on the community library staff. Because they had previously been workshop organizers with CoAct, they had experience being users of the data platform and could build the skills and capacity to fill a more intermediary role such as data curators or community liaisons.

We had two guiding questions for this module:

- What training, agreements, or guides might help engage those users (community library staff)?
- What policies, tools, or data practices are needed to better integrate those users into data governance processes?

As community liaisons, library staff must be able to understand the difference between open and closed data, and to explain these differences to others, such as members of the public who attend workshops. As data curators, CoAct wanted community library staff to be able to review submissions to the QPR data platform, publish community-submitted data that had been vetted as non-sensitive data, and archive the more sensitive data into a password protected repository. For this role, CoAct delineated a broader data governance training that would build capacity around data sensitivities and aggregation, and provide examples around disclosures. They also highlighted that it would be useful for library staff to grasp the technical aspects of the platform that en-

abled data sharing to support their understanding of the purpose and mechanics of the quality assurance and control tasks.

We also discussed how to incentivize and tailor this learning for the community library staff. For example, building up basic data governance and literacy skills would have benefits beyond this project in the staffs' personal and professional lives. Especially in today's data-saturated society, having an understanding of data privacy and security could support people in making more informed decisions about their own personal, health, or financial data. Adapting this learning for the community library staff brought up several key questions for the CoAct team to work through: How could they explain things in a simple and appealing way? How could they incorporate rewards or progress indicators within the platform? And how could they communicate the purpose or the benefits of the process?

We came to the conclusion that CoAct could take a few different next steps to build out their collaboration with the community library staff and other community actors in the future. First, they could pilot a data curator training program with a small number of testers from the community library staff and include modules on open and closed data and basic data security training. They could also develop a checklist that detailed the workflow for reviewing data submissions. Alongside this development, they should document clear roles for each collaborator and project member, with special focus on the data curator role.

TAKEAWAY #9

Truly participatory data governance requires deep understanding of each collaborator or actor, including their values, incentives, and capacity.

Data governance can seem like a difficult or nebulous concept—many approach the topic unsure of its meaning or purpose, or assuming data science understanding is needed to interact with its practices or frameworks. This is why it's critical to have clearly defined pathways or roles for community actors to fit into; creating those pathways and roles requires a thorough understanding of how a community actor may or may not want to participate. Creating user personas or interviewing interested actors to understand their values, incentives, and capacity can ensure that engagement can be productive for both the project and community actors. Ultimately, expanding governance to community actors allows for greater representation in decision-making, and can create greater wells of interest and awareness outside of the main project team.

CLICK OR SCAN THE CODE HERE TO HEAR MORE ABOUT COACT'S NEXT STEPS.

RESOURCES CREATED, NEXT STEPS, AND OUTCOMES

After the workshop, we created a synthesis of next steps for CoAct to consider. With each community partner, we held check-ins at the three and six month marks after our official collaboration ended. In these meetings, the CENIT team shared their progress in creating a project-wide governance document that provides clear definitions of the data, their existing governance and decision-making processes, and issues to be resolved. They also shared that they have held a workshop with FARN to discuss this document, and both reported that they are still in discussions to collaboratively finalize it.

SECTION TWO

A PLAYBOOK TO ADVANCE ENVIRONMENTAL DATA GOVERNANCE

As our case studies illustrate, communities that steward environmental data are approaching and using data governance in different ways depending on their capacity, existing social and technical infrastructures, political contexts, and data use goals. Some data stewards have access to a robust technical platform but need support to strengthen their organizational processes. Others have specific needs related to legal protection and data security, or specific data co-ownership requirements with members of their community. While each case is highly context-dependent, we have identified takeaways and corresponding “plays” that can be useful for data stewards with varying needs.

A QUICK PRIMER ON THIS PLAYBOOK

This playbook is designed for data stewards and intermediary organizations, who want data governance strategies to implement throughout processes of data collection and use. Each play has different considerations around time, money, and resources; some plays might be more appropriate than others for your specific context. But like a coach on the sidelines, a data steward or intermediary organization can reference the plays, and utilize them depending on what's needed to get the ball down the field.

In other words, these "plays" are modular strategies to support better data governance. They can be used individually or in combination with one another; they can be adapted and used no matter what relational or technical infrastructure a data steward might currently have. We've grouped the plays for each takeaway as relational, technical, or as a support play.

- **Relational plays** for data stewards refer to the who, how, and when of data governance; the "people" aspects, as well as the processes, practices, and workflows that people undertake.
- **Technical plays** for data stewards refer to the where, how and when of data governance; the software and digital systems employed by the data stewards.
- **Support plays** for intermediary organizations refer to external organizations that provide infrastructural support or interact with data in some way, including research and academic institutions, foundations, data and tech-focused non-profits, or government agencies.

The full slate of takeaways, as well as their related relational and technical plays for data stewards and intermediary organizations, is listed on the next page. This playbook is only intended to be a starting point; our goal is to add to this list of plays in our Resource Library for Community Data ↗ as we continue to collaborate with new partners and uncover new takeaways.

WHO CAN USE THESE PLAYS?

We've written these plays for environmental data stewards. In this section, a data steward refers to any community-based organization, collective, or individual that is managing community-generated environmental data. Each community partner in SECTION 1, for example, is a data steward.

We also recognize that data stewards do not work in isolation. We included plays for intermediary organizations that support data stewards since they, too, have a role in championing data governance approaches. An **intermediary organization** includes the peripheral infrastructure of civil society organizations, researchers, technologists, and government staff that engage with environmental data stewards. OEDP, for example, could be considered an intermediary organization.[36]

36. OEDP has already put some of these plays into practice, while other plays indicate strategies we plan to pursue in the near future. We are open to collaboration, so if you're interested in 1) submitting a play, or 2) partnering with OEDP to execute support plays, please fill out the form in our Resource Library for Community Data ↗

TAKEAWAY	RELATIONAL PLAYS FOR DATA STEWARDS	TECHNICAL PLAYS FOR DATA STEWARDS	SUPPORT PLAYS FOR INTERMEDIARY ORGANIZATIONS
TAKEAWAY 1 Data values statements are vital to create a basis for decision-making and team cohesion.	PLAY 1 Build and use data values statements.	PLAY 2 Develop a values-oriented ReadMe for the technical implementer.	
TAKEAWAY 2 A strong connection between social, legal, and technical components is key to cohesive data governance.		PLAY 3 Conduct scenario planning to better understand and prepare for risks.	PLAY 4 Create and distribute communication products that socialize the criticality of relational and technical dynamics within environmental data governance.
TAKEAWAY 3 Risk exists, but deliberately designed data governance can create pathways for opening up data in responsible ways.	PLAY 5 Establish clear roles, permissions, and moderators. PLAY 6 Conduct regular data audits for accuracy and security.	PLAY 7 Develop the infrastructure for tech-facilitated risk and privacy processes.	PLAY 8 Create templates for roles and permissions documentation. PLAY 9 Create a privacy policy template. PLAY 10 Create how-to guides for discussing risk and developing mitigation strategies.
TAKEAWAY 4 The technical design of a data system can, and should, reflect the data user's needs in order to unlock the value of the data.		PLAY 11 Define necessary technical features, review platforms, assess against required features, and the risks for technical overuse.	PLAY 12 Create templates to support metadata usage.
TAKEAWAY 5 Data ownership, in practice, has material requirements: somewhere to house the data and someone to maintain it.		PLAY 13 Implement a data storage structure using our templates. PLAY 14 Choose a hosting provider to store and backup the data. Consider alternatives to traditional hosting.	PLAY 15 Explore the role of public institutions and civil society actors in supporting data infrastructures in communities.

CONTINUED ON NEXT PAGE

TAKEAWAY	RELATIONAL PLAYS FOR DATA STEWARDS	TECHNICAL PLAYS FOR DATA STEWARDS	SUPPORT PLAYS FOR INTERMEDIARY ORGANIZATIONS
TAKEAWAY 6 Data sharing agreements are contracts. They don't build relationships, but rather maintain existing relationships.	PLAY 16 Develop a readymade workflow for employing and maintaining data sharing agreements.	PLAY 17 Regularly review platforms to ensure that data sharing agreements are being upheld.	PLAY 18 Create templates to support data sharing.
TAKEAWAY 7 Because data governance can be understood and employed in many different ways, it is essential to define what a data governance framework looks like within your project team.	PLAY 19 Facilitate conversations on defining data governance for your project.	PLAY 20 Ensure tools can adhere to the project definition of data governance.	PLAY 21 Create a playbook of a variety of data governance models depending on their desired governance definition and goals.
TAKEAWAY 8 Documentation underlies effective collaborative data governance and institutional knowledge.		PLAY 22 Ensure documentation is available in a central place. PLAY 23 Develop processes for maintaining documentation, including selecting the team member(s) responsible, cadence, approvals needed, folder structures, and file naming conventions.	
TAKEAWAY 9 Truly participatory data governance requires deep understanding of each collaborator or actor, including their values, incentives, and capacity.	PLAY 24 Conduct research through informational interviews to create user personas or a community audit.		

TAKEAWAY #1

DATA VALUES STATEMENTS CAN CREATE A BASIS FOR DECISION-MAKING AND TEAM COHESION.

Data values statements are documents that describe how and why an organization or project uses data. They describe the purpose of the data use, and what the organization will or won't do with the data it collects or uses. While data values statements are relatively new in the realm of data governance, they provide a normative space where data stewards can not only reach consensus on priorities but also provide a standard of what to expect in their collaborations.

PLAY 1

Relational play for data stewards: Build and use data values statements.

Just as an organization's mission and values guide their purpose and align a team's actions, a set of data values can provide a strategic direction for an organization that relies on using and sharing data to fulfill its goals. Internally, it creates a foundation for how to create data workflows, products, and infrastructures. Externally, these documents can serve as guides to set others' expectations and convey organizational commitments. OEDP has created a document template (RESOURCE 1 in the Appendix) and a zine for data stewards on facilitating conversations to brainstorm values, principles, and practices related to how data is used and shared.

PLAY 2

Technical play for data stewards: Develop a technical README for the technical implementer.

A values-oriented README serves as a guide and grounding document for technologists on your team. It communicates the expectations, principles, and commitments your program upholds, helping technologists align their work (whether within the cells of spreadsheets or entire databases) with your organizational or project goals.

- **Purpose Statement:** Explain why this README exists, including the importance of aligning technology with program goals and values.
- **Articulate Core Values:** List 3-5 core values that your program upholds. Ensure these reflect both organizational ethics and practical technology use. Explain why each value matters and how it applies to the work.
- **Detail Expectations for Technologists:** Describe the behaviors, practices, and outcomes your program expects from technologists (e.g., privacy and security practices, documentation requirements, accessibility needs).
- **Identify Ethical Boundaries:** Clearly articulate boundaries to prevent misunderstandings and protect program integrity. For example, in an environmental impact assessment project, an ethical boundary could be: 'No data on endangered species locations will be publicly disclosed so as to prevent potential poaching or habitat disturbance.'
- **Make it Actionable:** Include detailed information that technologists can use to implement their work (e.g., specify which users can have access to which datasets).
- **Metadata:** Commit to revisiting and updating the README as necessary. Include when it was last updated and by whom. Stay current on metadata practices in environmental data collection in scientific and regulatory fields. Regularly consult resources such as the National Environmental Methods Index (NEMI) ↗ to ensure alignment with industry standards.

The values-oriented README template is available in the Appendix as RESOURCE 5.

TAKEAWAY #2

A STRONG CONNECTION BETWEEN SOCIAL, LEGAL, AND TECHNICAL COMPONENTS IS KEY TO COHESIVE DATA GOVERNANCE.

This is an overarching takeaway that underscores all plays in this playbook. Each of these plays works to build that cohesive bond between relational and technical infrastructure, and can be employed according to the levels of capacity, interest, and need of the community generating environmental data.

PLAY 3

Technical play for data stewards: Conduct scenario planning to better understand and prepare for risks.

Scenario planning can help organizations anticipate potential risks associated with data misuse, loss, and breaches. By identifying vulnerabilities and strategizing around responses early on, organizations can strengthen their resilience and uphold trust with data stakeholders.

IDENTIFY DATA SENSITIVITIES. Start by cataloging the datasets your organization handles, classifying them based on sensitivity and potential misuse risks. Use this classification to determine high-risk datasets that may require stringent safeguards. Consider the following:

- **Personal Data:** Could this data be linked to individuals, leading to privacy violations or identity theft?
- **Commercially Sensitive Data:** Could this information give rise to unfair advantages or economic harm if misused?
- **Contextual Risk Factors:** How might cultural, political, or social factors influence the potential for data misuse? For example, location data from mobile sensors could be used to track individuals from specific ethnic groups, enabling authorities to disproportionately surveil and overpolice these communities.

MAP THREAT SCENARIOS. Collaborate with multidisciplinary teams to envision plausible misuse scenarios. Consider:

- **Data Breaches:** What if malicious actors gain unauthorized access to the data? Analyze pathways to unauthorized access such as weak encryption, insider threats, or unsecured endpoints.
- **Unintended Misuse:** Could the data be misinterpreted or repurposed in harmful ways? For example, if air quality data is collected primarily from sensors in wealthier neighborhoods, policies or interventions based on this incomplete data could neglect areas that experience the highest pollution exposure.
- **Data Loss:** How might human error, system failures, or cyberattacks compromise the availability or integrity of data?

EVALUATE PREVENTIVE AND MITIGATIVE MEASURES. For each scenario, outline strategies to mitigate risks, emphasizing both prevention and response. Consider:

- **Technical Safeguards:** Implement access controls, data encryption, and regular audits. Use anonymization or de-identification techniques for sensitive datasets.
- **Policies and Training:** Develop clear data-sharing policies, ensure compliance with legal policies such as GDPR ↗ or CCPA ↗, and conduct regular staff training on data ethics and security.
- **Incident Response Plans:** Establish protocols for detecting, containing, and communicating breaches or misuses. Assign roles and responsibilities to ensure swift action.

CONTINUED ON NEXT PAGE

ENGAGE STAKEHOLDERS. Effective risk management involves collaboration with data contributors, users, and affected communities. Facilitate open dialogues to:

- **Identify overlooked risks**,
- **Build trust** by demonstrating a commitment to ethical data practices, and
- **Gather insights** for co-creating safeguards that align with stakeholder needs.
- **Monitor and Adapt.** Risks evolve as data landscapes, technologies, and threats change. Continuously monitor emerging trends and update your scenario planning framework accordingly. Incorporate feedback from past incidents and lessons learned to enhance resilience.

PLAY 4

Support play for intermediary organizations: Create and distribute communication products that socialize the criticality of relational and technical dynamics within environmental data governance.

The interconnected nature of social and technical aspects of environmental data systems and data governance is becoming a more mainstream idea. But the stories documenting the practices, workflows, and processes that knit those two aspects together are hard to find and draw inspiration from.

Intermediary organizations can create **communication products**, like slide decks, op-eds, speeches, and social media posts to make the "case" for the foundational importance of interconnectedness within environmental data governance. For example, an organization could create a slide deck that includes supporting evidence on why this understanding of environmental data governance is needed and include use case scenarios of its efficacy within differing contexts. This slide deck's utility is in its application: it could be shared among other intermediary organizations aligned with community science, environmental justice, socio-environmental research, or the broader open source data movement. It could also be shared with organizations and teams working directly with community-generated environmental data. It could be referenced during fundraising and grant-writing, as a way to demonstrate critical understanding of this aspect of the data economy.

TAKEAWAY #3

RISK EXISTS, BUT DELIBERATELY DESIGNED DATA GOVERNANCE CAN CREATE PATHWAYS FOR OPENING UP DATA IN RESPONSIBLE WAYS.

Risk emerges in different ways when data sharing. Data misuse runs the gamut from truly harmful—misaligned actors can selectively choose data points to communicate harmful or untrue narratives—to merely inconsiderate—people could use data without acknowledging the original source.The risks of sharing data can rarely be fully eliminated, but there are practices and tools that can mitigate against potential threat.

PLAY 5

Relational play for data stewards: Establish clear roles and permissions for each kind of data user.

Different kinds of users interact with the data within any given organization in ways that may be implicit. For example, users from the general public may view a map of data inputs but not the raw CSV files, while a project manager makes decisions on what licenses apply to certain datasets. **Formalizing these roles and permissions**, and **creating documentation** that captures this information, is a way to delineate who is responsible for what and to minimize risk. For example, an organization may be more willing to share data with another organization that has a well-defined and transparent methodology of who can access their data at any given time in the data's lifecycle, and lists specific people who can be contacted if issues arise.

Roles will vary based on the governance framework and organization, but potential roles may include: data project manager, advisory councils or boards, individual data owners, technical implementers, contractors, programmatic staff, or the general public. Permissions will also vary based on the individual data project, but potential permissions include: perform analyses using raw data, manage data licenses, share data products, verify data from new sources, etc.

The list of permissions can also include what specific roles are not allowed to do, or if they need extra permission from another role to do. For example, a data project manager may not be able to share a specific dataset unless they receive permission from that individual dataset's owner. Establishing these roles and permissions should be done in a collaborative setting with each role represented. Creating a timeline for future review and revision can also be helpful.

PLAY 6

Relational play for data stewards: Conduct regular data audits for accuracy and security.

A **data audit** is a process of evaluating the social and technical infrastructures related to an organization's data at regularly scheduled intervals. This may include reviewing the location, quality, and consistency of the data, as well as the common workflows or uses of the data and who is able to do what with specific datasets. Data audits can support the minimization of risk in that an organization can review general workflows of sensitive data to ensure that access is available to permitted parties and the technical infrastructure supporting that data is still serving its intended purpose. Data audits can also be set up in ways that mirror organizational priorities. The basic considerations include: determining where the data are, if the data are complete, if the data can be used for a specific purpose, and how and by whom the data is being used.[37]

37. Everything in Its Place: A Guide to Data Audits. Strings Not Factors. Access here ↗.

CONTINUED ON NEXT PAGE

From the data audit, a report can be created to document the state of the data at that particular moment. This report can serve as a record of the data over time, including the issues that have arisen and been addressed, and who might serve as a contact point for a specific change or decision.

Technical play for data stewards: Develop the infrastructure for tech-facilitated risk and privacy processes.

PLAY 7

Create a system that facilitates the management of data-related risks and privacy concerns. A good system makes it easier to stick to data protection requirements consistently. This means less chance of staff accidentally sharing sensitive information or for data to be exposed in unintended ways. Relevant parts of the system should also be included in data sharing agreements. This includes:

Privacy and Security:

- Think about the ways in which you may want to use the data in the future so that data collection is intentional.
- If collecting data on individuals, ensure that you have a consent process in place that shares with users what their data will be used for and their rights in regards to their data.
- If needed, use methods like differential privacy ↗ to anonymize statistical analysis.
- For particularly sensitive information, consider regularly reviewing and deleting data that is no longer required.
- Spend time completing the Ford Foundation Cybersecurity Assessment Tool ↗ to better understand your level of risk and make a plan for implementing the recommendations provided over time.

Access Controls:

- Implement role-based access control (RBAC) to provide appropriate levels of access based on user roles and responsibilities.
- Regularly audit user access permissions to ensure they align with roles and data sharing agreements.

Security:

- Ensure that those with direct access to the data are using secure and unique passwords for platforms.
- Select platforms that use encryption to protect sensitive information.

Maintenance:

- Set up calendar events to schedule reminders for data audits, compliance checks, password changes, and policy reviews and revisions.
- Ensure that platforms receive regular updates and security patches.

Support play for intermediary organizations: Create templates for documentation of permissions and roles.

PLAY 8

It can be difficult to conduct conversations that formalize implicit understandings into more concrete governance policies. Intermediary organizations can create templates to serve as a starting point to guide organizations through conversations like this. A template like this can include a simple table with the roles listed in columns, and the activities that are usually allowed, require extra permission, or are always prohibited in rows alongside each role. The authors of Building A Fisherman-First Data Ecosystem ↗ provide a good example of this on page 24.

CONTINUED ON NEXT PAGE

PLAY 9

Support play for intermediary organizations: Create a privacy policy template.

People may be wary of sharing personal identifiable information without knowing how their data will be used and safeguarded. It is standard practice to share information about how this data will be used and protected at the start of any collaboration. When starting to collect new data, the organization can share their privacy policy detailing the information they will collect, how it will be used, information sharing and disclosure, users' rights and choices, data security practices, and how changes to the policy will be made. This is common practice with many digital platforms or apps that collect personal information. While all privacy systems are fallible to breaches, this privacy policy can include that disclaimer. This type of privacy policy is a formal guarantee of stated values regarding data use, and can be utilized when collecting data from individuals or when collaborating with another organization. OEDP has created a standard privacy policy template that can be downloaded and adapted to organizational needs, found in the Appendix (RESOURCE 6).

PLAY 10

Support play for intermediary organizations: Create how-to guides for conducting a risk evaluation.

Collecting, using, and sharing data can present legal, financial, and reputational risks, as well as the risk of misuse by a third party. Data stewards often know of these risks, but might not have examined them collaboratively, documented them, or identified strategies to address them.

Intermediary organizations can support data stewards by creating a how-to guide for conducting risk evaluations such as this one created by SafeTag ↗. This could detail different datasets, data products, and workflows, and determine if and how they present risk. This discussion shouldn't be limited to types of risks, but also include how specific social or technical approaches can mitigate the potential for risk. For example, an organization may be fearful that a particular staff member may share sensitive information with the media without knowing it. One mitigation strategy could be to host a training session for employees on how and when to interact with the media and to codify both the protocols for speaking with the media and spelling out the consequences of data misuse in an employee handbook. When we say consequences, we don't mean to imply that risk evaluations should be fear-based or use scare tactics, but rather that they should codify processes in a proactive desire to build relational or technical protocols and workflows that will secure the data against potential misuse.

TAKEAWAY #4

THE TECHNICAL DESIGN OF A DATA SYSTEM CAN, AND SHOULD, REFLECT THE DATA USER'S NEEDS TO UNLOCK THE VALUE OF THE DATA.

The technical aspects of data management can often be the most difficult to tailor to stewards' needs. Data storage requires a level of digital literacy above the average person's, digital infrastructure can be expensive, and customized systems or features require even more funding. Open source solutions may be available, but less findable or usable to the average user. Oftentimes, environmental data stewards are left with clunky or unnecessary technical systems that don't correspond with their data sharing or use needs. Recognizing this difficulty and misalignment, we've outlined methods to determine the highest priority needs in a technical model and make decisions accordingly, as well as an opportunity for intermediary organizations to support interoperability of any technical data system.

PLAY 11

Technical play for data stewards: Define necessary technical features for tool selection.

Clearly identify the technical capabilities needed to achieve your project's objectives. By clearly defining data usage rules, organizations can safeguard communities and critical ecosystems, build trust with stakeholders and local communities, and ensure high-quality, transparent data collection that supports environmental justice efforts.

1) Conduct a needs assessment: Gather requirements by consulting stakeholders and understanding project goals. In particular, aim to understand:
    a) What problem does the platform need to solve?
    b) How will we know if we've succeeded?
    c) What constraints (time, budget, or resources) do we have?
    d) What can be done successfully without the use of technology? What runs the risk of being over-engineered?
2) Categorize features into "must-have," "nice-to-have," or "future needs" to prioritize them.

CATEGORY	DEFINITION	EXAMPLES
MUST-HAVE	Features that are necessary for use	User authentication, Metadata tagging
NICE-TO-HAVE	Valuable features that improve user experience	Dashboards, API to connect to other platforms
FUTURE NEEDS	Features that are low priority and aspirational	Automation

CONTINUED ON NEXT PAGE

3) Research and document these requirements in a feature matrix. List selected platforms down the left hand of your matrix and evaluate each platform by scoring them against your features list.

PLATFORM	MUST-HAVE #1	MUST-HAVE #2	MUST-HAVE #3	NICE-TO-HAVE #1
PLATFORM 1	Yes	Limited	Yes	No
PLATFORM 2	Yes	Yes	No	Yes

4) Engage key stakeholders in reviewing the top choices and document all findings to ensure transparency.
5) Select the platform that best meets your project's needs.

PLAY 12

Support play for intermediary organizations: Create templates and guides to support metadata standards and documentation.

Metadata, as a tool for findability, interoperability, and understanding, may be one of the first technical components that data stewards learn and have to make decisions regarding. Yet, there are few if any succinct and useful metadata guides written for community environmental data stewards, specifically. This support work is essential: transparent data provenance ↗ is vital in any type of environmental regulation. There is an opportunity for intermediary organizations to create a suite of metadata support, pulling from existing resources, and adapting the metadata formats and standards to fit the needs and contexts of community environmental data stewards. This suite could entail:

- **A guide to different metadata standards:** Depending on the data, different standards might be more suited than others. A guide could lay out existing standards for various data types or agencies.
- **A guide to crosswalks:** To improve interoperability, intermediary organizations can provide information on how to use crosswalks ↗ or conversions to make sure that datasets with different metadata can be used together. For example, if an organization is collecting air quality data and recording the data in a 12 hour format, but an institution uses a 24 hour format, a key for conversion can be included in the documentation.
- **Documentation templates:** Data stewards can record their standards, crosswalks, and conversions in a streamlined documentation format that can be stored alongside their data and available to the public. Intermediary organizations could create a template for this type of document.

TAKEAWAY #5

DATA OWNERSHIP IN PRACTICE HAS MATERIAL REQUIREMENTS–SOMEWHERE TO HOUSE THE DATA AND SOMEONE TO MAINTAIN IT.

Many organizations that collect environmental data do so because they have questions that demand answers; by collecting their own data, using their own hardware and methods, they can create their own data-informed knowledge and evidence. The concept of data ownership recognizes this knowledge creation, and underscores who has authority to use that data and in what ways. This concept is powerful, but it loses its potential without the infrastructure to store, manage, and maintain that data. While open source solutions have become more available, each of the plays below either document easy-to-implement methods to retain data ownership and store data, or present opportunities to support data stewards in finding solutions.

PLAY 13

Technical play for data stewards: Implement a data storage structure using our cookie cutter templates.

We've created templates so you can begin implementing a standardized structure for data storage. These templates provide predefined folder structures, naming conventions, and metadata requirements that help organize data systematically. They can save time, ensure consistency, and reduce confusion about where to store files or how to name them.

1) **Download** the cookie-cutter template linked here ↗. Review the structure to see what works for your organization and what doesn't.
2) **Adapt.** Use the instructions ↗ in the file to customize the template to your needs.
    a) Rename folders to reflect the specific data and processes you work with.
    b) Adjust the folder hierarchies if needed (i.e., add, remove, or merge folders)
    c) Customize file naming rules to reflect your organization's workflows (i.e., include dates, versions, or author names)
3) **Test** the template with a small subset of data. Ask yourself:
    a) Is the folder structure logical?
    b) Is information easy to find?
    c) Will it work with other datasets?
4) **Pilot** the template with a small group of users, providing training for team members on its use.
    a) Encourage the active use of the template for a set period of time.
    b) Collect feedback from the pilot users. What are the pain points? What worked well?
5) **Roll out** the template to the rest of the organization.
    a) Provide training and clear instructions for use.
    b) Designate a point person to answer any questions.
    c) Schedule periodic reviews and feedback sessions to ensure that things are working as intended.

CONTINUED ON NEXT PAGE

PLAY 14

Technical play for data stewards: Choose a hosting provider to store and backup the data. Backing up your data ensures your valuable information is securely stored and regularly backed up, protecting against data loss.

- **Research:** Research hosting providers to store your data. Consider providers that prioritize sustainability and data ethics, such as Greenhost ↗ and GreenGeeks ↗ which emphasize transparency, renewable energy use, and privacy.
- **Evaluate:** Evaluate the providers based on factors important to your organization including data security (e.g., encryption standards), compliance with regional data regulations, scalability to meet project growth, and backup and recovery capabilities.
- **Select:** Select the provider most aligned with your needs.

PLAY 15

Support play for intermediary organizations: Explore the role of public institutions and civil society actors in supporting data infrastructures in communities.

Intermediary organizations can examine the feasibility of public institutions hosting or supporting relational and technical data infrastructures. The plays in this playbook will require at least some human or technical resources, and some will require a lot more than what small organizations have. Public institutions might be able to support capacity building or provide technical assistance. Such an examination should seek to identify where both trust and resources are available in a given space, whether that is a library, community hackerspace, or local university. Detroit Digital Justice Coalition's "DiscoTechs," ↗ or community learning spaces "where people can discover technology together, learn at their own pace, and learn from people who are accessible and understand the context of their neighborhoods and communities,"[38] is a model for utilizing local resources, in both the talents of the community and local trusted spaces like libraries to catalyze interactions with digital justice. This model could be replicated to explore data ownership with the community and share skills and tools related to data governance.

Guiding questions for this research could be: Where are there existing social or technical public resources that may support this work in specific locales? What does data ownership look like when multiple stakeholders steward environmental data? How can we build trust in order to galvanize meaningful community interactions with public institutions? There is a rich vein of opportunity here for research that could serve communities who are exploring different forms of data ownership and governance.

38. How To DiscoTech. Detroit Digital Justice Coalition. Access here ↗.

TAKEAWAY #6

DATA SHARING AGREEMENTS ARE CONTRACTS. THEY DON'T BUILD RELATIONSHIPS, BUT RATHER MAINTAIN EXISTING RELATIONSHIPS.

While data sharing agreements are essential tools in collaboration, they are only one aspect of a larger relationship or partnership. Especially with environmental data, there can be power asymmetries between the two data-sharing parties, such as community based organizations and university-affiliated researchers. These power asymmetries materialize as differing levels of capacity, funding, access to legal resources, and incentives, which have disadvantaged smaller community-based organizations and led to data misuse and extractive relationships. Power asymmetries don't necessarily stall a collaboration, but they should prompt the two parties to align on the central purpose and goals of the data sharing and use. In other words, they need to be able to build a strong working relationship before sharing.

PLAY 16

Relational play for data stewards: Develop a readymade workflow for employing and maintaining data sharing agreements.

When developing data sharing agreements, understanding what a working relationship looks like before and during the collaboration is just as important as what is contained in the agreement itself. Data stewards can build out their own Partnership Protocol, a basic rubric for determining the validity of a new sharing partner. Questions to include in a protocol like this include:

- In introductory or exploratory meetings, do you need to have a conversation about goals, purpose, and values, and how many of those need to align before sharing data?
- Should formalizing the partnership require a vote, consensus, or other explicit form of decision making?

Upon determining the validity of a new partnership and setting up a data sharing agreement (ideally with legal counsel), data stewards can create a checklist within the Partnership Protocol to maintain the agreement. At scheduled intervals, data stewards can review this checklist to detail a partner's compliance and uses of the data. The checklist can be derived from the language in the data sharing agreement, which may include questions regarding how the partner is sharing the data, how they are using it in research, or how and if they have credited the data's source.

PLAY 17

Technical play for data stewards: Regularly review platforms to ensure that data sharing agreements are being upheld.

Establish a systematic approach to reviewing platforms and ensuring compliance with data-sharing agreements.

1) Schedule regular (e.g., quarterly or biannual) reviews of platforms and processes.
2) Where possible, use monitoring tools to track data usage and access logs for suspicious activities.
3) Conduct audits to compare access privileges with data-sharing agreements.
4) Document findings and address any non-compliance. Communicate results with key individuals.

CONTINUED ON NEXT PAGE

PLAY 18

Support play for intermediary organizations: Create templates to support data sharing.

Intermediary organizations can create customizable templates to support data sharing. These could include templates for creating partnership protocols, data sharing agreements, and facilitating conversations on data sharing. OEDP has started this work: see our how-to zine for data stewards ↗ to think critically and prepare for discussions about data agreements.

TAKEAWAY #7

BECAUSE DATA GOVERNANCE CAN BE UNDERSTOOD AND EMPLOYED IN MANY DIFFERENT WAYS, IT IS ESSENTIAL TO DEFINE WHAT DATA GOVERNANCE LOOKS LIKE WITHIN YOUR PROJECT TEAM.

Data governance encompasses the decision making rules, processes, workflows, and tools that determine who owns, manages, shares, and uses the data and how, as well as where the data is stored and shared. Different aspects of data governance may be prioritized differently depending on data stewards' varying needs. For example, metadata and findability may be paramount to one organization, whereas protection of sensitive data may be the highest priority for another.

PLAY 19

Relational play for data stewards: Facilitate conversations on defining data governance frameworks or models for your project.

A facilitated conversation with pertinent members of a data-generating community can elevate the specific priorities that matter most. Getting a sense of each participant's understanding of data governance, as well as priorities for data use, can align your organization and support choices about technical infrastructure or governing structures within a framework or an existing governance model.[39]

For example, an organization may prioritize public participation and wider community use of collected data in their governance framework. This would lead them to explore data governance models that value findability and usability, like a data commons ↗ or data collaborative ↗. Another organization may value data protection and participant data co-ownership, which may be more suited for a data trust ↗. Another organization might have particular processes in place for data sharing and documentation within a framework, but it doesn't align with other more established models.

Below are some prompts to help parse out which elements are priorities for your framework or model:

- What are your current workflows or processes related to data collection?
- Which of your datasets or data products are open? Which are conditionally shared or closed to the public?
- What are your goals in sharing or using your collected data?
- What agreements do you have in place for co-ownership of data?
- How are decisions about data currently being made?

PLAY 20

Technical play for data stewards: Ensure your tools and workflows can adhere to your framework of data governance.

Evaluate each tool in your workflow to confirm that it aligns with your project's data governance framework. This includes ensuring tools support key principles like data access restrictions, auditability, and version control.

39. A data governance framework is the particular set of values, roles, rules, processes, and tools that are relevant to how a particular organization governs data. A data governance model is a more established set of relational or technical rules or structures that an organization can adopt, such as a data commons or data trust.

CONTINUED ON NEXT PAGE

1) Map your governance policies (e.g., data privacy, access controls) to tool capabilities. For example, check if the tool allows role-based access or has built-in features for tracking changes to datasets.
2) Review your findings to identify areas for improvement and implement updates or customizations as needed.
3) Train team members on how to use the tools in a way that aligns with governance policies.

PLAY 21

Support play for intermediary organizations: Create a guide detailing a variety of data governance frameworks and models depending on their desired outcome.

There is a plethora of writing on different data governance models, published by and tailored to various audiences. It is difficult to cut through the noise of all the potential avenues for learning about data governance models, and none that we're aware of focus specifically on the audience of data stewards or organizations collecting and using community-generated environmental data. Their goals are contextually driven and markedly different from corporate or academic data governance, and as the case studies demonstrate, require unique solutions to unlock the potential benefit of data use and sharing.

Intermediary organizations can create a user-friendly guide that links an organization's definition and goals to specific data governance models. This guide can include start-up implementation processes for employing these models, pros and cons of each model, and use case scenarios. Additionally, this guide can provide information on how to document a framework that doesn't fit neatly into a predetermined governance model.

TAKEAWAY #8

DOCUMENTATION UNDERLIES EFFECTIVE COLLABORATIVE DATA GOVERNANCE AND INSTITUTIONAL KNOWLEDGE.

Documentation about an organization's data can include information on provenance, metadata and other contextual information, how decisions are made, the methodology behind collection, and data access and use conditions. These types of documentation ensure that data can be understood and utilized by any user. Many organizations have an understanding of these aspects, but haven't made them explicit or published them alongside the data itself. Documentation allows for stewards' or organizational knowledge to be safeguarded and transparent for the public, and also provides an avenue to examine how these aspects have changed over time.

PLAY 22

Technical play for data stewards: Ensure documentation is available in a central place.

Ensure that users can easily find the information they need in a centralized location for storing all project documentation. Centralizing project documentation in one easily accessible location streamlines information retrieval, saving time and reducing frustration for team members who need to find specific documents quickly. Additionally, a centralized system makes it easier to track changes.

1) Select a central repository (e.g., GitHub, GitBook, or Google Drive) that is accessible to all stakeholders.
2) Organize documentation into clear categories, for example: project overviews, data dictionaries, technical manuals, policies, meeting notes, etc.
3) Use a version-control system to manage updates and track changes.
4) Make sure team members know where to find the documentation.

PLAY 23

Technical play for data stewards: Develop processes for maintaining documentation.

The most useful documentation isn't only created once, but rather updated regularly. Keeping documentation up-to-date ensures that everyone has access to the latest, most accurate information, reducing errors and confusion.

1) Assign responsibility for maintaining documentation to specific team members or roles.
2) Define a review cycle (e.g., quarterly or semi-annual) to ensure the documentation remains accurate and relevant.
3) Create a structured process for updates, including a checklist for content changes, stakeholder reviews, and approvals.
4) Use collaborative tools (e.g., a dedicated documentation platform like GitBook, or Google Docs) to track contributions.
5) Share the updated documentation with users and archive older versions as needed.

TAKEAWAY #9

TRULY PARTICIPATORY DATA GOVERNANCE REQUIRES DEEP UNDERSTANDING OF EACH USER, INCLUDING THEIR VALUES, INCENTIVES, AND CAPACITY.

Data governance as a concept can be difficult to understand, especially for those new to ideas surrounding data literacy. Data stewards and organizations can make governance more accessible by understanding how and why a user might want to interact with the data or its governance structures, and thus design appropriate avenues for engagement.

PLAY 24

Relational play for data stewards: Conduct research through informational interviews to create user personas or a community audit.

Organizations wishing to expand public engagement with their collected data or data governance can start by investigating the holistic picture of each actor, including their values, incentives, and capacity, through informational interviews. Questions may include:

- What are your goals in using the data?
- What is your understanding of data governance (or privacy, security, ownership, management, data's intended uses, etc.)?
- What is your capacity for engagement with the data or its owners?
- What incentivizes you to be involved or to use the data?
- What might concern you about using the data or being involved with its governance?

Organizations can hold informational interviews in order to 1) create user personas or 2) conduct a community audit. In creating user personas, data stewards can interview interested community actors to then create a persona that represents a type of actor. For example, if an organization interviews five teachers, they can "collate that research and personify specific trends and patterns in the data as a persona."[40] Using personas like these can help organizations understand patterns, especially if they have a large number of potential actors with whom they want to engage. Useful resources on how to create personas include those created by the Interaction Design Foundation ↗ and And Academy ↗.

Organizations can also use informational interviews to seed information for a community audit. This would be more useful if you plan to engage a more specific, smaller group of people and want to understand specific users. Interviews in this vein can focus more on descriptions of specific people, deeper motivating scenarios (what's the situation or challenge that brought them here?), and what they specifically need from the engagement. This kind of analysis can support understanding of what successful engagement looks like for one particular user, and if users can meet their particular goals in a way that brings value to the broader picture of data governance and use.

40. Personas - A Simple Introduction. Interaction Design Foundation. Access here ↗.

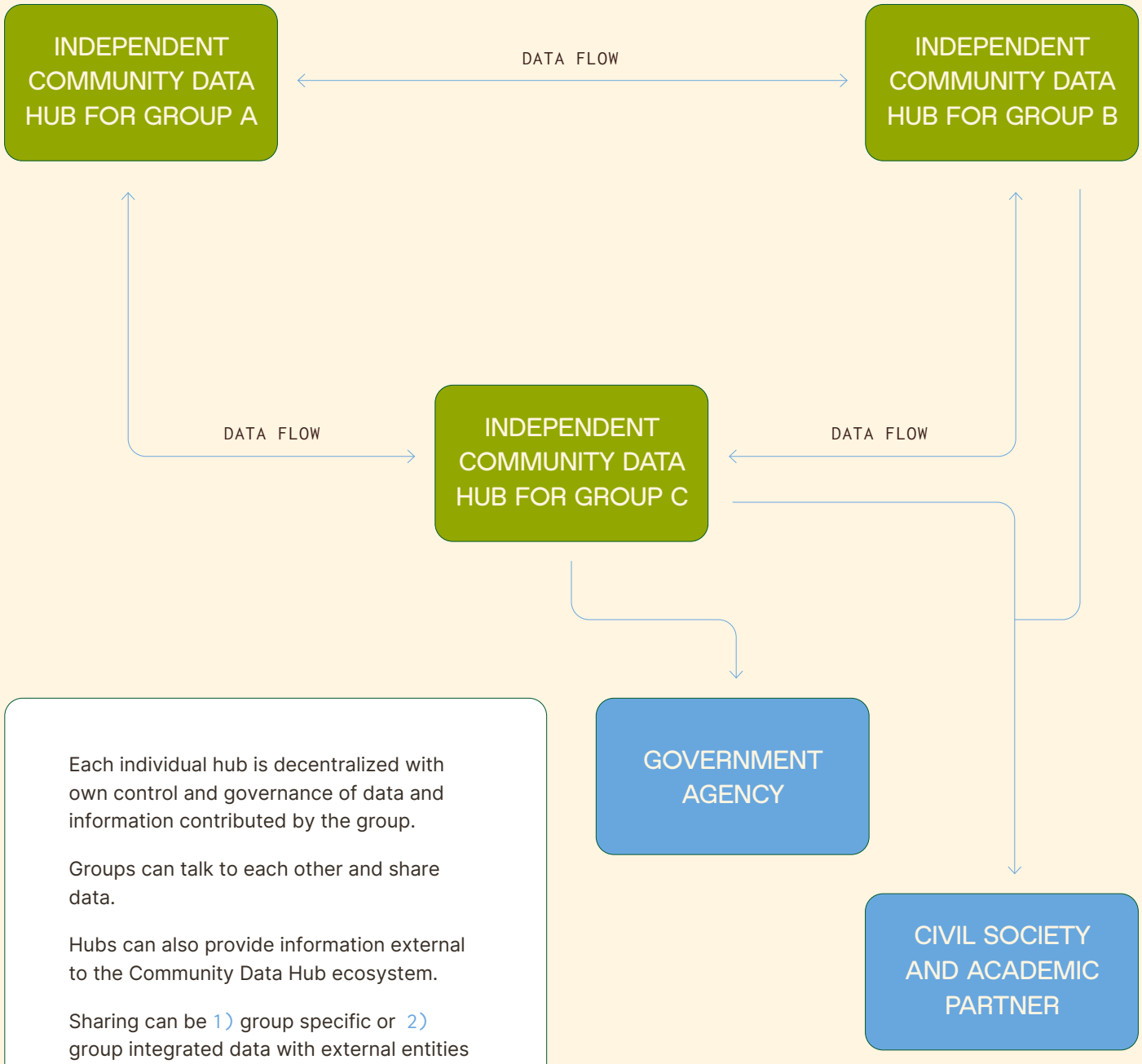
SECTION THREE

ASSESSING THE COMMUNITY DATA HUBS MODEL

At OEDP, we started exploring ideas around environmental data governance in 2021. Our brainstorms initially sparked an ambitious idea: a comprehensive framework for environmental data governance that can hold many types of data from varying sources, and can respond to relational and organizational needs, supported by a feature-rich technical infrastructure. We called it the Community Data Hubs (CDH) model, a prototype of a networked infrastructure for storing and sharing community-generated environmental data between communities, government, and researchers.

FIGURE 12

EXAMPLE OF A COMMUNITY DATA HUB



Each individual hub is decentralized with own control and governance of data and information contributed by the group.

Groups can talk to each other and share data.

Hubs can also provide information external to the Community Data Hub ecosystem.

Sharing can be 1) group specific or 2) group integrated data with external entities (government agencies, requests for information towards government policy, etc).

In our original conception, the CDH model would demonstrate how socially networked open data streams could preserve important local values while enabling timely, relevant, and trustworthy decision-making, both by and with community members. The model would feature a digital infrastructure with networked hubs, including:

- A place for communities to share and collectively manage data about their environment, providing mechanisms to decide how to share information;
- A schema through which each distributed hub could share data with other hubs or communities (see FIGURE 12);
- A resource library inclusive of data sharing agreements and policy templates which would a) serve communities' purposes and b) inform discussions with government on how to pilot new solutions; and
- A platform with feedback mechanisms for lawmakers to directly connect community information to policy development.

This model was built on a set of assumptions around what technological infrastructure would best support people in using environmental data in ways they wanted or needed. We first tested these assumptions with our advisory group, where our convened group of experts delved into particular case studies. In a similar vein, we conducted the collaborations we developed with our community partners to understand what their challenges were, identify what trends existed across the landscape, and determine if there were solutions that could be adapted, scaled, or integrated into the prototype. Some of our ideas and assumptions were affirmed, while others were invalidated or required further testing. Below, we share our learnings related to these assumptions, and outline opportunities for use cases in which a model like Community Data Hubs could meet data governance needs.

LEARNING

Many organizations already use specific technical infrastructures for their work and would prioritize other needs over creating a whole new technical model.

Our partners use established technical models or platforms; they are currently more focused on strengthening their relational governance practices or aligning their technical infrastructures to fit specific governance processes. This is what led us to developing the plays detailed in SECTION TWO, as "bite-sized" and modular strategies that organizations could employ, no matter what kind of relational or technical infrastructure they have in place.

LEARNING

For organizations that don't have a solidified platform, or are looking to pilot new digital infrastructure, the CDH model could be a good fit, but would require customization based on contextualized priorities.

In this specific collaboration and research endeavor, we focused on working with organizations that are already collecting data, which meant that most already had some kind of technical infrastructure to store and manage that information. To truly verify the benefit of the CDH model, we need to understand the perspectives of organizations that are just starting to collect data, or are actively seeking technical models or digital infrastructure for their efforts.

Still, our hunch around the potential benefits of a CDH model were confirmed: organizations do need straightforward and safe ways to not only decide on how information is shared, but also to share and collectively manage environmental data with other organizations and members of their community. A technical schema for sharing or pre-built digital infrastructure could model approaches to meeting these governance and sharing needs, but only if the specific context of a given organization is considered throughout its development.

LEARNING

The CDH approach requires buy-in from a group of organizations working in a geographically defined area.

Building on the previous learning, testing out this prototype would require collaboration with at least two organizations that 1) collect data and need a brand new technical infrastructure, and 2) would garner value from a shared data ecosystem. With the participation of two organizations, we could demonstrate the hub structure, albeit in a small manner, utilizing a sharing schema for data as they work toward similar or dissimilar goals. This need undoubtedly exists, but testing out a schema at an ecosystem level would require further research and partnership building to seek out multiple organizations that fit these criteria.

We sought out partnerships during this collaboration that were geographically separate and representing different environmental data types. As such, these organizations did not need to share data with each other. While this allowed us to better understand the broader environmental data governance landscape, we are excited to pursue future collaborations with geographically-linked organizations that want to explore creating a shared technical ecosystem.

LEARNING

A resource library could support environmental data stewards who are focused on community-generated data; in addition to data sharing and advocacy tools, this library should include templates and how-to guides.

While examining data governance challenges with our community partners, we uncovered the need for different kinds of resources that could support data sharing and advocacy. This broadened the scope of what topics could be included in a resource library, and how to package certain resources. For example, we discovered a need for resources for creating data values statements and data co-ownership templates. Our partners further elaborated that templates alone provide little utility unless they are accompanied by guidance on implementation—in short, templates increase accessibility and how-to guides increase usability. Along these lines, we recognized the need for basic explainer articles that establish environmental data governance and stewardship terms or concepts; shared, foundational definitions enable organizations to employ different approaches in order to unlock data's benefits. The combination of these learnings led us to create the Resource Library for Community Data ↗, a virtual space for community environmental data users to find resources, templates, and guidance for using and sharing environmental data safely, equitably, and effectively. We plan to continue to add to this Library based on input from our network of data stewarding organizations.

LEARNING

A future CDH prototype should feature a platform with feedback mechanisms and protection protocols that can be collaboratively used by data stewards and their partners in government and academia.

Organizations collecting environmental data interact with external partners, largely through sharing their collected data, with varying degrees of transparency, equity, and tangible value returning to the organization's community. Feedback loops within a repository could include an agreed-upon list of planned uses by the external partner and milestone indicators that demonstrate the progress toward these uses and their resulting benefit to the organization's community. For example, a university researcher may want to access an organization's air quality data to inform a particular study on pollution sources and impacts. In order to access the data, the researcher must demonstrate why this information in particular is integral to their research goals, and work with the organization to delineate a set of planned uses. When the researcher downloads or uses a particular dataset, this can be documented within the platform itself, and the community organization can be made aware of how the data are being used in real-time. This method of determining planned uses and documenting use can also serve as a protection protocol that dissuades external partners from misusing data, when used alongside appropriate data sharing agreements.

The Identifying Violations Affecting Neighborhoods ↗ (IVAN) Program in California is an illustrative example of how feedback loops can be built into government programs that utilize community data, beyond just cataloguing complaints. With IVAN, members of the public can monitor and report environmental violations and allow the person reporting to track what is happening with their reports. Learning from this and other models, OEDP and others can explore the design of systems for feedback with and for communities that collect data and interact with external partners to achieve their data goals, and determine if those systems of feedback could be incorporated into a CDH model.

CONCLUSION

In presenting our findings compiled over the past two years of learning and collaboration, we aim to leave readers with a central message: the challenges related to data usability within communities should not be understated, but there are a growing number of straightforward and easy-to-adapt tools, methods, and opportunities to pursue to build out functionality within environmental data governance infrastructures. Data governance is hard work—work that requires care and intentionality, a knitting together of the complex constellation of relational and technical infrastructures. Our community partners demonstrate that careful dedication in the collection, management, and sharing of their data.

We hope this playbook catalyzes conversations around what can be done: the avenues to explore, techniques to employ, and ways to adapt to challenges oftentimes found in the complex interactions between data and people using data to improve the environments that they call home. At OEDP, we are committed to investigating these interactions and working alongside communities as they re-invent their relationship to data and create governance systems that serve their own purposes. We will continue to learn and write new plays for data stewards, which can be found in our Resource Library for Community Data ↗. This playbook is not the culmination of this work but rather a milestone in collective understanding; it represents our team's commitment to explore environmental data governance and its ability to unlock data's value in pursuit of environmental justice.

APPENDIX

VIEW THIS APPENDIX ON GOOGLE DOCS ↗

RESOURCE 1
DATA VALUES STATEMENT TEMPLATE, CREATED BY EMELIA WILLIAMS 67

RESOURCE 2
DATA CO-OWNERSHIP RESOURCE, CREATED BY EMELIA WILLIAMS 68

RESOURCE 3
SELF-HELP ENTERPRISES HAULED WATER DATA MAPPING, CREATED BY KATE WING 69

RESOURCE 4
DATA USE AND SHARING AGREEMENT QUESTIONS, CREATED BY KATE WING 72

RESOURCE 5
VALUES-ORIENTED README, CREATED BY CATHY RICHARDS 75

RESOURCE 6
PRIVACY POLICY TEMPLATE, CREATED BY CATHY RICHARDS 76

RESOURCE 1

WHY AND HOW WE SHARE DATA VIEW ON GOOGLE DOCS ↗

Data supports _____ [organization/project]'s commitment to _____ _____[organization value #1], _____ [organization value #2], and _____ [organization value #3] . [If necessary, you can provide more context on organization values.] We collect, use, and share data in order to fulfill the following goals:

1) [Why #1]: [provide one sentence of clarification and context]

2) [Why #2]:

3) [Why #3]:

→ An example of a Why and clarification: "To connect our users to resources: Data supports our understanding of a given community or individual context, which allows us to tailor solutions to and generate resources for participants."

The way that we use and share data is based on our core values. The following values dictate our practices surrounding data of all types.

1) [Value #1]: [provide one sentence of clarification and context]

2) [Value #2]:

3) {Value #3]:

→ An example of a value and clarification: "Privacy and confidentiality: We respect your privacy and confidentiality and have practices that safeguaard your data as it relates to use and sharing. This is an extension of our principle to protect and serve communities, and so responsible stewardship of data is built into our workflows and governance."

We collect, use, and share different types of data, including [data types (e.g., aggregate data, individual and property-level data, personal identifiable information, and contact information)].

1) [Data type #1]: [provide one sentence of what this type of data is] [provide one sentence on how this data is used, what you will or won't do with this type of data]

2) [Data type #2]:

3) [Data type #3]

→ An example of a data type explanation and use: "Personal identifiable information (PII): Personal identifiable information is the most sensitive type of information and can include personal contact information like individual address, date of birth, or phone number. We share this data internally, and in some cases, with funders in required grant reporting. We do not share this information externally in any circumstance."

RESOURCE 2

DATA CO-OWNERSHIP DOCUMENT VIEW ON GOOGLE DOCS ↗

**Note that this is not a template for a legal document, but designed to be used more like a memorandum of understanding.*

→ Start with a statement that declares who owns the data, and if that is dependent on who collects it.

Example: Data collected by Self-Help Enterprises remains the property of Self-Help Enterprises and the program participant.

→ The following categories can be incorporated into the agreement based on what the data owner's prioritize in their partnership.

How can I access data collected?
Example: Participants will have access to their data through a portal, linked here. Participants can see data linked to their property but no other data from other properties. You can access the data for as long as we are using the data service, or for as long as the grant period lasts.

What are the ways it will be used?
Example: This is what we intend to do with the data: [list specific activities]. The data will not and can not be used to monitor individual properties or survey well levels. We will not monetize or sell your data under any circumstances.

What happens if the data needs to be shared with a new party?
Example: In the future, we may pursue new projects that will require sharing with new partners. We pursue partnerships that support our mission and goals, which you can read more about here. We will always notify you of potential changes and ask for consent, and if you choose not to share with this new partner, we will omit your data. If you do not respond to the consent request, after 60 days, we will continue with the proposed sharing with new parties.

What information has to be shared?
Example: We are required to share certain information with the California State Water Resources Control Board. We are required to share if a probe has been installed at a property and aggregate information about water levels at a regional scale.

What are the conditions of deletion?
Example: If the participant would like to discontinue working with Self-Help Enterprises well monitoring program in the future, they may notify Self-Help Enterprises in writing (by email or letter) or by calling the office [add phone number here] to revoke any permissions previously granted. Self-Help Enterprises will uninstall the probe and there will be no further tracking. Participants' personal identifiable information (e.g. contact information) will be retained due to grant reporting needs.

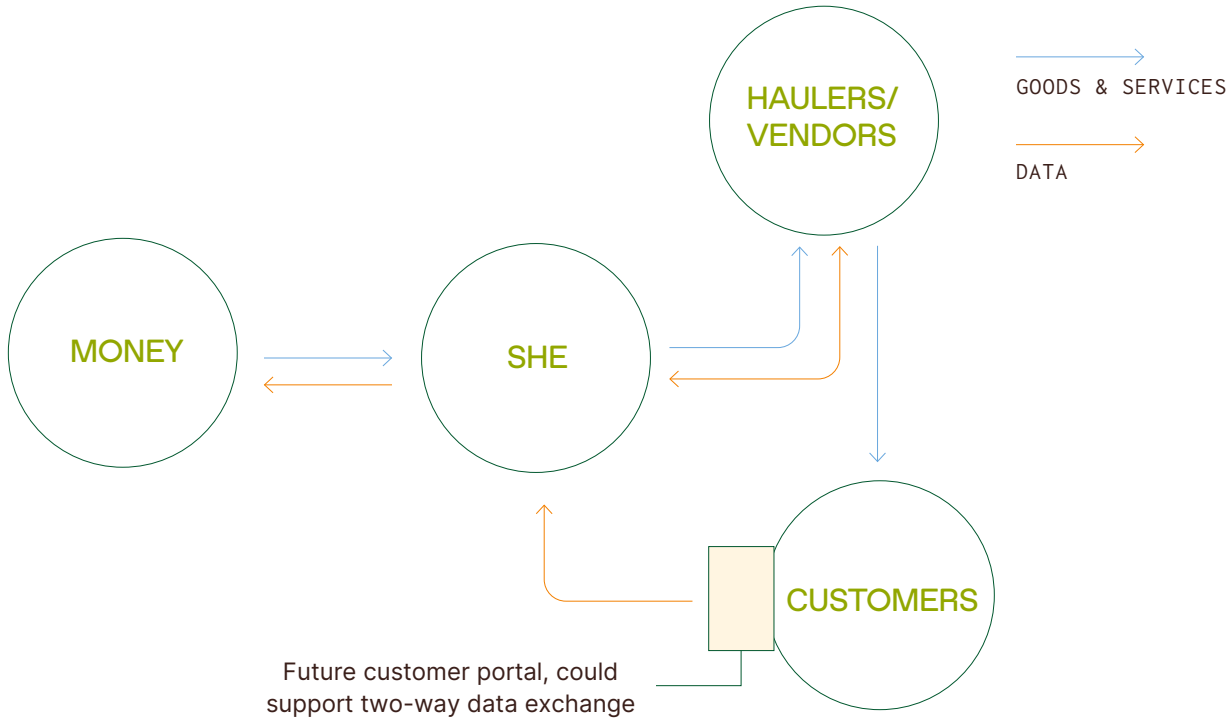
→ Include space for both data co-owners to sign and date.

RESOURCE 3

SHE HAULED WATER DATA MAPPING

VIEW ON GOOGLE SLIDES ↗

Data Ecosystem Map for Hauled Water



Data Exchanged

	FUNDERS	HAULERS / VENDORS	CUSTOMERS
DATA SHE RECEIVES		Price Estimated delivery time Delivery confirmation and actual delivery time Volume delivered, total and per address	Contact information for delivery recipient (name, address, phone, email)* Number of people at address
DATA SHE PROVIDES	Number of people served (per delivery, over X timeframe, by area) Demographics of people served Amount of water purchased	Delivery address Number of customers per address Contact information for delivery recipient (name, phone, email)*	Status of water delivery
OTHER SHE DATA NEEDS/ WANTS		Vendor reliability and availability	

CONTINUED ON NEXT PAGE

Data Questions & Considerations for Funders

What level of aggregation will they accept (time, location, maps vs. tables)?

"We provided water to 100 residences in Tulare County in July, with approximately 300 people"

OR

"Here's a heatmap of the most common delivery areas in 2022, where each dot = 50 people"

DISCUSSION TOPICS FOR CONTRACTS & FUNDING AGREEMENTS:

- A statement that SHE will only report aggregated data for customers and will not share individual contact information or data that could reveal individual identities, aka PII (if SHE develops an organizational data privacy policy, you could cite that)

Data Questions & Considerations for Vendors

- How much does SHE want to be the data hub, controlling what data vendors can see and access?
    - For example, if SHE develops a customer portal that supports tracking deliveries, this could also have a vendor view that shows customer data only until the delivery is completed, without actually transferring that data to the vendor
- If SHE is sending sensitive data to vendors (name, address, phone), does SHE want to check their data security practices?
- Is there other data vendors could help SHE collect to improve your services (e.g. 'this address has multiple residences' or comments on access to facilities)?

DISCUSSION TOPICS FOR CONTRACTS & FUNDING AGREEMENTS:

- Vendors will not share customer contact data and will notify SHE within 30 days of any data breach or legal request for data disclosure
- If there is a desire for real time tracking that can be shared with customers (e.g. a text alert when water is 10 minutes away), which vendors can support that and how will that data flow be managed?

CONTINUED ON NEXT PAGE

Data Questions & Considerations for Customers

- There may be multiple sets of personal data associated with each delivery address (e.g. property owner, person who will be home to receive delivery, other people at the address who are in SHE's database from other programs). Can SHE's data team make it easy for staff to see what data can be shared with which partners?
- If SHE is interested in (or required to) collecting additional data on customers, like demographic information, how will that be managed and linked to the contact data?
- Working with legal advisors on data minimization, data privacy, and PRA requests to understand how much is required to be provided, at what level of detail

DISCUSSION TOPICS FOR CONTRACTS & FUNDING AGREEMENTS:

- Statements of what data SHE collects, what it uses the data for, and what data vendors, partners, and funders can and cannot access. There may be subsections where people opt in to different data uses depending on the program, e.g. 'it's OK to text me at this number for water delivery but not for anything else'
- Explaining what data can be required to be provided under a court order or Public Records Act request (and making sure SHE's team is prepared to respond to those requests)

RESOURCE 4

DATA USE & SHARING AGREEMENT QUESTIONS FOR CHARRS

VIEW ON GOOGLE DOCS ↗

This document is a worksheet of questions and considerations for CHARRS and its partners around data sharing. It is not legal advice and should not be treated as legal advice. It is designed to be used for discussions by CHARRS and its partners, including discussions with attorneys and institutional contracting offices when exploring the terms of a data sharing agreement.

Roles and responsibilities as a data owner / data licensor

CHARRS staff may have specific responsibilities, obligations, and abilities to control how data are shared and used when CHARRS is the data owner. This may also be referred to as the data holder or data licensor. These are some situations when CHARRS may be treated as a data owner:

1) CHARRS collects the data directly (includes research with community leads - doing interviews, community using monitors sponsored by CHARRS, etc.)
2) CHARRS is part of a collaboration to collect data, in which case CHARRS may be collecting data for another partner (such as an academic researcher) and that partner is paying CHARRS. In this situation, it's important for CHARRS to understand:
    - What are the data requirements (if any) from the original funder or the institution paying CHARRS. For example, some federal grants have specific rules around data releases and the use of personal health information. Academic research institutions may also have rules around data, in addition to those of the funding source. These partners should communicate expectations around data to CHARRS as early as possible so CHARRS can prepare to meet them, including asking for additional capacity like staff time or cloud storage.
    - Will CHARRS staff or team members be expected to meet specific certifications or take training courses? This can be a requirement for certain institutions around information security.
    - How will CHARRS be able to use the data they collect during the project term as well as after the project is completed? If CHARRS wants to have access to the data during and beyond the project term, this may need to be explicitly spelled out in any contracts.
3) CHARRS hires someone else to collect data for CHARRS
    - How will the third party (such as a commercial sensor company or a consultant conducting surveys) store, use, and share the data?
    - Do you want them to delete the data once CHARRS receives it? Do you want them to store the data for you, and for how long?
4) CHARRS uses data from other sources to make a new data "product," such combining CHARRS' data with census and state agency data to create a map. CHARRS can be considered the "owner" of this new product and decide how to share and license it. Important: CHARRS must check the licenses on any data it uses to make the data product to be sure that CHARRS has the right to remix it into a new product, and, if it can be reused CHARRS must credit the original source appropriately (here is a guide to commonly used licenses).

When CHARRS is the data owner, what should CHARRS think about when sharing the data?

Who can use **what** and **how** they can use it

CHARRS may want to think about each of these categories independently, and how they intersect. The table on the following page shows some examples.

CONTINUED ON NEXT PAGE

WHO WANTS TO USE CHARRS' DATA	WHAT DATA CAN THEY ACCESS	HOW CAN THEY USE THE DATA
A university researcher and their students	Data *without* names and other personal details	Analyze public health issues and publish a paper on their findings, but not publish the underlying data
A university researcher and their students	Data *that includes* names and other personal details	Conduct follow-up surveys and interviews with CHARRS contacts for additional analysis, that will be shared back with CHARRS
A state public health agency	A map showing health issues and radon awareness, aggregated so that no fewer than five houses are summed into each data point	To develop new outreach materials and direct funding for remediation

What other conditions could CHARRS place in a data use agreement, or attach to data and data products?

Data use agreements are a type of contract and contracts can be very flexible, accommodating a wide range of conditions as long as all the partners agree to them. One thing to keep in mind with contracts is that you are responsible for monitoring the conditions to see if they're being met, and you can't take action unless a term of the contract is "breached." This is another reason why it's important to talk through the expectations for how data will collected, stored, and used at the beginning of any partnership and to keep checking in with your data partners to support the rules you've both agreed to. You want to try and work together to prevent any misuse of data before it happens, and then correct it if it does happen, before having to get the courts involved.

Basic provisions of a data use agreement

Most data sharing agreements will address these topics.

1) Identify who is covered by the agreement and how they will access or contribute data
   - This can be a general group (everyone in Dr. Scott's research lab) or people who meet certain conditions (users registered with the University's research data gateway and who are authorized by Dr. Scott)
2) Establish how data will be used and shared, both among the partners and with any external parties. This can include embargoes (ex. data will not be publicly released unto x days after the project is published) uses and disclosures and prohibit the recipient from using or further disclosing the data
3) Require appropriate safeguards to prevent an unauthorized use or disclosure
4) Require reporting of any unauthorized use or disclosure which the partner becomes aware of
5) Prohibit deanonymizing information or contacting or attempting to contact any individuals who participated in the research (ex. CHARRS surveys homeowners and a research partner cannot attempt to contact those homeowners directly)
6) Place requirements on data use and storage at the end of the contract, such as: all copies of data must be destroyed, data must be preserved in a specific repository, data products must be given to CHARRS, etc.

CONTINUED ON NEXT PAGE

Other provisions CHARRS could consider

1) Providing ongoing access to the community to data, information, and data products created with CHARRS data
2) Licensing data products to make them public or open, or require attribution to CHARRS
3) Regular meetings to check in on data use and management
4) Requiring that CHARRS be notified if the main point of contact changes, and setting an emergency contact/access option

Finally, CHARRS has a foundational commitment to equity that includes making sure communities are not exploited for their data, that they have access to data collected about them, and that data collected by and about them is used to support improvements in community quality of life. This may not be something CHARRS can require a partner to commit to in a contract (and, if CHARRS did, CHARRS would have to find a way to enforce it which could get complicated and time consuming). But, this is an important value CHARRS can talk about when signing contracts and consider how different contract provisions do or do not uphold that commitment.

For example, when talking with a new partner, asking about topics like:

- How will they protect sensitive data
- If CHARRS can review papers before they're published to make sure confidential data are protected
- If they will publish aggregate data on a free and open access platform

can be a way to confirm how CHARRS' values will be supported with actions during the partnership.

RESOURCE 5

VALUES-ORIENTED README VIEW ON GOOGLE DRIVE ↗

Introduction

Provide background information on the project that can provide useful context for the reader.

Purpose and Context

Explain why this README exists, including the importance of aligning technology with program goals and values.

Core Values

List 3-5 core values that your program upholds. Ensure these reflect both organizational ethics and practical technology use. Explain why each value matters and how it applies to the work.

Ethical Considerations

Clearly articulate boundaries to prevent misunderstandings and protect program integrity.

Expectations for Technologists

Describe the behaviors, practices, and outcomes your program expects from technologists. (e.g., privacy and security practices to adhere to, documentation requirements, accessibility needs).

Implementation Instructions

Include detailed information that technologists can use to implement their work (e.g., specify which users can have access to which datasets).

| Metadata

Last Updated: _____

By: _____

RESOURCE 6

PRIVACY POLICY TEMPLATE

VIEW ON GOOGLE DOCS ↗

Information We Collect

We collect the following types of information:

How We Use Your Information

We use the collected information for:

Information Sharing and Disclosure

We [do/do not] sell or rent your personal information to third parties. We may share information with:

Your Rights and Choices

You have the right to:

Data Security

Changes to This Policy

OPEN ENVIRONMENTAL DATA PROJECT

If you are interested in learning more about Open Environmental Data Project, or piloting or collaborating with us for one of the plays, please scan the codes below or email us at info@openenvironmentaldata.org ↗.

OPENENVIRONMENTALDATA.ORG ↗

SUBSTACK.COM/@OPENENVIRONMENTALDATAPROJECT ↗