
THE DATA ASSEMBLY
KEY FINDINGS ON THE RESPONSIBLE RE-USE OF DATA FOR CRISIS RESPONSE IN NYC
GOVLAB
HENRY LUCE FOUNDATION

CONTEXT

GOVLAB

THE DATA ASSEMBLY

A collaboration among and between citizens, civil rights organizations, key data holders and policymakers.

GOVLAB

HENRY LUCE FOUNDATION

Bklyn Public Library

New York Public Library

GOVLAB

WHY WE NEED A DATA ASSEMBLY

Policymakers and data holders often have little understanding of how different communities of users feel about the underlying issues—especially the tradeoffs between risk and benefit that are inherent to data re-use. As a result, regulators and government leaders often find themselves torn between competing impulses. On the one hand, they may adopt sharing and re-use policies that could endanger privacy and other rights of users, particularly those from traditionally marginalized communities. On the other hand, excessive caution may severely limit the options for data re-use out of fear of violating those rights, curtailing the wider societal benefits. This conflict between over-sharing and not sharing enough is the central conundrum faced by data governance today. It is one we hope to begin addressing with this project.

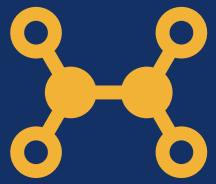
Signals about social attitudes and values toward data tend to come primarily from op-ed pieces in newspapers and broad surveys of public opinion, such as those contained within the Pew Research Center's Americans and Privacy report. Such surveys can provide a useful snapshot of public opinion, but they tend to lack the nuance and contextual discussion enabled by more deliberative methods. Deliberative public engagement methodologies (e.g. citizens' juries, citizens' assemblies, and public dialogues) offer a more context-rich approach, allowing us to understand how different constituencies make value judgements and how they perceive challenges and risks involved in data sharing.

GOVLAB

WHY WE NEED A DATA ASSEMBLY

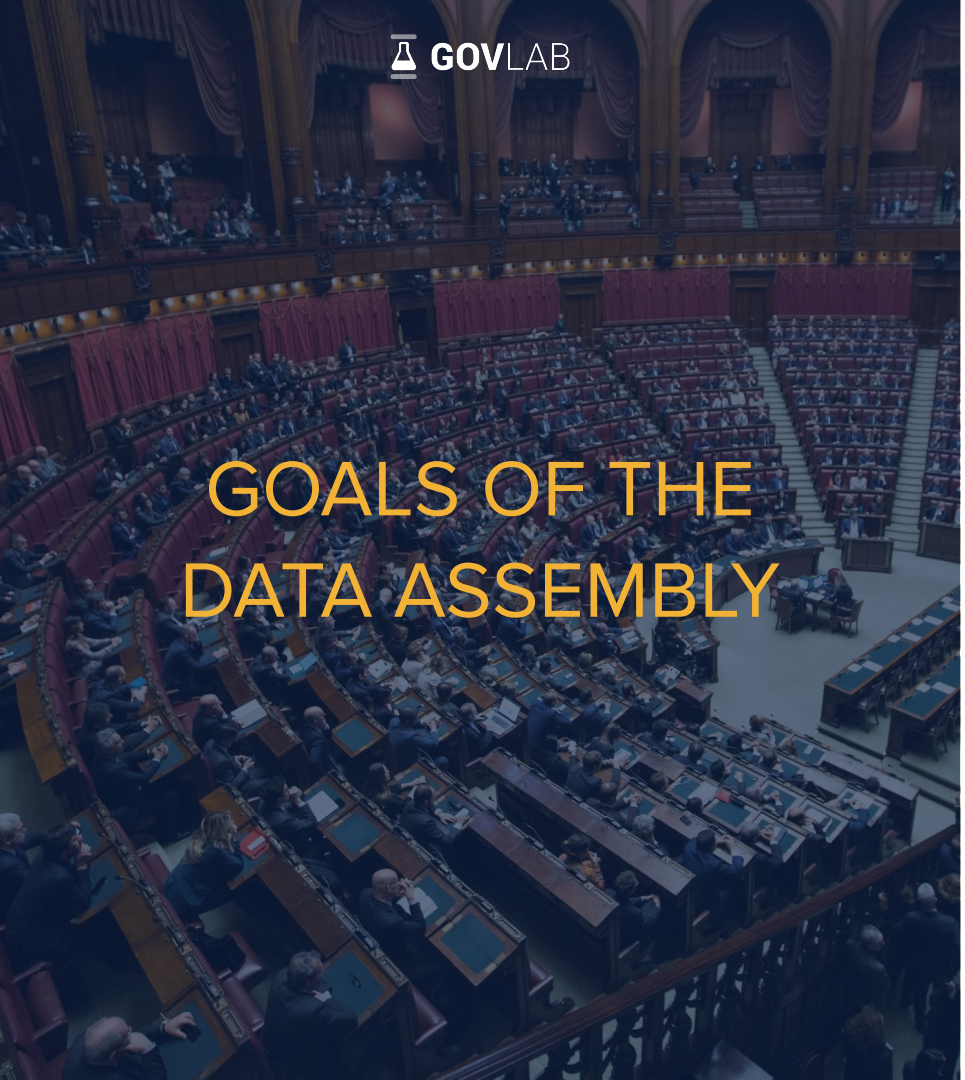
To achieve broadly acceptable policy solutions that harmonize and address the needs of as many stakeholders as possible, we initiated deliberations with three cohorts involved in or impacted by data re-use:

A mini-public of New York residents that aims to be representative of various community groups and individuals;



An assembly of rights groups and advocacy organizations; and

A selection of data holders and policymakers operating in New York City.



- Identify varying concerns, expectations, and opportunities surrounding data re-use;
- Produce cross-cutting recommendations to support policymakers and practitioners
- Co-design framework for responsible data re-use

METHODOLOGY

The Data Assembly deliberations took place during July and August of 2020. The GovLab and its partners at the New York Public Library and Brooklyn Public Library facilitated 90-minute remote video conferences with the data holders and policymakers mini-public and the rights groups and advocacy organizations mini-public. Both of these consultations involved between 15–20 experts curated using the GovLab's Smarter Crowdsourcing methodology.

The New Yorkers Mini-Public deliberation occurred on the Remesh online research platform. This consultation featured 55 New York City residents, sourced through the Remesh sampling methodology, with a focus on diversity across age, gender, income, and borough of residence. The Remesh platform provided participants with the ability to respond to polling questions, free-form text prompts, and to indicate their support for the contributions of their fellow participants.

CROSS-CUTTING RECOMMENDATIONS

GOVLAB

Match Urgency with Accountability

Participants in all three mini-publics expressed a willingness to tolerate increased surveillance for public health purposes. However, this expanded support for data collection and re-use does not excuse organizations from abiding by responsible data practices and other basic duties of care. Organizations should provide mechanisms that guarantee public oversight of their actions and provide opportunities for public input and accountability.

GOVLAB

Support and Expand Data Literacy

Though recent events have prompted more awareness of data re-use, many government leaders, community groups, and members of the public might lack knowledge of certain data practices and terminology. As such, meaningful public participation (including informed consent) in a data re-use effort depends on all communications being clear, well-justified, and broadly understandable. Various actors, including public libraries, could play an important role in fostering data literacy.

GOVLAB

Center Equity

Data re-use can yield substantial benefits for a community, but these benefits are not always distributed to those who need them most. In the mini-publics, participants noted the ability for data projects to miss subsets of the population or otherwise exacerbate existing inequalities. To address these problems, organizations should consider whether the data they intend to re-use misses or under-represents any groups or whether the methods have the potential to otherwise cause harm.

GOVLAB

Engage Legitimate, Local Actors

Participants in the mini-publics highlighted the need for effective public engagement and leadership from local actors in government and civil society. The deliberations also pointed to the importance of involving trusted intermediary organizations working at the local level that can help to engage with and solicit input from target beneficiary communities.

GOVLAB

Develop Positions for Responsible Data Re-Use

Data re-use projects are complex undertakings that require coordination with various actors inside and outside an organization. Dedicated positions devoted to these issues can allow organizations to better respond to new circumstances as they arise. In The GovLab's work, we call the people in these positions "data stewards."

THE RESPONSIBLE DATA RE-USE FRAMEWORK

- LOCALIZED VALUE CREATION
- PLACE-BASED OPPORTUNITIES AND RISKS

- PURPOSE-DRIVEN RE-USE
- EQUITABLE BENEFITS
- MINIMUM VIABLE ANALYSIS

- FIT-FOR-PURPOSE DATA RETENTION
- END-TO-END DATA RESPONSIBILITY

- DATA PROVENANCE
- AGGREGATED AND ANONYMIZED DATA

- PARTICIPATORY ENGAGEMENT, CONSENT, AND DATA LITERACY
- COMMON FRAMEWORKS, METRICS, AND GUIDANCE
- TRANSPARENCY AND COMMUNICATION

- COMMUNITY ENGAGEMENT
- DATA STEWARDSHIP
- LOCAL ACTORS
- TRUSTED INTERMEDIARIES

GOVLAB

WHY

- **Purpose-Driven Re-Use:** The re-use of data in the context of COVID-19 should be tied to a clear and well-defined purpose.

- **Equitable Benefits:** Practitioners should prioritize data re-use that benefits all people, including under-served populations and those who are “invisible” in many institutional datasets.

- **Minimum Viable Analysis:** Practitioners should only re-use data when it is the most direct, least invasive means to obtain the desired outcome.

WHAT

- **Data Provenance**: Practitioners should capture and communicate the origin, potential biases, limitations, and previous uses of datasets to ensure that those re-using the data are clear on what insights the data can and cannot provide.

- **Aggregated and Anonymized Data:** While recognizing that risks can never be fully erased, practitioners should ensure an adequate level of data aggregation to guard against group privacy harms and re-identification of individuals.

WHO

- **Community Engagement:** Community leaders and members of the general public should be involved in the planning stages of data re-use to help clarify what is “mission critical” and valuable to them.
- **Data Stewardship:** Data re-use should never be a fully automated process. Human actors need to be involved to ensure data quality and accuracy, and provide oversight throughout the data lifecycle.
- **Local Actors:** Where possible, the re-users of data should be actors in local governments, nonprofits, businesses, or academia in close proximity to the problems at hand and intended beneficiaries.
- **Trusted Intermediaries:** Beyond data suppliers and data re-users, trusted third parties should be empowered to help support responsible, ethical, and legally sound data re-use.

HOW

- **Participatory Engagement, Consent, and Data Literacy**: Practitioners should engage data subjects and community leaders at the planning stage of a project to steer responsible data re-use. They should also seek meaningful consent, with the ability to opt-out prior to the initiation of data re-use. Clear and accessible language and data literacy education can support these efforts.

- **Common Frameworks, Metrics, and Guidance**: No one-size-fits-all approach will suffice for responsible data re-use. Nonetheless, practitioners should seek out best practices and engage with stakeholders to create repurposable public resources to support peer-learning and collaboration.

- **Transparency and Communication:** Throughout the data re-use lifecycle, practitioners should communicate regularly with data subjects regarding how their data is being handled and how it is (or is not) contributing to the intended purpose of the work.

WHEN

- **Fit-for-Purpose Data Retention:** Data should only be held for as long as necessary to address the core issue or to answer the key question that is driving the re-use project. Future-oriented or exploratory analyses require new consent.

- **End-to-End Data Responsibility**: Opportunities, risks and challenges exist at all stages of the data re-use cycle. Policies, procedures, and oversight should be designed and deployed with a focus on navigating inevitable shifts in circumstance over time.

WHERE

- **Localized Value Creation:** Practitioners should prioritize re-using data to address local, community-based problems and opportunities first and foremost.

- **Place-Based Opportunities and Risks:** The re-use of geolocation data in particular can lead to emergent or unexpected risks and challenges. Data stewards should be tasked with assessing and mitigating place-based risks on a regular basis.

TOWN HALL AND EXPERT PANEL

On October 14, 2020, The GovLab facilitated a 90-minute expert panel and virtual town hall to share findings from the Data Assembly and generate additional insights on the opportunities and challenges of responsible data re-use. The exchange sought to provide a platform for key stakeholders and leaders in the space, including co-hosts of the series and representatives from the mini-public deliberations, to reflect on the report and share their own experience in working to advance responsible data use and re-use.

THE DATA ASSEMBLY

Interested in learning more about the Data Assembly; collaborating on the implementation of these recommendations and the Responsible Data Re-Use Framework; or co-organizing a Data Assembly in your city?
Visit http://thedataassembly.org

GOVLAB

HENRY LUCE FOUNDATION