Open.Data.Policy.Lab

GOVLAB

A FACILITATORS GUIDE AND WORKSHEET

Operationalizing a Social License for Data Re-Use

Questions to Signal and Capture Community Preferences and Expectations

Stefaan Verhulst and Adam Zable

Introduction

A social license for data reuse refers to the collectively expressed conditions under which communities consider the reuse of data about them to be acceptable, legitimate, and appropriate over time. Unlike legal compliance alone, a social license embeds community signals, preferences and expectations about purpose, safeguards, oversight, and accountability as data is reused across contexts and evolves.

This guide seeks to support organizations and facilitators who want to capture and document community priorities, concerns, and conditions for reuse. It was prepared by The GovLab, building on prior work developed in collaboration with Agence Française de Développement and The DataTank, including *Responsible Data Reuse in Developing Countries* and *Reimagining Data Governance for AI*.

Why a Social License for Data Re-Use Is Needed

Data reuse - the practice of combining, sharing, or repurposing data that was collected for one purpose to serve another - is increasingly central to how institutions deliver services, generate insights, and address public challenges. When done responsibly, data reuse can strengthen governance and public services, empower communities, support economic opportunity, accelerate innovation, and advance scientific and social problem-solving.[1]

Yet many data reuse initiatives face a legitimacy gap. Existing governance tools, often built around individual, point-in-time consent, do not reflect the ongoing and collective nature of many secondary uses.[2] They offer limited ways for communities to influence decisions once data is shared, particularly as new risks, technologies, partners, or uses emerge. As a result, communities increasingly seek meaningful influence over how data about them is reused, while organizations seek practical guidance to build trust, reduce risk, and establish fair conditions for reuse.

A social license for data reuse responds to this gap by framing legitimacy as an ongoing relationship rather than a one-time transaction. It can be established through participatory engagement, clear documentation of expectations, and mechanisms that support accountability and review over time. This enables communities to define acceptable uses, boundaries and red lines, safeguards, and how reuse should adapt as circumstances change.

Operationalizing a Social License for Data Re-Use

There is no single model for establishing a social license. Approaches vary depending on context, risk, and capacity. However, they can be understood through three core phases, outlined below. While these phases often overlap in practice, each involves distinct mechanisms and methods. Together,

1 Verhulst, Stefaan and Andrew Young. "The Potential of Social Media — Intelligence to Improve People's Lives: Social Media Data for Good." The GovLab. September 24, 2017. https://ssrn.com/abstract=3141457

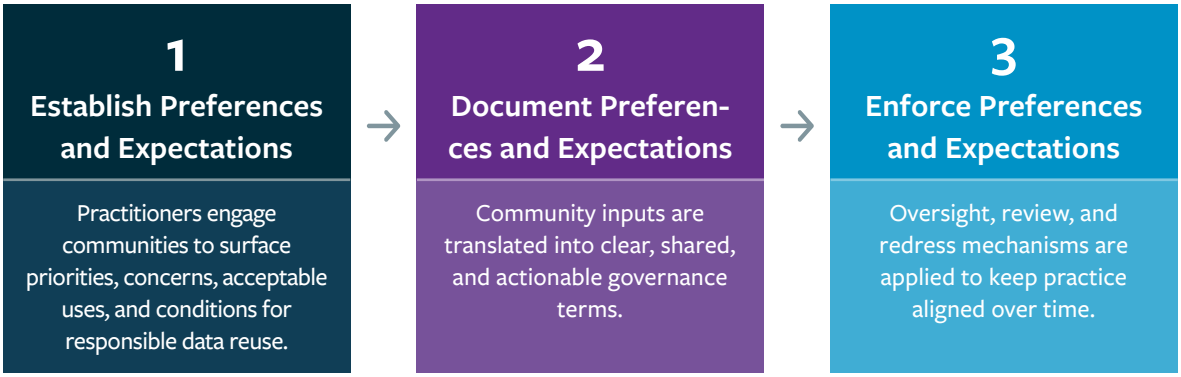
2 Addo, Peter, Adam Zable, Andrew Zahuranec, and Stefan Verhulst. "Reimagining Data Governance for AI." Agence Française de Développement Technical Report. April 2025. https://www.afd.fr/en/ressources/reimagining-data-governance-ai

they produce clear, community-defined parameters for data reuse that can guide decision-making, agreements, and oversight even as circumstances evolve.

1. **Establish Preferences and Expectations:** Practitioners engage communities to surface priorities, concerns, acceptable uses, and conditions for responsible data reuse.
2. **Document Preferences and Expectations:** Community inputs are translated into clear, shared, and actionable governance terms.
3. **Enforce Preferences and Expectations:** Oversight, review, and redress mechanisms are applied to keep practice aligned over time.

This toolkit seeks to support the implementation of Phase 1 by providing a guide on how to signal and establish preferences and expectations.

THE SOCIAL LICENSE PROCESS



Examples of Social Licensing in Practice

Social licensing for data reuse can take many forms depending on context, risk, and institutional setting. The examples below illustrate how communities have shaped data reuse in practice—highlighting different approaches to engagement, documentation, governance, and ongoing stewardship. Each example emphasizes a distinct pathway through which legitimacy has been established and maintained over time.

New York City Data Assembly (The GovLab, 2020): During the COVID-19 pandemic, The GovLab partnered with New York City and Brooklyn public libraries to convene three remote Data Assembly mini-publics, bringing together 55 randomly selected New Yorkers to deliberate alongside policymakers, data holders, and civil rights advocates on concrete data reuse scenarios related to the pandemic. Participants articulated conditions, safeguards, and priorities for pandemic-related data use. The resulting recommendations informed a public framework for New York City's response.[3]

3 Young, Andrew, Stefaan G. Verhulst, Nadiya Safonova, and Andrew J. Zahuranec. "Responsible Data Re-Use Framework." The GovLab. November 2020. https://thedataassembly.org/files/nyc-data-assembly-report.pdf

openSAFELY (UK NHS and partners, 2020): openSAFELY is a secure health data analytics platform that enables large-scale reuse of sensitive patient data without providing researchers access to raw records. Public legitimacy was initially established through independently commissioned citizen juries during the COVID-19 pandemic and is sustained through a standing public advisory group known as the Digital Critical Friends. Ongoing workshops, regular engagement meetings and "office hours," a coding club, and extensive user support continue to shape boundaries, safeguards, oversight, and platform expansion decisions.[4]

Language Data Commons of Australia (LDaCA) (Australian Research Data Commons, 2024-) LDaCA is a national research infrastructure initiative aimed at improving access to language data while respecting Indigenous data governance and cultural authority. The project has relied on co-design workshops and targeted consultations with Indigenous language custodians, researchers, and cultural heritage organizations to define acceptable uses, access conditions, and stewardship expectations. Planned cycles of public testing and consultation are intended to ensure the project remains responsive to community needs as they evolve.[5]

Choral Data Trust Experiment (Serpentine Arts Technologies, 2023–2025): The Choral Data Trust experiment explored collective governance of voice and cultural data created for training choral AI models. 15 choirs from across the UK contributed recordings through a purpose-built data collection process and participated in a multi-stage governance process led by an independent data steward. This included relationship-building conversations with choir representatives, open "Choral Data Conversations," surveys capturing preferences around data use, credit, compensation, and governance, and structured deliberation using Pol.is to reveal areas of consensus and disagreement about future reuse. These inputs were translated into concrete licensing preferences and governance requirements, informing the design of legal mechanisms, including performance rights agreements, a data rights mandate, and a trusted intermediary with representative oversight.[6]

Liverpool City Region Civic Data Cooperative (University of Liverpool and LCRCA, 2023): The Liverpool City Region Civic Data Cooperative is a regional data stewardship initiative that connects public sector organizations, researchers, businesses, and residents to enable data reuse for health and social benefit. The Cooperative combines inclusive governance structures with ongoing public engagement, including residents' assemblies on data and AI, community conversations with equity-seeking groups, participatory research on local data hubs, and continuous patient and public involvement in health data initiatives.[7]

4 Macaulay, Emily. "Data and AI Public Engagement Community of Practice Meeting Notes." Connected by Data. July 21, 2025. https://docs.google.com/document/d/1aNg-uLH1R6QpYQIcrjqfE4n4co-oyzNDpzialswnqVQ/edit?tab=t=0

5 Language Data Commons of Australia (LDaCA). https://www.ldaca.edu.au/

6 Ivanova, Victoria, and Jennifer Ding. "Choral Data 'Trust' Experiment: White Paper." Serpentine Arts Technologies. February 2025. https://d37zoqglehb907.cloudfront.net/uploads/2024/08/SerpentineArtsTechnologies_ChoralDataTrustExperiment_WhitePaper.2025.02.17.pdf

7 Liverpool City Region Civic Data Cooperative. https://civicdatacooperative.com/

Worksheet and Questions to Signal and Establish Preferences

The below worksheet is a facilitation tool designed to guide discussions during the phase of Signaling and Establishing Preferences. It is intended for use by community representatives, policymakers, researchers, and organizations engaging communities in data sharing initiatives to ensure data governance aligns with community values, expectations, and cultural considerations.

Trained facilitators - with expertise in data governance, legal frameworks, and participatory methods - can use the worksheet to lead structured dialogues, workshops, or focus groups to ensure community preference and expectations are captured in a way that can subsequently be used to document as enforceable agreements. It is not exhaustive but serves as a guide to shape participatory data governance processes.

The questionnaire is divided into six categories:

Why: Defining the purpose, scope, and limitations

What: Determining the appropriate data assets

Who: Identifying actors, roles, and responsibilities

How: Establishing operational governance and strategy

When: Addressing duration, retention, and review

Where: Considering contextual and jurisdictional factors

Each section includes key prompts to guide discussions and decision-making..

Data Management Literacy and Scenarios

Prior to engaging in the deliberation phase, it will be important to ensure that all participants share a baseline understanding of key data management and governance concepts. Providing a concise primer on topics such as data access models, data quality, privacy-enhancing technologies, stewardship roles, and governance mechanisms will help establish a common vocabulary and expectations.

In addition, we recommend structuring the discussion around a set of concrete use cases or scenarios. Anchoring the dialogue in real-world examples will make the deliberations more tangible, relevant, and actionable for participants, ensuring that insights translate into practicable pathways and not remain at the level of abstraction.

SECTION 1

WHY
Purpose, Scope, and Boundaries of the Project

Explore why data is being reused and what the community believes is acceptable, appropriate, or off-limits.

1.1 PURPOSE

Main Question:

What goals is the data reuse intended to serve, and how do these goals align with community priorities?

Explain to participants:

"This question helps us understand if the project's goals match what the community cares about. We want to ensure the project serves real needs, not assumptions."

Select all that apply and add details where relevant:

☐	**Improving public health**
	Examples: disease surveillance, maternal health tracking, vaccination planning.
☐	**Strengthening education**
	Examples: attendance insights, learning gaps, resource allocation.
☐	**Supporting infrastructure planning**
	Examples: mobility patterns, traffic flow, energy/water planning.
☐	**Protecting the environment**
	Examples: air quality, deforestation alerts, climate risk mapping.
☐	**Promoting economic development**
	Examples: small-business insights, labor market needs.

☐	**Preserving cultural or Indigenous knowledge**
	Examples: documenting practices, safeguarding ecological knowledge
☐	**Enhancing safety or emergency response**
	Examples: disaster alerts, evacuation planning, heatwave response
☐	**Other (please specify):**

Reflection Questions (for participants to fill in):

What problem should data reuse help solve?
Who should benefit the most from data reuse?
What positive outcomes do you hope to see?

1.2 SCOPE OF ACTIVITIES

Main Question:

Which data reuse activities are planned, and have these activities and their objectives been clearly explained? What remains unclear?

Explain to participants:

"This helps us identify whether people have enough information to meaningfully participate and give informed views."

Clarity Check (tick all that need explanation):

☐	**What data will be collected or reused**
☐	**Who will have access to the data**
☐	**How the data will be analyzed**
☐	**Which technologies will be used (AI, sensors, etc.)**
☐	**How results will be communicated**
☐	**What safeguards are in place**
☐	**Who benefits and who might be affected**
☐	**Timeline for project activities**

What is clear so far?
What still needs clarification?

1.3 LIMITS & RED LINES

Main Question:

Which data uses, activities, or outcomes should be considered unacceptable under any circumstances?

Explain to participants:

"This section captures community boundaries—what people consider unacceptable, harmful, or beyond what they consent to."

Select all community red lines:

☐	**Commercial exploitation of community data**
☐	**Surveillance or monitoring of individuals**
☐	**Use by law enforcement or intelligence agencies**
☐	**Political profiling or targeted influence**
☐	**Training AI systems without permission**
☐	**Sharing data with third parties without approval**
☐	**Identifying individuals or groups without consent**
☐	**Combining datasets in ways that increase risk**
☐	**Any use that could stigmatize the community**
☐	**Other:**

Explain why these are unacceptable:

SECTION 2

WHAT Data Assets and Protections Needed

Identify what types of data may be reused and what constraints should apply.

2.1 DATA TYPES

Main Question:

What types of data may be reused, and should personal vs. non-personal data be governed differently?

Explain to participants:

"We want to understand which types of data the community is comfortable using, and which require stricter rules or should not be reused at all."

Select all that apply:

☐	**Health data**
	Clinic visits, immunization, disease trends.
☐	**Mobility data**
	Travel patterns, transit usage, GPS traces.
☐	**Environmental data**
	Air quality, water levels, pollution.

☐	**Financial data**
	Income, transactions, household economics
☐	**Linguistic/text data**
	Speech patterns, oral histories, transcripts.
☐	**Biometric data**
	Fingerprints, facial data, DNA.
☐	**Other (please specify):**

2.2 SENSITIVE OR INAPPROPRIATE DATA

Main Question:

Which types of data should be restricted or off-limits due to cultural, ethical, or privacy concerns?

Explain to participants:

"This helps identify what the community considers too sensitive, private, or culturally inappropriate to reuse."

Select all that apply:

☐	**Sacred or traditional knowledge**
☐	**Biometric or genetic data**
☐	**Geospatial data linked to sacred/protected lands**
☐	**Identifiable health or financial records**
☐	**Data that could stigmatize groups**
☐	**Other:**

Why should these be restricted?

2.3 GRANULARITY & ANONYMIZATION

Main Question:

At what level of aggregation should data be shared, and under what conditions?

Explain to participants:

"This helps decide whether data should only show group-level patterns or whether individual-level information is ever acceptable with safeguards."

Select preferences:

☐	**Aggregated data only**
☐	**Individual-level allowed with strict protections**
☐	**Individual-level not allowed**

If individual-level is allowed, under what conditions?

SECTION 3

WHO
Actors, Roles, and Responsibilities

Clarify who should steward, access, and oversee data, and how accountability and redress should work.

3.1 DATA CUSTODIANSHIP

Main Question:

Who should be responsible for managing and safeguarding the data?

Explain to participants:

"This question helps decide who the community trusts to hold, manage, and protect the data over time."

Select all that apply:

☐	**Government agency**
☐	**Local community organization**
☐	**Community data trust**
☐	**Independent third-party steward**
☐	**Other:**

Why is this option preferred?

3.2 ACCESS RIGHTS

Main Question:

Who should be allowed to access the data, for what purposes, and under what conditions?

Explain to participants:

"We want to clarify who can use the data and make sure access matches public-interest goals and community expectations."

Specify access by group:

☐	**Researchers**	
	Conditions:	
☐	**Government agencies**	
	Restrictions:	
☐	**Private companies**	
	Oversight required:	
☐	**Community members**	
	Level of access:	

Additional notes:

3.3 THIRD-PARTY USE & DERIVATIVES

Main Question:

Under what conditions, if any, may third parties reuse the data or create derivative products?

Explain to participants:

"This helps determine whether others can create new datasets, models, or tools from the original data, and what rules should apply."

Select all that apply:

☐	**Third-party use allowed with approval**
☐	**Derivatives allowed with conditions**
☐	**Attribution required**
☐	**Revenue-sharing required**
☐	**Synthetic data allowed**
☐	**AI model training allowed only with consent**
☐	**No third-party or derivative use allowed**

Conditions or requirements:

3.4 ACCOUNTABILITY & REDRESS

Main Question:

How should misuse, violations, or breaches be addressed, and who should be responsible for oversight and redress?

Explain to participants:

"This helps define how the project will be monitored and what happens if rules are broken."

Select mechanisms:

☐	**Independent audits**
☐	**Community or participatory oversight board**
☐	**Complaint or grievance mechanism**
☐	**Legal enforcement or penalties**
☐	**Public reporting of violations**

Additional expectations:

SECTION 4

HOW Operational Strategy and Governance

Define how data will be accessed, protected, governed, and monitored in practice.

4.1 ACCESS & SHARING MECHANISMS

Main Question:

How should data be accessed or shared?

Explain to participants:

"This clarifies how data will be made available in ways that protect people while enabling responsible use."

Select preferred approaches:

☐	**Secure environment (no downloads)**
☐	**Tiered or restricted access**
☐	**Community-led cooperative or data trust**
☐	**Open access (where appropriate)**
☐	**Other:**

Notes:

4.2 PRIVACY & SECURITY MEASURES

Main Question:

What privacy and security protections should be applied?

Explain to participants:

"This helps ensure the data is protected against misuse, breaches, or unintended harm."

Select all that apply:

☐	**De-identification or anonymization**
☐	**Encryption**
☐	**Strong access controls**
☐	**Privacy-enhancing technologies**
☐	**Monitoring and access logs**
☐	**Other:**

Additional requirements:

4.3 TRANSPARENCY & OVERSIGHT

Main Question:

What reporting, monitoring, or participatory oversight mechanisms should be in place?

Explain to participants:

"This ensures the community can see how the data is used and hold the project accountable."

Select all that apply:

☐	**Regular audits**
☐	**Public reporting**
☐	**Participatory oversight board**
☐	**Monitoring dashboards**
☐	**Dispute resolution process**

Notes:

4.4 OWNERSHIP & INTELLECTUAL PROPERTY

Main Question:

Who should retain ownership of the data and resulting outputs, and how should usage and intellectual property rights be managed?

Explain to participants:

"This clarifies ownership, attribution, and whether commercial or external use is allowed."

Select preferences:

☐	**Community retains full ownership**
☐	**Shared ownership under agreed terms**
☐	**Licensing allowed with conditions**
☐	**No commercial use**
☐	**Attribution required in all uses**

Specify conditions:

SECTION 5.

WHEN
Duration, Retention, and Review

Set expectations around how long data may be used, when it should be reviewed, and how permissions can change over time.

5.1 DURATION OF USE

Main Question:

For how long may the data be used, and under what conditions can permission be paused or revoked?

Explain to participants:

"This helps determine whether data use is time-limited and how control can be exercised over time."

Select options:

☐	**Fixed period:**
☐	**Renewable with community approval**
☐	**Revocable at any time**
☐	**Ongoing with regular review**

Notes:

5.2 RETENTION & DELETION

Main Question:

Under what conditions should data be archived, deleted, or returned?

Explain to participants:

"This ensures data is not kept longer than necessary or against community wishes."

Select all that apply:

☐	**At project completion**
☐	**After a defined time period**
☐	**Upon community request**
☐	**After a breach of terms**
☐	**Archive for approved future use**

Notes:

5.3 REVIEW & RENEWAL

Main Question:

How often should the terms of data use be reviewed, and what events or changes should trigger a review?

Explain to participants:

"This allows the rules to adapt if risks, technologies, or project goals change."

Select triggers:

☐	**Scheduled (e.g., annual)**
☐	**Technology or risk changes**
☐	**Project expansion or new use cases**
☐	**Community request**

Notes

5.4 EVALUATING BENEFITS & RISKS

Main Question:

How should project impacts be evaluated over time, and which indicators matter most?

Explain to participants:

"This helps track whether the project is delivering benefits and avoiding harm."

Select indicators:

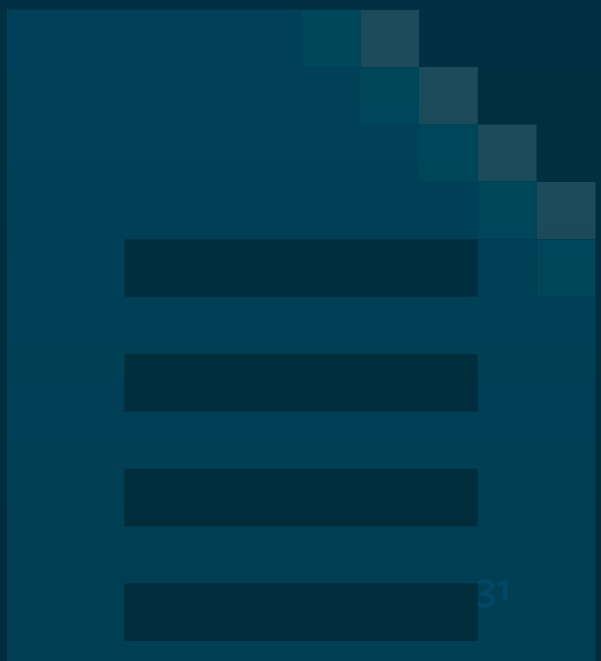
☐	**Social benefits**
☐	**Economic impacts**
☐	**Environmental outcomes**
☐	**Capacity-building or skills**
☐	**Equity and inclusion**

Notes:

SECTION 6.

WHERE
Jurisdiction and Context

Address where data is stored, how cross-border use is handled, and which legal and cultural norms apply.



6.1 DATA STORAGE LOCATION

Main Question:

Where should data be stored, and which jurisdiction should govern it?

Explain to participants:

"Where data is stored affects legal protections, community control, and oversight."

Select preferences:

☐	**Local storage only**
☐	**National storage**
☐	**Cross-border storage with conditions**
☐	**Approved cloud jurisdictions**

Notes:

6.2 CROSS-BORDER MOVEMENT

Main Question:

Under what conditions may data or its derivatives be transferred across borders?

Explain to participants:

"This clarifies whether data can move across borders and what safeguards are needed."

Select options:

☐	**Not allowed**
☐	**Allowed only with approval**
☐	**Allowed for research purposes only**
☐	**Only derivatives may be transferred**

Notes:

6.3 LEGAL & REGULATORY COMPLIANCE

Main Question:

Which laws, rights frameworks, or community protocols must the project follow?

Explain to participants:

"This ensures legal compliance and respect for community governance rules."

List applicable frameworks:

6.4 CULTURAL & CONTEXTUAL SENSITIVITY

Main Question:

Which cultural norms, languages, and governance traditions should guide project decisions?

Explain to participants:

"This ensures the project aligns with local values and decision-making practices."

Cultural norms to respect:
Linguistic considerations:
Governance traditions:

APPENDIX:

Facilitator Cheat Sheet

For facilitator use only (not for participants)

Facilitator Cheat Sheet

For facilitator use only (not for participants)

Purpose

This tool supports meaningful community engagement around data reuse. Its goal is to:

- Surface community values, preferences, boundaries, and conditions
- Identify red lines early
- Create a documented basis for legitimacy and trust
- If core concerns cannot be addressed, the correct outcome may be not to proceed.

What to Listen For

- Capture and distinguish between:
- Agreements (shared expectations)
- Disagreements (conflicting views)
- Uncertainty (lack of clarity or confidence)
- All three are valuable governance inputs.

How to Use the Workbook

- Introduce each section in plain language
- Read the ***Main Question*** aloud
- Use the ***Explain to participants*** text verbatim or paraphrased
- Let participants respond before offering examples
- Capture answers in participants' own words

Do not rush to consensus.

Key Facilitation Principles

- Use plain language; avoid technical jargon
- Encourage reasoning, not just yes/no answers
- Make space for dissent and discomfort
- Do not correct or reframe participant views
- Be honest about what is negotiable and what is not

Your role is listener and recorder, not advocate.

Red Flags to Flag Immediately

Escalate if you hear:

- Clear opposition to the project's purpose
- Non-negotiable red lines that conflict with project design
- Fear of harm, surveillance, or exploitation
- Mistrust based on prior experiences

These require follow-up, not reinterpretation.

Documenting Inputs

When recording responses:

- Use participants' own words
- Avoid summarizing too early
- Flag strong conditions or red lines clearly

After the session, translate inputs into:

- Requirements
- Preferences
- Open questions

Managing Power Dynamics

Watch for:

- Dominant voices
- Deference to authority figures
- Silence around sensitive issues (e.g. surveillance, law enforcement, commercialization)

Helpful techniques:

- Round-robin sharing
- Small group breakouts
- Written or anonymous inputs

Closing Question (Optional)

If time allows, ask:

"Based on everything discussed today, do you feel this project should proceed? If so, under what conditions?"

This often synthesizes the entire discussion.

When to Re-Engage the Community

Re-engagement is recommended if:

- New data types are introduced
- New actors request access
- Technologies change
- Scope expands
- Risks or harms emerge

A FACILITATORS GUIDE AND WORKSHEET

Operationalizing a Social License for Data Re-Use

Open.Data.Policy.Lab

GOVLAB