Reimagining Data Governance for AI: Operationalizing a Social License for Data Reuse

April 2025
No. 78

Technical Reports

Authors Stefaan G. VERHULST (The Data Tank and The GovLab)
Adam ZABLE (The GovLab)
Andrew ZAHURANEC (The GovLab)
Peter ADDO (Agence française de développement)

AFD éditions
AGENCE FRANÇAISE DE DÉVELOPPEMENT

Reimagining Data Governance for AI

Operationalizing a Social License for Data Reuse

Authors

Stefaan G. VERHULST
The GovLab, The Data Tank, & New York University

Adam ZABLE
The GovLab

Andrew J. ZAHURANEC
The GovLab

Peter ADDO
Agence française de développement

Abstract

Artificial intelligence offers great opportunities to drive inclusive and sustainable development—expanding access to healthcare, improving the allocation of scarce resources, and providing education at scale. Yet realizing this potential depends on vast amounts of data, much of it sourced from individuals and communities who often have little say in how it is used or reused. While this data fuels innovation, current governance frameworks frequently fall short in ensuring that those most affected can meaningfully inform decisions or share in the benefits of data reuse. Traditional tools such as individual consent and static licensing models are poorly equipped to manage the complexities of modern data ecosystems, especially in AI contexts where data is continually repurposed and recombined.

This report re-imagines data governance by focusing on social licensing. Social licensing provides a participatory framework through which communities can collectively define, document, and—where possible—enforce conditions for how data about them is reused, particularly in AI development. It shifts decision-making power away from extractive data practices and toward community-centered governance.

The report operationalizes social license for data re-use through a practical Social Licensing Toolkit which includes tools for facilitating community engagement, structuring enforceable agreements, and identifying emerging methods for ensuring compliance. Together, these tools are designed to help development agencies, policymakers, and civil society actors embed social licensing principles into their data practices.

In addition to outlining the social licensing process, the report briefly reviews existing governance mechanisms—consent, data provenance, licensing, and participatory AI approaches—and explains why they fall short in addressing the power asymmetries and governance gaps in AI. It concludes with recommendations for piloting, scaling, and institutionalizing social licensing in development contexts. By re-imagining data governance as a collective, ongoing process, this report offers a path toward more inclusive, transparent, and equitable AI systems that reflect the values and priorities of the communities whose data makes them possible.

Acknowledgements

We would like to thank all of the participants of our Interactive Studio Sessions for their engagement and insights: Jen Dumas, François Sabot, Bogdana Rakova, Mélanie Dulong de Rosnay, Tapiwa Ronald Cheuka, EM Lewis-Jong, Ramya Chandrasekhar, Paul Riviere, José Antonio Guridi, Charlie Martial Ngounou, Viraj Samir Desai, Anais Bayrou, Alek Tarkowski, Ambassador Bitange Ndemo, Melissa Omino, Eve Wambui Njiri, Ian Thomson, Renata Avila, Natascha Gerlach, Jim Fruchterman, Doreen Abiero, Caroline Running Wolf, and Dorcas Chebet Koros.

We are also grateful to the peer reviewers who provided invaluable input on a semi-finalized draft: Thomas Linder, Eloïse Gabadou, Jörg Hoffman, Kaveh Bazargan, Ana Catarina de Alencar, Andrea Duechting, Renato Berrino Malaccorto, Katy Mckinney-Bock, Sherman Kong, Marieke Göttsch, Marco Marini, François Sabot, and Angkana Whiley.

Your participation and contributions have significantly

enhanced the quality of this work, and we express our sincere gratitude for your assistance. Additionally, this paper would not have been possible without the support and expert guidance from our collaborators at the Agence Française de Développement. In particular, we would like to thank Amandine Kashani-Poor, Alexis Frémeaux, Laura Landrein, and AnastesiaTaieb.

A note on how we used generative AI

To improve the efficiency, clarity, and accessibility of our work, we used generative AI (in particular ChatGPT) in the following ways:

- Language refinement: Assisting with grammar, clarity, and readability, particularly in early drafts.
- Idea organization and brainstorming: Exploring alternative ways to structure the report within space constraints.
- Summarization of background materials: Synthesizing source materials to inform literature reviews and situate our research within the broader scholarly landscape.

In all cases, AI-generated output was subject to careful human review and revision. Nothing was used verbatim without critical evaluation and editing by our research team. We did not use AI to generate original findings, conduct analysis, or draw conclusions. All substantive content was developed by researchers with relevant expertise and subject to a rigorous open peer review process. We did not use AI to fabricate, manipulate, or misrepresent data, nor to replace researcher voice, judgment, or intellectual contribution.

Keywords:
Agency, AI Governance, Asymmetries, Citizen Participation, Consent, Data Governance, Data Reuse, Data re-use, Development, Digital Self-Determination, Open Data, Participatory Framework, Public Engagement, SDGs, Sustainable Development Goals, Social License, Social License Framework, Use Cases

Areas:
Global focus with emphasis on low- and middle-income countries.

Glossary of Terms

Community: A group bound by geography, culture, values, or interests, directly or indirectly affected by data collection and reuse.

Consent: A voluntary, informed, and revocable agreement by individuals or communities to data collection, use, or sharing under specific conditions (Longpre et al., 2024).

Data Provenance: The documentation of a dataset's origin, history, and transformations, preserving transparency, accountability, and auditability (Longpre et al., 2024).

Data Reuse: The use of collected data for purposes beyond the original intent for which it was collected.

Data Sharing: The controlled practice of making data available to other parties for collaboration, transparency, and innovation, governed by legal, technical, and organizational safeguards (Mucci, 2024).

Data Sharing Agreement (DSA): A written agreement that establishes the terms for how data is exchanged between parties; important for establishing accountability and trust (The GovLab, n.d.)

Digital Self-Determination: The principle that individuals and communities retain agency over their digital data across its lifecycle, ensuring rights, interests, and preferences are respected (Verhulst, 2023)

Low-and-Middle-Income Countries (LMICs): Nations classified by the World Bank as having a gross national income per capita below a defined threshold (Hamadeh et al., 2023)

Social License: The ongoing acceptance and approval of an organization's operations by stakeholders and the public, beyond legal compliance, based on trust, legitimacy, and alignment with societal values (Kenton, 2024)

Social License for Data Reuse: A participatory governance framework that enables communities to set enforceable conditions for how data about them is reused, balancing ethical, legal, cultural, and practical considerations.

Stakeholders: All entities with a vested interest in data governance, including data holders, users, affected communities, policymakers, civil society, and the private sector.

Executive Summary

As data increasingly powers artificial intelligence (AI) applications, communities in low- and middle-income countries (LMICs) often find themselves excluded from decisions about how their data is collected, reused, and monetized. These communities also receive little of the value their data generates. Existing data governance frameworks—centered on individual consent—have proven inadequate in addressing these systemic inequities. This report addresses the agency asymmetry challenges related to data reuse in LMICs, which have become even more pronounced at a time of AI. It seeks to re-imagine data governance and suggest social licensing as a pathway forward.

The Challenge: Structural Inequities in AI Governance

AI technologies offer opportunities to accelerate development, but existing governance structures concentrate data control in the hands of a few actors—primarily technology companies and institutions based in high-income countries. This imbalance limits the ability of LMICs and marginalized communities to shape how data from and about them is used. Several governance challenges exacerbate this imbalance:

- **Limited Agency and Representation:** Communities often have no meaningful influence over how their data is collected, shared, and reused. Data governance decisions are centralized, with limited transparency and almost no participation from those most affected.
- **Unequal Benefit Distribution:** While marginalized communities generate valuable data, they lack the bargaining power to demand fair benefit-sharing. AI-driven innovations built on their data often deliver economic and social returns elsewhere.
- **Exclusion and Bias:** AI models trained primarily on data from wealthier countries often fail to work effectively in LMIC contexts—further excluding these communities from the benefits of AI while exposing them to the risks of biased or harmful algorithms.
- **Weak Governance Infrastructure:** Many LMICs lack the regulatory capacity, technical expertise, and institutional frameworks necessary to establish robust, contextually appropriate data governance—making them dependent on foreign technologies and governance models that do not reflect their needs or realities.

Why Current Approaches Fall Short

Traditional governance tools—especially consent-based frameworks—have serious limitations in the context of AI and large-scale data reuse. While consent aims to give individuals control over their data, it does not account for the collective nature of many datasets (such as environmental, health, or community-generated data) or the ways data flows across global AI ecosystems long after initial collection. Consent mechanisms tend to be transactional, static, and focused on individual decision-making—failing to reflect the ongoing, collective governance needs of communities.

Consent frameworks also put unrealistic burdens on individuals, requiring them to understand highly complex data ecosystems and predict future uses of their data at the point of collection. These models were not designed for AI contexts, where data can be recombined, repurposed, and reinterpreted in unpredictable ways.

Social Licensing for Data Reuse: A New Governance Tool

To address these shortcomings, the report operationalizes **social licensing for data reuse**—a governance framework that allows communities to collectively define the terms, purposes, and conditions under which data about them is reused. A social license for data reuse is meant to embed community preferences into enforceable agreements that help ensure data governance reflects social, ethical, and cultural considerations alongside legal requirements.

Social licensing does not replace consent—it complements it. While consent focuses on individual agreement at the point of data collection, social licensing creates a process for communities to shape and oversee ongoing data reuse. This is especially important for data that reflects collective interests, such as environmental monitoring data,

community health data, and cultural knowledge. Together, consent and social licensing form a layered governance system, balancing individual rights with community-level governance.

Key Contributions: The Social Licensing Toolkit

The report introduces a practical **Social Licensing Toolkit** designed to help development agencies, governments, and practitioners apply **social licensing for data reuse**. The toolkit supports a three-part process:

1. **Establishing Preferences:** Facilitating participatory engagement to document community-defined terms and conditions for data reuse.
2. **Documenting Preferences:** Translating those preferences into **data sharing and use agreements** that embed community priorities into binding terms.
3. **Enforcing Preferences:** Exploring ways to make sure those agreements maintain compliance in practice, including legal mechanisms, community oversight, certification programs, and emerging technologies like automated compliance tools.

While the report introduces these three steps as essential to social licensing, the toolkit itself only directly develops the first two—establishing and documenting preferences. The third step, enforcement, is identified as a critical area for future research, as the tools provided do not yet include direct mechanisms for monitoring and ensuring compliance.[1] The toolkit also includes a set of scenarios which offers real-world examples to help practitioners apply the questionnaire and sample clauses in context, demonstrating how social licensing could work across different data types and governance settings.

Recommendations for Scaling Social Licensing

To make social licensing scalable and sustainable, the report recommends:

- **Piloting social licensing across sectors** to refine the approach and generate evidence on what works.
- **Innovating compliance mechanisms** by combining legal contracts, community-led oversight, certification programs, and technology-enabled monitoring to ensure adherence to community-defined terms.
- **Building capacity** among communities, policymakers, and practitioners to engage meaningfully in participatory processes that seek to establish preferences and conditions for data re-use.
- **Establishing a Center of Excellence** to provide ongoing technical support, share best practices, and conduct further research on participatory data governance.
- **Embedding social licensing into global development programs, funding criteria, and procurement policies** to encourage responsible data practices at scale.
- **Advance research on social licensing** to generate evidence and refine models.

[1] Enforcement can take different forms across jurisdictions, and this report takes a holistic perspective rather than focusing on a specific legal system's enforcement model. In some legal traditions, enforcement applies to contractual clauses within agreements, while in others, it involves broader legal, regulatory, and ethical mechanisms governing AI data use. Although it may appear that we conflate these perspectives (e.g., common law vs. contract-based continental law approaches), our intent is to present general principles rather than a jurisdiction-specific legal framework. The realization of a social license may take various legal forms depending on context, and this report is not intended as a legal guide but rather a framework for embedding community-defined governance principles into enforceable data practices. Further research is needed to develop scalable enforcement models that balance legal, community-driven, and technological approaches.

How to Navigate This Report

This report provides a framework for integrating social licensing into data reuse governance, aligning AI and data-driven initiatives with community-defined preferences and ethical standards. It is designed for development agencies, government officials, legal experts, researchers, facilitators, community representatives, and data practitioners involved in data governance, AI policy, and participatory engagement. Readers can focus on sections most relevant to their expertise:

For Development Agencies and Policymakers

If you fund, regulate, or guide ethical AI and data reuse, focus on:

- **Executive Summary:** Key findings and policy implications.
- **Chapter 1 (Introduction):** AI's role in global development and existing governance gaps.
- **Chapter 4 (Operationalizing Social Licensing):** Practical steps for integrating social licensing into governance.
- **Chapter 5 (Recommendations):** Actionable policy and funding strategies.

For Facilitators and Practitioners

If you engage communities and translate their input into governance agreements, focus on:

- **Chapter 4.1 (Establishing Preferences and Expectations):** Step-by-step instructions for running participatory processes.
- **Chapter 4.2 (Documenting Preferences and Expectations):** Overview of the toolkit.
- **Appendix A:** Full Social License Questionnaire to guide community conversations.
- **Appendix B:** Sample contract clauses for social license-compliant data agreements.
- **Appendix C:** Example scenarios showing how to apply the toolkit in real world contexts.

For Researchers

If you study **participatory AI governance, data rights, or responsible data use**, focus on:

- **Chapter 5 (Recommendations):** Discusses key research gaps.
- **Appendices A, B, & C:** Tools for guiding and formalizing community data governance.

Introduction

AI and other data-driven technologies are reshaping global development, offering new solutions across sectors such as healthcare, education, agriculture, and infrastructure. AI-driven tools can optimize resource distribution, enhance public service delivery, and improve decision-making (The Alan Turing Institute, n.d.; IEEE Public Safety, 2024). These innovations hold the potential to improve lives, especially in LMICs, where improved efficiency can help overcome resource constraints.

However, realizing AI's potential requires accessing and reusing vast amounts of high-quality data to train the models—much of which is generated by individuals and communities who often have little influence over how that data is collected, shared, used, or reused. In many cases, data governance frameworks do not reflect local priorities, offer limited transparency, and provide few mechanisms for accountability. These gaps can lead to exploitative data practices, reinforcing global power asymmetries and limiting the benefits AI provides to those contributing to its development.

To ensure AI serves the public interest, governance models must evolve beyond extractive approaches and embed community participation in decisions about data reuse. This report explores how **social licensing** can provide a structured, participatory framework for governing data reuse in AI, ensuring that decision-making power is shared with affected communities and that data practices align with their values and needs.

Focus and Objectives

This report, led by The Governance Lab at New York University's Tandon School of Engineering (The GovLab) in partnership with Agence Française de Développement (AFD), presents a set of practical tools for development agencies, governments, and civil society organizations to operationalize community-centered decision-making in local data reuse initiatives. It responds to the urgent need to empower communities in LMICs to influence how data about them is collected, processed, shared, and reused—particularly for AI systems that increasingly shape decisions affecting their lives.

At the core of this effort is the concept of **social licensing**, a participatory approach originally developed in the extractive industries. A social license represents the collective acceptance of a project or activity based on its perceived legitimacy and alignment with community values, needs, and interests (Thomson, 1999). However, unlike in mining—where a company and a community must coexist in the same physical space, and social license is often tied to visible impacts—data is intangible, its use is opaque, and communities rarely have ongoing relationships with the entities reusing their data.

Traditional social licensing models—focused on securing initial community acceptance—do not fully address the fluid, decentralized, and often global nature of modern data governance. In the context of data and AI, social licensing should function as a continuous governance process, where community preferences are regularly revisited, updated, and reaffirmed to ensure ongoing alignment with evolving local values, new risks, and emerging uses of data (Verhulst et al., 2023). While this report emphasizes the importance of continuous engagement, it focuses primarily on the initial processes and agreements and does not yet provide a full framework for long-term permission management—an area for further research**.**

Box 1: Types of Data Appropriate for Social License

Social licensing is most relevant when data reuse for AI has broad societal implications, requires ongoing oversight, or involves power imbalances that consent-based mechanisms alone cannot address. This includes:

- **Community-Generated Data**: Data from citizen science, participatory mapping, or crowdsourcing.
- **Public Interest Data**: Environmental, health, and infrastructure data.
- **Indigenous and Culturally Sensitive Data**: Data tied to traditional knowledge or cultural practices.

- **Shared and Relational Data**: Data involving multiple stakeholders (e.g., cooperative records, communal land use).
- **AI Training Data**: Large-scale personal, non-personal, or de-identified data used in AI models.

Building on existing work, including Creative Commons' exploration of preference signaling for data reuse and the Data Tank and AFD report *Responsible Data Re-Use in Developing Countries: Social Licence Through Public Engagement*, this report frames preference signaling for data reuse as a process with three core phases (Stihler, 2023; Verhulst et al., 2024b). These phases are explored in greater detail in Chapter 4:

1. **Signaling & Establishing Community Preferences**: Organizations engage communities through participatory, transparent, and context-specific processes to define values, concerns, and priorities, ensuring that governance aligns with local needs and ethical considerations.
2. **Documenting Preferences**: The results of participatory engagement are formalized into enforceable agreements, embedding community-defined governance terms into data sharing structures.
3. **Enforcing the Social License**: Mechanisms are developed to ensure adherence to documented preferences, using context-specific compliance strategies.

We stress that social licensing does not replace legal and regulatory frameworks. Instead, it adds an ethical, social, and community-defined layer to complement existing data governance structures (see Box 2). Social licensing ensures that data reuse aligns with the values, priorities, and cultural contexts of affected communities while reinforcing regulatory compliance.

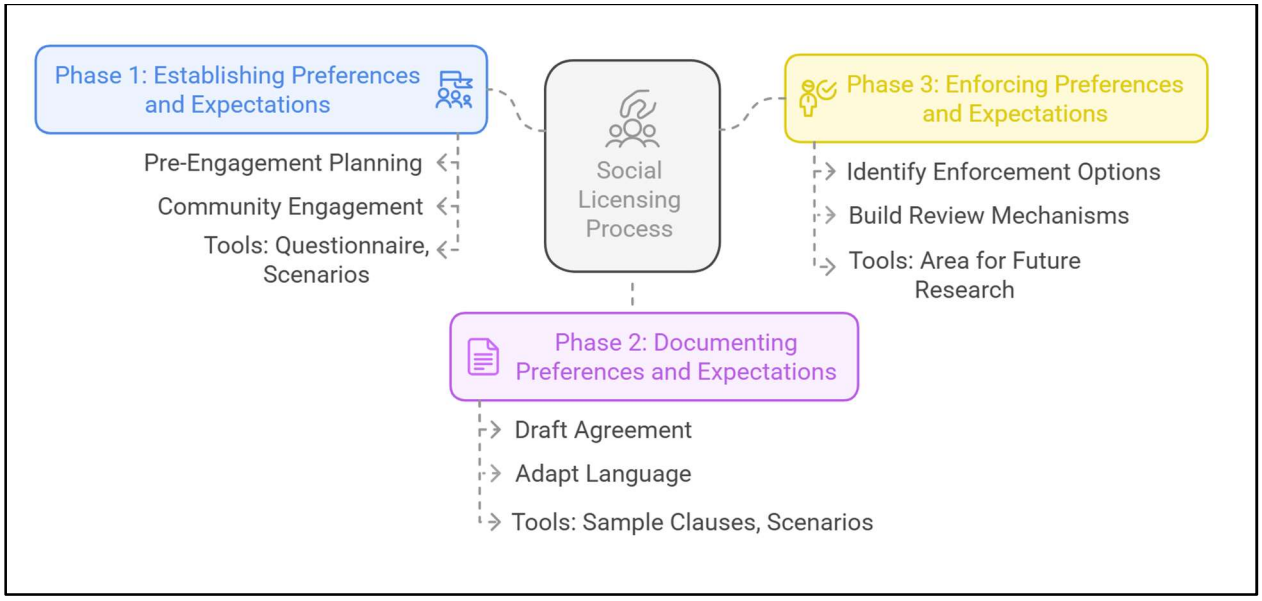
Box 2: Social Licensing and Legal Frameworks – How They Fit Together

- **Social licensing does not replace existing legal frameworks;** it adds a complementary layer.
- **Consent-based frameworks remain critical for personal data;** social licensing enhances governance for collective, community-held, or non-personal data.
- **Legal and social legitimacy work together;** strong licensing processes can reinforce regulatory compliance (e.g., supporting legitimate interest claims by demonstrating meaningful community engagement).
- **Where regulatory frameworks are weak, social licensing can serve as a voluntary governance standard**, raising ethical accountability even without strong legal enforcement mechanisms.

While significant work has been done on how to obtain a social license through participatory engagement, little guidance exists on what a successful social license should look like in practice. Existing models often stop at preference signaling, leaving questions about how to formalize, enforce, and adapt social licenses over time. This report seeks to bridge that gap by introducing tools that help document and operationalize community-defined preferences, ensuring social licensing leads to tangible governance outcomes rather than symbolic agreements. These three tools are:

- **A Questionnaire for Capturing Community Preferences**: Organized around six pillars (Why, What, Who, How, When, and Where), this tool helps facilitate participatory discussions that define data reuse conditions and set governance boundaries.
- **Sample Clauses for Social License-Compliant Data Sharing and Use Agreements**: Adaptable legal templates that translate community preferences into enforceable agreements.
- **Scenarios:** Practical examples illustrating how these tools can be applied across different governance contexts.

Diagram 1: Social License Process: Phases and Tools



Source: The author, with the assistance of Napkin.ai. Original creation.

Methodology

This report builds on earlier research and stakeholder engagement, drawing in particular from the Data Tank and AFD report *Responsible Data Re-Use in Developing Countries*, which introduced a community-led model for data reuse and governance, emphasizing participatory engagement as the foundation for responsible data practices (Verhulst et al., 2024b). To translate this concept into a practical, adaptable framework, we combined three methodological approaches:

1. **Expert Engagement and Peer Review:**
    a. Hosted two expert studios with legal scholars, AI researchers, policymakers, and data governance practitioners to refine governance strategies.
    b. Conducted an open peer review process, incorporating insights from 13+ interdisciplinary specialists to ensure global relevance.
    c. For a full list of contributors, see the Acknowledgments section.
2. **Literature Review**:
    a. Assessed existing approaches to embedding community control into data governance, including data licenses, provenance tools, and participatory AI frameworks.
    b. Identified the need to move beyond *transparency artifacts* (which disclose data practices) to *decision artifacts* (which establish enforceable governance conditions) (Pushkarna & Zaldivar, 2022).
3. **Prototype Development and Refinement**:
    a. Developed practical tools that translate social licensing principles into enforceable agreements for data sharing and reuse.
    b. Designed a structured yet adaptable framework that embeds equity, accountability, and sustainability into AI governance.

Why This Approach?

➔ **Why focus on agreements?** Social licensing is often treated as a vague ethical obligation. This framework ensures participatory outcomes result in actionable agreements rather than mere aspirational commitments.

- ➔ **Why sample clauses?** Standardized agreements reduce transaction costs, improve legal clarity, and rebalance power asymmetries by shifting governance control to data-contributing communities (Global Partnership on AI, 2022).
- ➔ **How does this improve existing models?** This model extends beyond transparency by embedding decision-making power directly into the contractual structure, fostering responsible, systematic, and sustainable data governance.
- ➔ **Why prioritize governance at this stage of the AI lifecycle?** Embedding social licensing at the point of data collection and preparation enables communities to shape how AI systems are built, trained, and deployed before biases or harmful practices become entrenched. Early interventions provide opportunities for communities to define how their data is used, who has access, and under what conditions AI systems are developed. Proactive governance mechanisms align AI development with community priorities before models are finalized and deployed rather than attempting to correct issues later (Verhulst & Schüür, 2023).

1. How the Governance of Data Reuse for AI Perpetuates Inequities

The reuse of data plays a crucial role in the development of AI systems, yet current governance models often fail to ensure fairness, accountability, or community agency. While AI models rely on vast datasets (many sourced from LMICs), affected communities have little control over how their data is repurposed, shared, or monetized. This imbalance exacerbates existing inequalities, as AI systems trained on globally sourced data often benefit a small group of dominant actors while reinforcing extractive data practices (Sharma et al., 2020).

Governance structures for AI data reuse are largely shaped by high-income countries, with limited consideration of the diverse socio-political contexts in which AI operates (Bremmer & Suleyman, 2023). Without mechanisms for meaningful participation, LMICs risk becoming passive data providers rather than active stakeholders in AI development. Moreover, as AI models continuously evolve, data can be reused in ways that extend beyond the initial context in which it was collected, further disconnecting communities from decisions that affect them.

Limited Agency and Centralized Decision-Making

Communities in LMICs often lack opportunities to influence how their data is governed, particularly when it is reused for AI development (Verhulst & Young, 2022). A small group of powerful actors—large technology firms and governments in high-income countries—maintain centralized control over data access, usage, and governance, typically without transparency or meaningful avenues for community engagement (Mulligan & Godsiff, 2023; Osseiran et al., 2023).

Traditional governance mechanisms, such as consent frameworks, do little to challenge this power imbalance. Designed primarily for individual decision-making, consent structures often fail to reflect the collective nature of data governance in many LMICs. Additionally, they rarely account for the evolving nature of AI, where data can be repurposed for new models without further input from the original contributors. Without viable alternative governance mechanisms, communities have little agency over how their data is reused in AI systems that shape public services, economic policies, and social outcomes. These limitations, explored further below, highlight the need for governance models that extend beyond individual consent to ensure meaningful community agency over data reuse in AI.

Exclusion and Unequal Distribution of Benefits

While LMICs generate substantial data that fuels AI advancements—through mobile usage, social platforms, geospatial systems, biodiversity information systems, and labor-intensive work like data labeling—they receive limited benefits in return (Shakir, 2024). Data-intensive industries, including AI development, remain highly concentrated in a handful of countries, creating an economic and knowledge gap between those who provide data and those who profit from it (Waithira et al., 2024; Alonso et al., 2020).

Infrastructure disparities further entrench these inequities. Many LMICs lack the computing power, regulatory frameworks, and institutional capacity to fully engage in AI development (Fan & Qiang, 2024; Stephens et al., 2025). As a result, data governance decisions are often dictated by external entities, leaving LMICs with little leverage to demand fairer terms of data reuse.

Even when communities do participate in AI data governance, their involvement often reinforces extractive dynamics. While collaborating with AI developers can help ensure models are more contextually relevant, it frequently requires sharing data, cultural knowledge, or expertise—contributions that disproportionately benefit external actors rather than local communities (Cardona-Rivera et al., n.d.). At the same time, opting out is not a viable alternative, as exclusion from AI datasets can leave LMICs marginalized in technological advancements. Generative AI models predominantly trained on Western-centric data often produce outputs that are inaccurate, misleading, or even harmful in indigenous and rural contexts (Ghosh et al., 2024; Adams et al., 2023).

Limitations of Existing Data Governance Models for AI

Efforts to govern AI and data use more equitably have largely relied on mechanisms such as consent frameworks, data provenance tracking, licensing models, and participatory approaches. While these strategies aim to enhance transparency, accountability, and community involvement, they fail to address key challenges, including power imbalances, long-term oversight, and enforceability. The following sections examine why these models fall short.

1.1. Consent-Based Mechanisms

Consent remains the dominant approach to governing data collection and reuse. Individuals are typically required to agree to terms through privacy policies, click-through agreements, or explicit opt-in processes.[2] Designed to uphold individual rights and transparency, consent frameworks theoretically enable users to make informed choices about how their data is used and serve as a baseline for accountability, requiring data controllers to obtain explicit permission for data collection and use (Information Commissioner's Office, n.d.). However, in practice, they are poorly suited for AI ecosystems that rely on continuous data reuse, large-scale aggregation, and evolving applications.

Limitations of Consent-Based Governance

- **Unrealistic Burden on Individuals:**
  - Informed consent assumes that individuals can meaningfully evaluate and manage their data choices. However, reading every privacy policy an average person encounters would take weeks annually (Wagner, 2022). Complex language, excessive consent requests, and manipulative design tactics (such as dark patterns) make informed decision-making nearly impossible—particularly in LMICs, where digital literacy and legal resources are often limited (Centre for Information Policy Leadership & BKL, 2024; Cheesman, 2024).
- **Failure to Address Collective and Third-Party Impacts:**
  - Consent mechanisms prioritize individual decision-making, failing to account for the broader social, cultural, and economic consequences of data use. Many forms of data—such as public health records, environmental monitoring data, or geolocation data—have collective implications that cannot be addressed through individual agreements.[3] For example, cybersecurity and fraud prevention depend on cross-referencing data across platforms, but

[2] For a discussion of the differences between the two main approaches to consent in privacy law, see Solove (2024). Notice-and-choice, common in the U.S., presumes consent from inaction or continued use after privacy notices are provided, often failing to capture meaningful agreement. In contrast, the EU's GDPR mandates express, informed, and specific consent, requiring affirmative action from individuals. While the latter is more rigorous, both approaches are deeply flawed.

[3] This does not suggest that communities should override individual rights (e.g., privacy), but rather that governance cannot rely on millions of individuals voluntarily making the same choice.

strict consent requirements can hinder proactive security measures by restricting data processing without explicit user approval (Centre for Information Policy Leadership & BKL, 2024).

- **Failure to Address Information Asymmetries:**
  - AI systems are frequently described as "black boxes," not necessarily due to technical complexity but also because their deployers choose not to prioritize explainability. The core issue is not inherent technical limitations but rather a lack of incentives, proprietary restrictions, and intentional opacity. This creates significant power imbalances, where only those with privileged access can fully understand or regulate AI's impact. Without transparency into system design, decision-making, or bias mitigation, affected communities lack the ability to challenge, negotiate, or influence how these technologies shape their lives (Kim, 2020). Consent mechanisms do not resolve this issue, as individuals cannot meaningfully consent to data use they do not—and often cannot—fully understand.
- **Scalability and Static Nature of Consent:**
  - AI models continuously evolve, integrating new datasets and being repurposed for new applications. However, consent is typically a one-time, binary decision made at the point of data collection. Once granted, consent is often treated as an open-ended license, allowing data to be reused in unforeseen ways without further user input (Frechtling, 2023). Consent frameworks also struggle to scale across the vast datasets and interconnected networks of stakeholders on which AI systems operate, as reauthorization for every downstream use is impractical.
- **Power Imbalances and Lack of True Choice:**
  - Many individuals "agree" to data collection under coercive conditions, where refusing consent means losing access to essential services such as healthcare, education, or digital platforms. While some frameworks attempt to provide granular consent options, they are often complex, difficult to navigate, and lack mechanisms for tracking evolving user preferences (Verhulst et al., 2023). In many cases, consent is reduced to a legal fiction that legitimizes exploitative data practices (Solove, 2024).
- **Weak Regulatory Frameworks:**
  - In many LMICs, weak or absent data protection laws expose individuals and communities to exploitation by corporations, governments, and foreign entities. Without strong legal and compliance mechanisms, local stakeholders have little power to influence global data governance norms. As AI-driven decision-making becomes more pervasive, LMICs risk being further marginalized in economic and policy discussions.

Together, these challenges reduce consent to a checkbox exercise, offering a false sense of agency while failing to protect individuals and communities. As AI systems reshape data governance, consent alone is insufficient—it often results in superficial engagement, prioritizing data extraction over meaningful collaboration or equitable benefit-sharing. Without mechanisms for sustained participation, communities remain vulnerable to exploitation, privacy violations, and governance decisions that do not reflect their needs or values.

1.2. Data Context and Provenance

A key challenge in AI governance is the loss of data context. As datasets are collected, integrated, and reused across multiple applications, their original meaning, limitations, and ethical considerations often become obscured. Provenance—the documentation of a dataset's origins, history, transformations, and applications—is essential to ensuring responsible data use, but it remains difficult to maintain in AI-driven ecosystems (National Network of Libraries of Medicine, n.d.).

The opacity of AI models further amplifies these challenges. Training data is often aggregated from diverse sources without clear lineage, making it difficult to assess bias, prevent model drift, or ensure fair decision-making. As data is repurposed across different jurisdictions and regulatory environments, questions of ownership, licensing, and ethical obligations become increasingly complex.

While efforts to improve data provenance—such as metadata documentation standards, content authenticity techniques, and opt-in/opt-out registries—offer incremental improvements, they share many of the same limitations as consent frameworks. These tools place a high burden on individuals, fail to account for collective governance, and remain difficult to enforce at scale. Without mechanisms for community oversight, compliance, and accountability, provenance tracking alone is insufficient for ensuring responsible AI data governance (Longpre et al., 2024).

1.3. Data Licensing and Indigenous Data Sovereignty

Data licensing frameworks attempt to regulate access and use by specifying legal terms and conditions. Traditional licensing models—such as open-source software and Creative Commons licenses—have been adapted for datasets, but they were not originally designed for the complexities of AI-driven data ecosystems. As a result, these models introduce ambiguities in ownership, create jurisdictional conflicts, and provide limited safeguards against unethical applications (Carroll, 2015).

The transnational nature of AI development complicates enforcement. Many AI models rely on aggregated, cross-border data, making it difficult to ensure compliance with licensing terms or protect community interests (Global Partnership on AI, 2022). Emerging licensing frameworks, such as the Montreal Data License, Responsible AI License (RAIL), Allen Institute for AI (AI2) ImpACT License, and Nwulite Obodo Open Data License, seek to incorporate ethical restrictions and innovative governance and accountability mechanisms (Benjamin et al., 2019; Contractor & Muñoz Ferrandis, 2022; Dumas et al., 2023; Licensing African Datasets, n.d.). However, these approaches remain difficult to scale and enforce, particularly when data providers have limited leverage over AI developers.

Indigenous Data Sovereignty (IDS) presents an alternative model, recognizing data as a cultural and collective resource rather than an individual asset. IDS frameworks emphasize self-determination, cultural preservation, and reciprocity, enabling Indigenous communities to retain control over the collection, use, and application of their data. While this approach offers important protections, IDS frameworks are highly context-specific and not easily transferable to broader AI ecosystems.

1.4. Participatory Approaches in AI and Data Governance

Participatory approaches in AI data governance aim to embed community engagement, deliberation, and decision-making into governance structures. Unlike top-down models that centralize control among AI developers and policymakers, participatory approaches seek to directly involve affected communities in shaping how data is collected, used, and governed (Corbett et al., 2023).

Engagement can take multiple forms, ranging from consultation—where communities provide input but have no formal decision-making authority—to co-decision models, where communities actively shape governance processes and hold enforcement power (Verhulst et al., 2024b). Recent initiatives—including citizens' assemblies, trusted data intermediaries, and Indigenous governance protocols—demonstrate how participatory models can enhance transparency, trust, and social accountability in AI decision-making (Birhane et al., 2022; Nayebare et al., 2023). Some of these models are explored further in Chapter 3.

Despite their potential, most participatory approaches remain limited in impact. Many serve as advisory mechanisms rather than binding governance structures, giving communities a voice but little actual authority. Others lack enforcement mechanisms, making them susceptible to tokenization or neglect once initial consultations conclude. Additionally, most remain small-scale and experimental, with limited adoption in mainstream AI governance frameworks.

1.5. Toward a More Equitable Model for Data Governance for AI

Current governance models for AI data reuse fail to ensure meaningful community agency, equitable benefit distribution, and long-term accountability. Consent frameworks place excessive responsibility on individuals, provenance tools lack enforcement mechanisms, and licensing remains fragmented. While participatory approaches show promise, they often stop short of granting true governance power. These limitations reinforce

existing power imbalances, leaving communities without mechanisms to shape how their data is reused in AI systems that influence public services, economic policies, and social outcomes.

AI governance must move beyond one-time consent and symbolic engagement toward models that embed continuous participation, negotiated decision-making, and enforceable agreements. This shift requires governance structures that respect cultural and social norms, align AI applications with locally identified needs, and hold organizations accountable for data usage. Development agencies, policymakers, and civil society should work to create frameworks that actively include affected communities in data governance processes.

At the core of this transformation is **Digital Self-Determination**, which prioritizes community-driven governance structures that enable informed decision-making and negotiated access to data (International Network on Digital Self-Determination, 2023–2025). Achieving this requires reimagining AI governance to center equity, community control, and social justice. Investing in alternative governance models that challenge extractive data practices and elevate the rights and interests of LMIC communities can enable AI to be harnessed as a tool for collective benefit—rather than one that deepens global inequalities (Benjamin, 2019).

2. Social Licensing as a New Norm and Practice

2.1. What is Social Licensing?

A social license for data reuse offers an emerging alternative to provenance tracking, conventional licensing, and participatory governance frameworks. Instead of relying on one-time consent agreements, social licensing ensures that data reuse aligns with community expectations through ongoing public engagement and negotiated decision-making. This framework shifts governance from individual control to collective agency, embedding transparency, accountability, and democratic principles into data governance (Verhulst et al., 2024b). Social licensing aims to address power imbalances, particularly for marginalized communities, by enabling affected communities to meaningfully participate in decisions about data governance.

The concept of social licensing originated in the extractive industries in the late 1990s, emerging as a response to mounting public backlash against environmentally and socially harmful corporate practices. In 1996, the mining industry faced intense scrutiny following a series of dam failures in the Philippines, which buried a village in toxic waste and displaced hundreds of families (De La Cruz, 2017). In the wake of the disaster, James Cooney, a Canadian mining executive, described the industry's growing need to secure a "social license to operate"—a form of public acceptance extending beyond legal compliance (Gehman et al., 2017). By 2005, 90% of mining industry professionals recognized social licensing as an essential but intangible factor in maintaining community support (Boutilier, 2014).

This perspective reframed corporate responsibility, shifting from a focus on securing regulatory approval to an ongoing obligation to engage local stakeholders, including community organizations, nonprofits, and residents, to maintain and renew public acceptance (Moffat et al., 2016). This approach claimed to encourage deeper engagement with community concerns, particularly regarding the environmental and social impacts of their operations. Advocates argued that sustained, proactive engagement could surface and address concerns early, reducing opposition and providing communities with meaningful channels to voice grievances before resistance escalated. Critics countered that the concept of social license risked becoming a tool for companies to deflect stricter regulation and burnish their reputations without making substantive changes (Owen & Kemp, 2012).

2.2. Challenges and Critiques of Social Licensing

Achieving a social license requires organizations to actively engage with communities in meaningful, transparent, and accountable ways, ensuring that affected stakeholders can determine the extent of their involvement. Whether communities prefer to be informed, consulted, included in deliberations, or granted decision-making authority, the process should ensure inclusive participation while allowing stakeholders to engage safely and constructively.

Superficial or performative efforts—where organizations claim engagement without meaningfully shifting power—risk eroding trust and reinforcing existing inequalities (Meesters et al., 2021).

Since its inception, social licensing has been framed as a tool for ensuring social accountability, particularly in contexts of power asymmetries or weak regulatory frameworks. However, the concept remains highly contested and has faced significant criticism both within and beyond the extractive industries where it first emerged (Mercer-Mapstone et al., 2017; Demajorovic et al., 2019). Some critics see it as an industry-driven ploy, used to ethics-wash harmful practices and manipulate public perception through token engagement efforts. Others argue that social licensing grants excessive power to activist groups, enabling them to unilaterally disrupt legally compliant business operations. Still others question whether a concept so broad and adaptable risks becoming an empty buzzword—invoked to mean everything but delivering little (Breakey, 2023). These criticisms largely fall into two categories:

1. **Social License as a PR Strategy:** A major concern is that organizations treat social licensing as a public relations tool rather than a genuine accountability mechanism. Companies may engage communities only as much as necessary to maintain public approval, reduce operational risks, or avoid disruptions—without truly addressing community concerns (Ihlen & Raknes, 2020). This practice, often referred to as ethics-washing, allows corporations to frame their activities as socially beneficial while continuing extractive or exploitative practices (Breakey, 2023). For example, oil companies claim their work strengthens the welfare state and supports unions, while smart city developers justify mass data collection by citing efficiency and quality-of-life improvements (Mann et al., 2020). In practice, social license has often been used to protect corporate interests rather than empower communities, with engagement efforts designed primarily to mitigate reputational risks rather than meaningfully shift power dynamics (Ali, 2015; Owen & Kemp, 2012).
2. **Social License as a "Rubber Stamp":** Another critique is that social licensing may weaken formal regulations or bypass legally binding agreements. By engaging communities in superficial, irregular, or non-binding ways, organizations can claim to have sought public input while avoiding concrete legal obligations to enact meaningful protections (Meesters et al., 2021). In these cases, communities may be invited to express concerns, but without mechanisms to ensure that institutions implement commitments made, social licensing risks being purely symbolic rather than a genuine tool for accountability and empowerment.

2.3. Establishing a Social License for Data Reuse

A social license for data reuse reimagines traditional licensing models by shifting the focus from legal contracts and individual consent to ongoing public engagement and collective decision-making. Unlike its origins in the extractive industries, where the term has often been criticized as a corporate PR tool, the concept is repurposed as a **participatory governance framework that enables communities to set enforceable conditions for how data about them is reused, in alignment with community-defined values and expectations**. Rather than being a vague or symbolic notion, a social license for data reuse is a process-driven mechanism for fostering trust, agency, and accountability in data collection, sharing, use, and reuse.

Social licensing for data moves beyond transactional consent models that leave individuals and communities disempowered. It introduces participatory governance methods that enhance community agency, build trust between data providers and users, and embed flexibility into data governance. Importantly, a social license does not replace legal or regulatory frameworks. It serves as a complementary governance layer, ensuring that data sharing and reuse reflect ethical, social, and contextual considerations identified by the communities directly impacted. This dual approach—legal compliance alongside community-defined governance—creates stronger accountability, reduces risks of harmful reuse, and ensures data contributes to equitable development outcomes.

2.4. Real-Life Examples of Social Licensing in Data Reuse

While social licensing for data governance is still an evolving concept, several initiatives have already shown that they can successfully align data practices with community values and participatory governance.

1. **NYC Data Assembly:** Launched by the GovLab in collaboration with New York public institutions during the COVID-19 pandemic, this initiative engaged New Yorkers in discussions on responsible data reuse. Participants deliberated on data sharing strategies, with their input helping to shape local policies. The Data Assembly was a type of citizens' assembly—a public engagement forum that, in the context of AI and data governance, brings together demographically representative groups to deliberate on policy and ethical challenges related to data use (Young et al., 2020).
2. **Language Data Commons of Australia (LDaCA)**: An initiative of the Australian Research Data Commons, LDaCA addresses challenges in making language data accessible for research while respecting Indigenous cultural sensitivities. The initiative integrates co-design workshops and public consultations, engaging Indigenous language custodians and cultural heritage organizations. Continuous public testing and consultation will allow the project to remain responsive to evolving community needs, aiming to set a new standard for socially sustainable research infrastructure (Australian Research Data Commons, 2024).
3. **Karya Data Cooperative**: Karya, an Indian data cooperative, challenges traditional data collection models by ensuring fair compensation and ethical data stewardship for marginalized communities. Unlike extractive data models, Karya pays contributors upfront at above-market wages and ensures they retain control over how their data is used. Its cooperative structure enables workers to collectively decide on future data applications and benefit from downstream use. Additionally, Karya integrates local language preservation into its data collection, helping AI systems better reflect linguistic diversity (Douglas, 2022).
4. **Choral Data Trust Experiment:** This UK-based pilot explored ethical data sharing practices for generative AI training, applying a Trusted Data Intermediary model. Conducted with 15 community choirs, the initiative developed a governance framework that allowed contributors to collectively determine licensing terms. Participants negotiated licensing terms via Performance Rights Agreements, Data Rights Mandates, and voting, and engaged in "Data Conversations," expressing a strong preference for collective credit at the choir level rather than individual attribution (Ding et al., 2024).

2.5. Enabling Factors for Social Licensing in Data Reuse

The effectiveness of social licensing for data reuse depends on key enabling factors that enhance credibility, community empowerment, and practical governance outcomes. Strengthening these conditions can increase adoption, improve sustainability, and ensure meaningful participation:

- **Legitimacy**: Communities are more likely to engage when the process is perceived as fair, transparent, and inclusive.
- **Capacity**: Access to technical, legal, and organizational resources supports informed decision-making and active participation.
- **Mutual Benefit**: When both communities and data users see clear value, engagement becomes more collaborative and sustainable.
- **Flexibility & Adaptability**: Social licensing is most effective when it can adjust to different data types, governance needs, and community contexts rather than following a rigid template.
- **Accountability & Compliance**: Clear mechanisms for monitoring, dispute resolution, and adherence to agreements reinforce trust and enforceability.

When these factors are in place, social licensing can evolve from a one-time consultation into a credible, lasting governance mechanism. Without them, it risks becoming symbolic or too difficult to sustain in practice.

2.6. Barriers to Implementation of Social Licensing in Data Reuse

Implementing a social license for data reuse in AI, data governance, and international development presents unique challenges. Unlike traditional social licenses, data governance involves abstract, distributed, and highly technical decision-making. Addressing these complexities is essential to securing legitimacy and long-term trust.

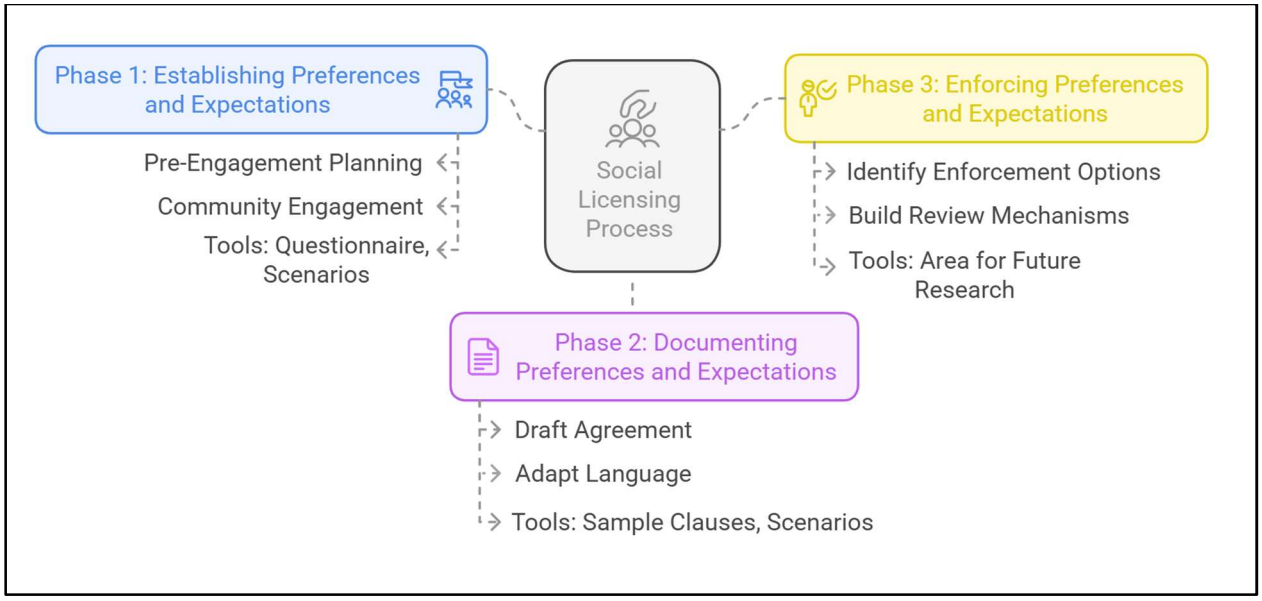
- **Balancing Complexity and Simplification:** Securing a social license requires navigating data processes while remaining accessible to communities with varying levels of technical literacy. Organizations must decide what information to share, how to engage communities, and how many choices to present. Too many options can overwhelm participants, while too few risks excluding critical perspectives. Frameworks should be actionable yet reflective of local priorities, with participatory discussions that explore preferences around data control, conditions, and duration yet do not require deep technical expertise.
- **Defining the "Community" in Data Governance:** Unlike extractive industries, where affected stakeholders typically form a community of place, data governance transcends geographic boundaries (Moffat et al., 2016). Organizations should identify who is directly impacted—data contributors, AI users, or those affected by AI decisions—and clarify who can serve as legitimate representatives for the affected population (Aitken et al., 2020). Equally important is choosing the right engagement channels, which can change depending on the community in question.
- **Scalability:** Public engagement demands significant investments in time, staff, funding, and institutional commitment. Acting on engagement outcomes often requires long-term resource allocation, such as revising policies, retraining staff, or restructuring governance mechanisms. In data and AI governance, these challenges are amplified by the technical complexity of data systems and the high upfront costs of infrastructure, expertise, and sustained community participation. Scaling social license efforts across regions presents additional challenges, as success in one community may not translate elsewhere.
- **Defining When to Update, Renew and Sunset:** In traditional sectors, social licenses are often tied to clearly defined projects with noticeable start and end points. For example, communities can easily observe when a mining operation begins or when it violates agreed-upon environmental protections. However, in data and AI systems, changes are often opaque, making it difficult for communities to recognize when their relationship with data users has shifted. In many cases, people may only realize the impact of data use after harm has occurred, such as discriminatory algorithmic outcomes or unauthorized data repurposing.

  Because of this opacity, maintaining a social license should be an ongoing, adaptive process rather than a one-time exercise. Yet sustaining engagement over time comes with high transaction costs, especially when third-party rights and cross-jurisdictional governance complicate oversight.
- **Avoiding Participation Fatigue:** When communities are repeatedly asked for input without seeing meaningful benefits, engagement deteriorates, distrust increases, and participation eventually ceases (Casal–Ribeiro et al., 2024). Over-engagement can become burdensome, making communities feel that consultation is merely a box-checking exercise rather than a genuine effort to integrate their perspectives.
- **Transnational Issues and Regulatory Complexities:** Cross-border data flows, localization laws, and intellectual property rights complicate social licensing. Countries with data sovereignty requirements impose restrictions on where data can be stored and processed, while exclusive rights in data can trigger specific legal and contractual obligations under international frameworks.
- **Third-Party Rights and Intellectual Property (IPR):** Clarifying third-party rights is complex in social licensing, with many legal regimes for privacy excluding these rights upfront. Unresolved issues include:
  - Data originally collected under one agreement being repurposed under different terms.
  - IP claims (e.g., copyrights, AI patents) complicating enforcement and compliance.
  - Varying contractual interpretations affecting downstream users.
- **Openness vs. Community Control**: A potential tension in social licensing for data lies in its intersection with open data principles. While social licensing aims to ensure community control over data, this can conflict with the goal of open access, particularly in fields that have historically championed data sharing for public benefit.

3. Operationalizing Social Licensing for Data Reuse

Implementing social licensing for data reuse involves a structured, multi-phase process that enables communities to signal their preferences, ensures those preferences are formally documented, and establishes mechanisms for enforcement. The sections below provide a detailed overview of each phase, offering practical guidance for implementation in LMIC contexts.

Diagram 1: Social License Process: Phases and Tools



Source: The author, with the assistance of Napkin.ai. Original creation.

3.1. Phase 1: Establishing Preferences and Expectations

Phase 1 of social licensing focuses on identifying and capturing community-defined preferences for how data should be used, shared, and governed. This phase lays the groundwork for trust and accountability by ensuring that governance decisions reflect stakeholder values, concerns, and conditions for data reuse.

The process includes two key steps:

- **Pre-Engagement Planning:** Organizations define the purpose and scope of engagement, identify relevant stakeholders, and select appropriate engagement strategies.
- **Community Engagement:** Facilitated discussions gather concrete preferences on acceptable data uses, governance boundaries, and oversight mechanisms.

Together, these steps help ensure participation is both procedurally fair (through inclusive and well-structured engagement) and substantively meaningful (by shaping enforceable agreements). **Pre-Engagement Planning** is guided by the Public Engagement Toolkit developed by AFD and The Data Tank (Verhulst et al., 2024b, p. 36), while **Community Engagement** is further supported by the Social Licensing Questionnaire (Appendix A), introduced in this report.

Box 3: Types of Preferences and Expectations

Community preferences and expectations refer to stakeholder input on how data should be reused and governed. These inputs can take different forms, including:

- **Ranking & Prioritization**: Preferred governance models, benefit-sharing mechanisms, and access conditions.
- **Experiential Insights:** Lived experiences, concerns, and values related to data use, trust, and past harms.
- **Open-Ended & Evolving Feedback:** Unstructured input that informs long-term governance adaptations.
- **Governance & Oversight Priorities:** Expectations for transparency, accountability, and compliance.
- **Risk & Harm Considerations:** Concerns about privacy, security, and potential misuse.

3.1.1. Step 1: Pre-Engagement Planning

Before engaging communities, organizations should develop an engagement strategy. Crucial elements include:

1. **Defining the Purpose:** Clarify project objectives, expected outcomes, and how data use aligns with community interests and ethical standards. Consider early on how the social license will be reviewed, enforced, and adapted.
2. **Identifying Stakeholders:** Identify who is impacted, who holds legitimate authority to represent community interests, and how different groups should be included. This should balance specificity (ensuring the right voices are heard) and inclusivity (ensuring broad representation) (Aitken et al., 2020).
3. **Selecting the Type of Public Engagement:** Choose the appropriate level of engagement based on project complexity, risks, and desired outcomes. Levels of engagement include:
    a. **Informing & capacitating**: Providing clear, accessible information to ensure communities understand how data is used and can make informed choices.
    b. **Consultation & feedback**: Gathering insights through polls, surveys, or structured engagements to refine proposals.
    c. **Deliberative discussions**: Holding co-design workshops or citizen assemblies for deeper input into decisions.
    d. **Shared governance**: Establishing mechanisms where communities directly influence decision-making (e.g., through data stewardship models or formalized consent structures).

    Reference frameworks such as the OECD participation spectrum or Fung's Democracy Cube can guide selection (OECD, 2022; Fung, 2006).
4. **Innovative Tools of Engagement:** Tailor engagement methods to the local context—combining in-person and digital formats as needed. Use hybrid approaches where possible to broaden accessibility.
5. **Building on Local Infrastructure and Collaborative Networks:** Embedding social licensing efforts within existing community structures fosters trust and long-term sustainability. Engaging trusted intermediaries, aligning with ongoing development projects, and integrating with local governance frameworks enhance legitimacy, streamline communication, and facilitate negotiation between stakeholders.
6. **Ensuring Information Sharing:** Establish clear channels for two-way communication and ongoing updates. Transparency is key to building trust and sustaining participation.
7. **Engagement Strategy Development and Planning:** Develop a detailed engagement plan that outlines activities, timelines, and responsibilities. Set benchmarks for participation and inclusivity, identify the metrics for impact, and allocate resources accordingly.
8. **Implementation:** Put the engagement plan into action, using the Social Licensing Toolkit (Verhulst et al., 2024b, p. 36) to guide discussions and support co-development of outputs—such as data sharing agreements, consent protocols, or governance frameworks. Facilitators should remain responsive to community feedback and adjust approaches as needed.
9. **Continuous Monitoring and Evaluation:** Establish feedback mechanisms and track key indicators such as participation rates, community satisfaction, and follow-through on commitments. Incorporate processes for lodging complaints, resolving disputes, and making iterative adjustments to ensure the process stays aligned with evolving community expectations.

10. **Reporting and Documentation:** Share progress updates and outcomes with community members and other stakeholders throughout the engagement. Documentation should capture not only preferences but also the rationale behind decisions.

Data governance often involves competing claims over shared datasets, particularly in cases where multiple communities have a stake in language data, cultural knowledge, or environmental records. Engagement should first identify legitimate representatives of all affected groups and attempt to negotiate consensus. If consensus cannot be reached, it may indicate that a social license—or even the engagement itself—is not an appropriate governance model for that data.

Box 4: Continuous Engagement—Why It Matters and What It Could Look Like

Social licensing is not a one-time event but a continuous relationship. Communities must be able to revisit and revise agreements as priorities shift. A future guide could help define:

- **When to review:** Scheduled intervals, important lifecycle events, or triggered by new risks, grievances, or changes in use.
- **What to review:** Purpose, scope, access, terms, benefits, and safeguards.
- **How to review:** From informal check-ins to structured, participatory processes.

3.1.2. Step 2: Community Engagement

Once planning is complete, structured engagement activities begin. These sessions are designed to surface community preferences and governance expectations that will subsequently be documented in formal agreements (see Phase 2).

The **Social Licensing Questionnaire** (Appendix A) supports this step by organizing discussions around six dimensions—Why, What, Who, How, When, and Where. Facilitators should have experience in participatory methods, data governance, and legal frameworks to guide these conversations and ensure informed, inclusive input.

More than feedback, these engagements can produce concrete terms of governance. When well-executed, this phase allows communities to co-determine how their data will be used and reused, under what conditions, and with what protections in place. These preferences become the foundation of subsequent social licensing steps.

3.2. Phase 2: Documenting Preferences and Expectations

Phase 2 of the social licensing process focuses on translating community-defined preferences into formal, enforceable governance instruments. While Phase 1 centers on participatory engagement, Phase 2 is about ensuring that what communities articulate is captured in concrete terms—agreements that guide how data is reused, who can access it, for what purposes, and under what conditions.

To support this, we introduce the **Social Licensing Toolkit**: a set of prototype tools designed to help operationalize social licensing principles by embedding them into data sharing and use agreements. These tools are designed to

promote participatory and adaptable governance—not fixed templates, but flexible instruments that reflect evolving community priorities and allow for ongoing oversight.[4]

The toolkit comprises three components:

1. **Social Licensing Questionnaire** (Appendix A): A facilitation tool for the **Community Engagement** phase of preference signaling. Organized around six core questions—Why, What, Who, How, When, and Where—it guides communities in defining acceptable conditions for data use and governance boundaries.

Diagram 2: Social License Framework Overview

Source: The author, with the assistance of Napkin.ai. Original creation.

2. **Sample Clauses for Social License-Compliant Agreements** (Appendix B): Once preferences are established, these sample clauses help formalize them into enforceable agreements. The language is adaptable to various legal and cultural contexts.
3. **Scenarios** (Appendix C): To show how these tools can be used in practice, Appendix C offers real-world examples illustrating both risks and opportunities across the data lifecycle—from collection and aggregation to AI model training and downstream use. These scenarios can help practitioners anticipate governance challenges and apply social licensing to mitigate harm and enhance value.

It is important to clarify: **the social license is not the resulting agreement itself**. Rather, it is the ongoing, community-granted permission for data reuse under agreed conditions. While a written agreement is one way to document this license, the social license may also take other forms, depending on the context. Still, in settings with significant power asymmetries, codifying the license into a formal agreement strengthens protections for communities and provides a shared reference point for accountability.

The toolkit was developed through an iterative process. Drawing from the social licensing principles first outlined in the *Responsible Data Re-Use in Developing Countries* report by AFD and The Data Tank, those principles were translated into facilitation questions (the **Social Licensing Questionnaire**), distilled into enforceable governance conditions, and then developed into **sample contractual clauses**. The clauses were informed by legal language from the Contracts for Data Collaboration (C4DC) library and tailored to reflect social licensing values and requirements (The GovLab, n.d.).

[4] This version of the toolkit focuses on initial agreements. Mechanisms for managing long-term engagement—such as structured review cycles, renewal processes, or revocation pathways—are not yet fully developed and remain a priority for future research and iteration.

3.3. Phase 3: Enforcing Preferences and Expectations

The final phase of social licensing—ensuring compliance with community-defined terms—is essential for transforming participatory engagement into meaningful accountability. Without mechanisms to uphold the agreed-upon conditions, social licenses risk becoming symbolic rather than binding.

This section outlines preliminary strategies for enforcement that could be adapted across legal systems, organizational types, and data contexts. These strategies combine formal legal instruments, community-based oversight, intermediary models, and emerging technological tools. Though still in early stages, these mechanisms point toward a more enforceable, scalable, and rights-respecting approach to data reuse governance.

Like the rest of the social licensing process, enforcement should be seen as iterative and adaptable—not a one-size-fits-all solution. The tools described here are not prescriptive, but rather building blocks to support communities, policymakers, and practitioners in holding data users accountable and ensuring that data reuse continues to reflect community priorities over time.

Anchoring in Formal Legal Mechanisms

- **Embedding as Contracts:** Social licenses can be formalized through data sharing agreements that embed community preferences into contractually enforceable terms. These agreements, supported by the sample clauses in Appendix B, provide clear obligations and legal recourse in cases of non-compliance.
- **Incorporating into National Laws and Policies:** Governments can institutionalize social licensing by embedding it in national data protection laws, AI strategies, and digital governance frameworks. This could include mandating participatory governance processes for certain high-risk data projects, requiring organizations to demonstrate compliance with community preferences when seeking regulatory approval, or making social licensing a legal precondition for public data access or certain types of cross-border data transfers.
- **Linking to Consent Frameworks**: Social licensing can supplement individual consent by establishing collective conditions for data reuse. For non-personal or community-held data, social licensing can serve as the procedural basis for lawful processing under alternative legal bases such as legitimate interest or contractual necessity (Centre for Information Policy Leadership & BKL, 2024).

  Importantly, embedding social licensing into consent frameworks requires treating consent not as a one-time event but as an ongoing process. This means building in periodic community reviews to reaffirm, revise, or revoke the social license based on evolving community preferences, new risks, or shifting data uses. Such a structure creates dual accountability: enforcement by the community through participatory governance, and regulatory oversight by data protection authorities or other responsible bodies. Aligning social licensing with existing legal processes both enhances compliance and strengthens its legitimacy within formal governance systems.

Community-Based and Collective Governance

- **Local Governance Structures**: Community-appointed representatives or oversight committees can monitor implementation, mediate disputes, and review ongoing alignment with social license terms.
- **Data Stewardship**: Data stewardship is the systematic, sustainable, and responsible management of data to maximize public value while ensuring ethical and legal compliance. Unlike chief data officers and other data managers, data stewards prioritize collaboration and responsible data use over operational oversight. As AI adoption grows, data stewards play a critical role in overseeing data governance across its lifecycle—from collection to reuse, including its integration into AI models. In the context of social licensing for data for AI, they serve several key functions:
  - **Stewarding Data:** Identifying minimum viable data needed to develop AI systems that align with community priorities.
  - **Stewarding Relationships:** Connecting internal teams and external partners to facilitate ethical data reuse and AI development.

    - **Stewarding Resources:** Supporting AI development by securing necessary data, expertise, and financial assets, while also assisting in drafting data sharing agreements.
    - **Stewarding Collaboration:** Institutionalizing partnerships to scale and sustain AI models while ensuring alignment with public expectations.
    - **Stewarding Insights:** Communicating AI development outcomes to ensure public understanding, encourage adoption, and guide future improvements (Verhulst, 2025; Verhulst & Saxena, 2020).
- **Data Stewardship Through Intermediaries:** Intermediary-based data stewardship bridges the gap between data holders, users, and communities, working to ensure ethical, transparent, and participatory governance. These organizations act as trusted facilitators, negotiators, and neutral intermediaries, helping to establish long-term oversight beyond one-off agreements, reduce transaction costs for individual contributors, and formalize accountability in AI data ecosystems.

  Models of stewardship-through-intermediaries include:

    - **Data Trusts:** Legal entities with fiduciary responsibilities that can enforce licensing conditions through contractual agreements and legal protections (Benhamou & Dulong de Rosnay, 2023).
    - **Data Cooperatives:** Member-owned organizations that allow individuals and communities to pool, manage, and share data collectively while retaining control over its use. These cooperatives promote data sovereignty, equitable access, and responsible data sharing for economic, social, and public interest benefits by negotiating terms on behalf of members (Bühler et al., 2023).
    - **EU Data Intermediaries:** Neutral third parties registered under the EU Data Governance Act that facilitate data sharing between individuals, businesses, and organizations without profiting from the data itself (European Commission, n.d.).
- **Data Commons with Access Requirements**: A data commons model could be established in which access to shared datasets is conditional on meeting governance standards defined through the Social Licensing Toolkit. Data commons could incorporate ethical obligations, benefit-sharing mechanisms, and participatory governance structures, ensuring that organizations using the data respect community-defined conditions. These commons could also employ tiered access controls, granting permissions based on intended use, organizational credibility, or adherence to social license principles (Verhulst et al., 2024a).

Certification and Compliance Incentives

- **Ethical AI Training Certification:** A certification model could require organizations to demonstrate compliance with community-defined preferences. Similar to Fairly Trained, which certifies generative AI companies for ethically sourcing training data, a certification system for socially licensed data reuse could involve due diligence processes, regular audits, public compliance disclosure, and market incentives where certification signals responsible data practices, encouraging wider adoption (Fairly Trained, n.d.).
- **Open-Source Compliance Tools and Repositories:** Developing and promoting open-source tools can help communities track, document, and ensure compliance with social licenses. Platforms could include registries for social license agreements, automated monitoring tools for compliance verification, and repositories of best practices for ethical data governance. These resources would facilitate transparency and accountability, making it easier for organizations to align their data practices with social licensing commitments.

Decentralized and Technology-Driven Mechanisms

- **Trusted Digital Agents:** Automated agents could help manage and negotiate access rights, track usage, and negotiate permissions in real time, reducing friction and increasing visibility across data lifecycles (World Economic Forum, 2022).
- **AI as Virtual Stewards**: Assuming AI summarization can extract faithful and meaningful summaries (a contested subject), AI tools could be trained on individual opinions or organizational documentation, in addition to licensing terms. These models could provide automated assessments of proposed data uses and evaluate their alignment with community-defined values and conditions (Revel & Penigaud, 2025).
- **Blockchain-Based Governance Mechanisms:**

- **Smart Contracts and Decentralized Data Storage:** Smart contracts—self-executing agreements that trigger actions when predefined conditions are met—could be used to enforce data sharing conditions automatically by granting or revoking access based on compliance. In addition, they could be integrated with decentralized data storage models. For example, personal data stores, which allow individuals to securely collect, manage, and control access to their personal information, could be governed by a data trust or cooperative, ensuring that data remains protected and shared only under agreed-upon conditions. If an entity breaches the terms of the social license, smart contracts could automatically revoke access. The EU Data Act formalizes the use of smart contracts to automate data sharing agreements, potentially legitimizing this approach within regulatory frameworks (European Union, 2023).
- **Decentralized Autonomous Organizations (DAOs):** DAOs—internet-based entities collectively owned and managed by their members, operating without centralized authority, with rules encoded in smart contracts on a blockchain—could formalize social licensing as a collectively managed framework. DAOs could facilitate deliberative governance through blockchain-based voting mechanisms, allowing communities to securely and transparently decide on data governance rules.

Box 5: Encouraging Voluntary Compliance and Sustainable Adoption

While enforcement mechanisms are essential, long-term sustainability also depends on fostering voluntary compliance. Incentives, behavioral design, and trust-building strategies can drive adoption across sectors:

Social Norms & Behavioral Nudges: Highlighting positive examples of ethical, community-centered data practices can normalize responsible behavior. Default settings favoring transparency, participation, and benefit-sharing can help reinforce these values.

Incentives: Offer tangible benefits—such as preferential funding, procurement eligibility, or reputational rewards—for actors who adopt and maintain social license standards.

Adaptive Change Management: Encourage organizations to integrate social licensing into internal policies and workflows. Capacity building and peer learning can help embed these practices as standard components of ethical data reuse.

4. Recommendations to Advance Social Licensing for Data Reuse

To effectively implement social licensing for data reuse in LMICs, key stakeholders should take coordinated action to fund pilots, develop scalable models, and integrate social licensing into governance frameworks. The following recommendations outline priority actions and the actors best positioned to lead their implementation.

Fund Social Licensing Pilots and Scalable Prototypes

To establish social licensing for data reuse in LMICs, development practitioners can support pilot projects across health, agriculture, education, and finance to generate evidence, refine models, and inform policy adoption. These pilots could also explore applications in stakeholder consultations within government offices and regulatory frameworks, as well as official statistics where social licensing can enhance data accuracy, public trust, and reduce costs by ethically reusing existing datasets.

Who Could Lead?

- Development agencies, multilateral organizations, and philanthropic entities should provide funding and technical assistance for pilot programs.
- Governments could integrate pilots into national data strategies and regulatory processes.
- Academic institutions and civil society could evaluate pilots and extract best practices.
- Private sector partners can incorporate social licensing within data sharing agreements and test scalable methods of implementation.

Key Actions:

- Identify priority areas where social licensing can improve governance and create impact.
- Develop impact metrics to measure effectiveness, support accountability, and refine approaches over time.
- Disseminate lessons learned through policy briefs, case studies, and toolkits to support broader adoption.
- Maintain adaptability to evolving community needs and establish review procedures and grievance mechanisms to allow communities to modify agreements.
- Balance community agency with practical feasibility through lightweight implementation that minimizes administrative burden while maintaining meaningful oversight (Fruchterman et al., 2024).

Explore and Innovate Compliance Mechanisms

Development practitioners should develop flexible, scalable, and contextually appropriate compliance mechanisms that uphold social licensing principles. These mechanisms could balance contractual enforcement, incentives, community governance, and technology-driven tools such as automated compliance monitoring, as outlined above. Some of the most critical include:

a. **Develop Certification Bodies for Social Licensing Compliance**

Independent certification bodies can promote adherence to social licensing principles, enhance trust in data governance, and create incentives for ethical data use. A certification model—similar to Fair Trade standards—can help establish accountability benchmarks and signal compliance to funders, regulators, and the public.

Who Could Lead?

- Independent organizations such as community councils, civil society groups, multilateral regulatory bodies, and specialized certifying entities should oversee certification processes.
- Development agencies can fund pilot certification programs, create global benchmarks, and integrate certification into capacity-building initiatives.

Key Actions:

- Develop a certification framework for organizations, data sharing agreements, or AI models that align with social licensing principles.
- Implement compliance audits and public reporting to strengthen trust and transparency.
- Incentivize organizations by tying certification to funding eligibility, reputational benefits, and preferential procurement opportunities.

b. **Integrate Social Licensing into Procurement and Funding Requirements**

Embedding social licensing principles into procurement policies, funding requirements, and project selection criteria can make compliance a prerequisite for receiving public or institutional funding. This approach creates strong incentives for adoption and accountability.

Who Could Lead?

- Governments could incorporate social licensing into public sector contracts and procurement frameworks.
- Development agencies and multilateral organizations could require compliance in AI and data-related funding and grants.
- Philanthropic organizations and private sector funders can prioritize investment in projects that adhere to social licensing principles.

Build Capacity and Awareness

To support widespread adoption, implementation, and sustainability of social licensing, development practitioners should invest in capacity building and public awareness efforts. Training programs can equip facilitators, community representatives, policymakers, and technical experts with the knowledge needed to apply social licensing principles effectively. At the same time, broad outreach initiatives should drive demand for community-driven data governance and foster informed public engagement.

Who Could Lead?

- Development agencies, universities, and civil society organizations could lead training programs, and embed social licensing into digital literacy, AI governance, and ethical data use education.
- Advocacy groups, civil society organizations, and academic institutions could spearhead awareness campaigns, while governments and multilateral organizations amplify messaging through policy channels.
- Governments and multilateral organizations can integrate training into policymaker education programs to support regulatory adoption.
- Private sector actors can contribute through workforce development initiatives.
- Grassroots movements can mobilize community engagement and advocacy.

Key Actions:

- **Develop Training Programs for**:
  - **Facilitators**: Strengthen skills in participatory engagement, cultural sensitivity, and legal frameworks.
  - **Community Representatives**: Provide resources on digital rights, data privacy, and AI ethics so communities can advocate for their data rights.
  - **Procurement Officers & Policymakers**: Ensure decision-makers can apply social licensing principles in funding, regulatory oversight, and project design.
  - **Technical and Legal Experts**: Expand training to cover best practices in data documentation (e.g., datasheets for datasets) and other legal frameworks referenced in Appendix B.
- **Raise Public Awareness:**
  - Promote social licensing as an ethical standard.
  - Develop accessible materials (e.g., videos, infographics, and case studies) to communicate social licensing principles to diverse audiences.
  - Empower communities to critically engage with technology and shape the data governance systems that affect them.
- **Foster a Community of Practice and Data Stewardship Network:**
  - Establish a global community of practice to connect practitioners, researchers, and policymakers, facilitating collaboration, knowledge exchange, and continuous improvements to the social licensing framework.
  - Create a network of trained data stewards to support communities in understanding the social, legal, and technical implications of data reuse.

- Position data stewards as independent representatives who provide guidance, oversight, and cross-national learning to strengthen community-driven data governance.

Establish a Center of Excellence for Social Licensing in Data Governance

A global hub—or multiple regional hubs—dedicated to advancing research, innovation, and adoption of social licensing principles could be a key driver for responsible and equitable data governance. This Center of Excellence would function as a knowledge hub, technical resource, and advocacy platform, fostering collaboration among governments, private sector actors, civil society, and local communities to ensure context-driven and scalable social licensing frameworks.

Who Could Lead?

- Multilateral organizations (e.g., UN, World Bank, regional development agencies) could provide institutional backing and funding.
- Multinational tech companies managing large-scale data ecosystems could demonstrate leadership by co-creating hubs that promote responsible data reuse.
- Development agencies, academic institutions, and local organizations could collaborate to support broad legitimacy and responsiveness to local contexts.
- Governments could integrate the center's outputs into national data governance strategies.

Key Functions of the Center

- **Technical Assistance:**
  - Provide hands-on support to organizations implementing social licensing.
  - Offer advisory services on embedding social license principles into national and local governance frameworks.
- **Repository Development:**
  - Host a centralized collection of social license-compliant agreements, templates, clauses, and case studies tailored to different data and regulatory contexts.
- **Capacity Building:**
  - Train facilitators, policymakers, and organizational leaders to integrate social licensing into governance and procurement.
  - Provide educational resources on digital rights, data privacy, and AI ethics to equip communities for informed decision-making.
  - Develop certification programs to establish social licensing as a recognized standard in responsible data management.
- **Advocacy and Outreach:**
  - Publicize successful social licensing implementations to drive replication and scaling.
  - Engage in policy discussions to encourage regulatory adoption of social licensing principles.

Partner with Large Players to Promote Adoption

Scaling and institutionalizing social licensing requires collaboration with international organizations, governments and responsible actors from the private sector. These entities often have the reach, resources, and influence to embed social licensing into governance frameworks, industry standards, and procurement policies. They can drive adoption across both public and private sectors by incorporating social license-compliant agreements into data supply chains, AI development models, and global development initiatives.

Who Could Lead?

- Multilateral organizations can integrate social licensing into global development frameworks and funding requirements.
- Governments can embed social licensing into national data laws, procurement policies, and AI governance frameworks.
- Responsible private sector leaders and digital platforms can incorporate social license-compliant agreements into corporate data governance policies.
- Industry alliances and standards bodies can establish technical and legal benchmarks to secure interoperability and compliance.

Key Actions

- **Engage with Multilateral Organizations:**
  - Integrate social licensing into AI governance policies and responsible data use frameworks.
  - Promote social license compliance as a criterion for funded digital development projects.
- **Encourage Private Sector Adoption:**
  - Collaborate with companies and industry groups to integrate social licensing agreements into corporate data governance policies.
  - Provide technical guidelines, case studies, and best practices to facilitate adoption.
  - Establish market-based incentives (e.g., certification programs) to encourage compliance.
- **Addressing Supply Chain Misalignment**:
  - Adopt dynamic data provenance frameworks to track data flows and verify licensing compliance across AI supply chains.
  - Apply social licensing principles consistently across organizational structures and supplier networks, embedding them into internal governance mechanisms.
  - Conduct supply chain audits and publish transparency reports to assess adherence to social license terms and responsible data practices.
  - Promote community-oriented considerations in data reuse partnerships by integrating community voices into the ongoing iteration of social licensing best practices (Partnership on AI, 2021).

Advance Research on Social Licensing

Expanding research on social licensing is critical to refining governance models, strengthening compliance mechanisms, and enabling scalable implementation. Many aspects of social licensing remain underdeveloped, requiring further exploration to create adaptable, enforceable, and community-driven frameworks.

Who Could Lead?

- Academic institutions and research organizations could conduct empirical studies, pilot evaluations, and theoretical development.
- Development agencies and philanthropic organizations can provide funding and infrastructure for long-term research initiatives.
- Civil society organizations and community networks can shape research agendas and validate findings to ensure alignment with local realities.

Key Areas for Future Research

- **Continuous Engagement and Renewal Processes:**
  - Social licensing should be an evolving process rather than a one-time agreement, yet scalable models for ongoing engagement, periodic review, and community-driven audits are still lacking. Research should define triggers for review, methods for re-engagement, and solutions for practical challenges such as revoking or modifying data rights after it has already been shared.
- **Tailored Social License Models:**
  - A single, standardized social license is unlikely to suit all communities and data ecosystems. Future research should explore flexible licensing models with adaptable tiers or customized

provisions, including mechanisms for Indigenous communities to exercise collective governance and long-term data sovereignty, as well as simplified processes for smaller organizations to reduce compliance burdens while maintaining community trust.

- **Social Licensing in Complex, Multi-Stakeholder Ecosystems**:
    - The social licensing model introduced in this report primarily addresses bilateral relationships between a community and a single data user. In practice, data ecosystems involve multiple contributors with competing interests, intermediaries that facilitate access and aggregation, and supply chains that span multiple jurisdictions. Research should explore how social licensing can function in multi-community settings and data ecosystems where governance structures must account for diverse actors, legal frameworks, and evolving data flows.
- **Balancing Social Licensing with Open Data and Public Interest Needs**:
    - Social licensing prioritizes community control, while open data principles emphasize broad access. Future research should examine how these models can coexist and whether licensing structures can be designed to support public interest objectives while preserving community governance, particularly in cases where data has significant societal value beyond its originating community.
- **Metrics and Evidence Base for Impact:**
    - Social licensing remains largely experimental, with limited empirical evidence on its effectiveness. Research should focus on developing measurable indicators to assess its long-term impact, generating case studies, and collecting longitudinal data to refine governance models and support broader adoption.

6. Reimagining Data Governance: Embedding Social License as a New Norm and Practice

As AI and data-driven technologies expand, they present both opportunities and risks, particularly for low- and middle-income countries. While AI has the potential to drive economic development, existing governance systems—built on rigid, individual-focused consent models—fail to address systemic power imbalances. Without deliberate interventions, AI risks reinforcing extractive data practices, widening inequalities, and excluding marginalized communities from decisions that directly impact them.

To address these concerns we need to reimagine data governance. Social licensing offers a community-centered alternative that shifts decision-making power back to the people most affected by data reuse. Unlike static consent models and conventional licensing agreements, it provides an adaptive governance framework that embeds community agency, participatory governance, and ethical safeguards into data use agreements. By ensuring that data contributors are active stakeholders, social licensing helps align AI and data governance with community priorities, responsible practices, and local development goals.

This report has outlined three key steps to operationalizing social licensing:

- **Signaling and Establishing Community Preferences and Expectations**: Engaging communities in participatory processes to define their expectations, priorities, and concerns around data governance.
- **Documenting Preferences into Enforceable Agreements:** Translating community-driven decisions into structured agreements that guide responsible data use.
- **Exploring Future Enforcement Mechanisms:** Strengthening compliance through multiple pathways, including formal legal frameworks, community-led oversight, technical tools, and incentive structures.

Key Recommendations

To implement social licensing for data reuse, stakeholders should:

1. **Fund Social Licensing Pilots and Scalable Prototypes**: Test and refine models to develop lightweight, adaptable frameworks with functional grievance and review mechanisms.

2. **Explore and Innovate Enforcement Mechanisms**: Develop compliance pathways through certification bodies, procurement requirements, and technical tools.
3. **Build Capacity and Awareness**: Train communities, policymakers, and organizations while expanding public engagement.
4. **Establish a Center of Excellence**: Establish a hub for technical support, research, and best-practice sharing.
5. **Partner with Large Players to Promote Adoption**: Collaborate with international organizations, AI industry leaders, and governments to integrate social licensing into global development, AI strategies, and industry governance policies.
6. **Advance Research on Social Licensing**: Generate empirical evidence, impact metrics, and case studies to refine models and ensure effective, long-term governance.

Finally, to realize the promise of AI and data-driven innovation while safeguarding the rights and interests of communities, social licensing must become a common practice embedded in the design of digital governance. By institutionalizing mechanisms that elevate collective agency and community voice, social licensing offers a path toward digital self-determination—particularly for historically marginalized groups. Making social license standard practice not only addresses structural agency asymmetries but also lays the foundation for more inclusive, trustworthy, and equitable digital ecosystems. As governments, companies, and civil society chart the future of AI governance, embracing social license as a foundational norm is both a moral imperative and a practical necessity.

Appendix A: Social License Questionnaire

The **Social License Questionnaire** is a facilitation tool designed to guide discussions during the Community Engagement Phase of Signaling and Establishing Preferences, after organizations have completed Pre-Engagement Planning using the **Public Engagement Toolkit**. 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—should lead dialogues, workshops, or focus groups to ensure informed decision-making. Social licensing requires ongoing, structured engagement and cannot be achieved through a single consultation.

This questionnaire provides a flexible framework for capturing community input that can be documented as enforceable agreements (see **Appendix B: Sample Clauses**). It is not exhaustive but serves as a guide to shape participatory data governance processes. For examples of how this can work in practice, see Appendix C: Scenarios.

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.

1. WHY: Defining Purpose, Scope, and Limitations

This section explores the **purpose of data reuse** and ensures alignment with community priorities.

- **Purpose:** What is the intended goal of the project, and does it align with your community's priorities?
  Examples: Public health, education, infrastructure, environmental protection, economic development, indigenous knowledge preservation.
- **Scope:** Have the objectives and intended outcomes been clearly communicated and justified?
  Considerations: Projects should provide transparent documentation, consult communities early, and allow for meaningful community input before implementation.
- **Limitations:** What activities should be explicitly prohibited when using this data?
  Examples: Commercial exploitation, surveillance, law enforcement use, political profiling, AI model training without consent, privatization of public resources.

2. WHAT: Determining Appropriate Data Assets

This section focuses on ensuring the **types of data** reused align with community standards and expectations.

- **Data Types:** What types of data will be reused, and should different governance models apply to personal vs. non-personal data?
  Examples: Health, mobility, environmental, financial, linguistic, biometric. Personal data often requires stricter protections, while non-personal data may have broader reuse potential.
- **Restrictions:** What types of data are sensitive, off-limits, or culturally inappropriate for reuse?
  Examples: Sacred or traditional knowledge, biometric data, geospatial data tied to protected lands, personally identifiable financial or health records.

- **Granularity and Anonymization:** Should data be aggregated or used at an individual level? Should data be de-identified before reuse?
  Considerations*: Aggregation reduces privacy risks but may limit utility; anonymization should prevent re-identification.*
- **Accuracy & Relevance:** Is the data up to date and high-quality? How should data quality be maintained?
  Considerations: Data provenance tools can ensure that reused data is current, representative, and appropriate for its intended purpose.

3. WHO: Identifying Actors, Roles, and Responsibilities

This section clarifies **who manages, accesses, and oversees the data** to ensure accountability and trust.

- **Data Custodians:** Who should be responsible for managing and safeguarding the data?
  Examples: Government agencies, local organizations, community data trusts, or independent third-party stewards.
- **Access Rights:** Who is allowed to access the data, and under what conditions? Should different levels of access be granted based on purpose and stakeholder type?
  Considerations*: Researchers may have broader access for public-interest studies, while private companies may need stricter oversight. Governments may have conditional access depending on policy frameworks.*
- **Third-Party Use & Derivatives:** Can third parties create derivatives based on community data? If so, under what conditions?
  Considerations: Derivatives may include modified datasets, statistical models, AI models, or synthetic data generated from the original dataset. Some communities may permit derivatives under conditions like attribution, revenue sharing, or explicit approval.
- **Accountability:** How should violations of agreed terms be addressed? Who should monitor compliance and resolve disputes?
  Examples: Oversight mechanisms can include independent audits, grievance redress mechanisms, participatory review boards, or legal enforcement measures.

4. HOW: Establishing Operational Strategy and Governance

This section defines the **governance mechanisms** that ensure responsible data reuse.

- **Data Access & Sharing:** How will data be accessed or shared?
  Examples: Data can be shared through models including open access, secure repositories, or community-led cooperatives.
- **Privacy & Security:** What privacy-preserving measures should be applied to protect individuals and sensitive information?
  Examples: De-identification, encryption, access controls, and Privacy Enhancing Technologies like differential privacy techniques can mitigate risks. Communities may also require transparency on how privacy is maintained over time.
- **Governance Mechanisms:** What transparency, accountability, and grievance redress measures should exist?
  Considerations: Transparent governance can include regular audits, public reporting, participatory oversight boards, and dispute resolution pathways. Communities may want formal grievance mechanisms for addressing misuse or violations.
- **IP & Ownership:** Should the community retain intellectual property rights over datasets and derivative outputs? How should IP-related claims be managed?
  Considerations: Some communities may wish to maintain full control, while others may allow limited rights to external actors under predefined conditions (e.g., licensing fees, attribution, or restrictions on commercialization).

5. WHEN: Addressing Duration, Retention, and Review

This section ensures the **social license remains dynamic** and adapts to evolving community needs.

- **Duration:** How long should the data be used? Should the community have the power to revoke the license?
  Considerations: Data use may be time-limited or ongoing. Communities may require expiration clauses, renewal mechanisms, or revocation rights.
- **Retention & Deletion:** Under what conditions should data be deleted?
  Considerations: Data retention should balance utility with privacy and security concerns. Some communities may require automatic deletion after a set period or upon project completion, community request, or breach of terms, while others may allow continued use with oversight.
- **Periodic Review:** How frequently should the terms of the social license be reassessed? What should trigger a review?
  Considerations: Reviews may be scheduled (e.g., annually) or event-driven (e.g., technology changes, new risks, project expansion).
- **Measuring Impact:** When and how should the benefits of data reuse be evaluated? Should the agreement include provisions for monitoring progress?
  Considerations: Impact assessments can track community benefits, mitigate harms, and inform adjustments to the license. Metrics may include economic, social, or technological outcomes.

6. WHERE: Considering Contextual and Jurisdictional Factors

This section ensures compliance with **legal, cultural, and geographic considerations**.

- **Data Storage Location:** Where should data be stored, and under whose jurisdiction?
  Considerations: Storage location affects legal oversight, data sovereignty, and accessibility. Some communities may require local storage to maintain control, while others may permit cross-border transfers under specific conditions.
- ***Data Mobility:*** Under what conditions can data or its derivatives be moved to another country or jurisdiction?
  Considerations: Cross-border data transfers may impact jurisdictional oversight, legal compliance, and community control. Licensing terms can restrict movement, require additional approvals, or mandate that data remains under local governance.
- **Legal Compliance:** To what local, national, and international regulations should the project adhere?
  Considerations: Compliance may involve GDPR, data localization laws, indigenous data sovereignty frameworks, or sector-specific regulations. Communities may require additional contractual protections beyond legal mandates.
- **Contextual Sensitivity:** How will the initiative respect cultural norms, and socio-economic realities?
  Considerations: Cultural protocols, language preferences, and governance traditions should be considered.

Appendix B: Sample Clauses for Social License-Compliant Data Sharing and Use Agreements

These sample clauses provide a practical template for embedding social licensing principles into data sharing and use agreements. They are intended for facilitators, legal advisors, and community representatives responsible for formalizing agreements after completing participatory engagement using the **Social Licensing Questionnaire** (Appendix A).

The clauses reflect community-defined preferences, creating enforceable terms that protect community rights and hold data users accountable. They are designed to be flexible, adapted to fit the project's sector, cultural and legal context, and technical environment.

Where data crosses jurisdictions or involves third-party access, additional considerations—such as cross-border enforcement, jurisdictional submission, and technical safeguards—may be required. The clauses include optional language to support these needs, but users should work with local legal counsel to ensure agreements align with applicable national and Indigenous legal frameworks.

Finally, these clauses emphasize a layered approach to enforcement—combining contractual accountability, technical safeguards like audit logs and machine-readable licenses, and participatory oversight through community bodies. Where disputes arise, the clauses recommend structured, staged resolution processes, starting with community-based mechanisms and escalating if necessary to arbitration or judicial proceedings, with parties submitting to a jurisdiction chosen by the community.

These clauses are not prescriptive templates, but rather starting points—practical tools to help operationalize social licensing in ways that reflect each community's values, governance structures, and legal realities.

[Insert Project Title]

Data Sharing and Use Agreement

This Agreement is made and entered into by and between:
Party A: [Insert Organization Name, Address, and Contact Information]
Party B: The [Insert Community Name] Community, represented by [Insert Representative Names/Titles, e.g., elected leaders, cultural advisors, and community-designated representatives].

Effective Date: [Insert Date]
Duration: [Specify term or conditions for termination]

WHY: Defining Purpose, Scope, and Limitations

- **Purpose**
  This agreement establishes conditions for the ethical and responsible use of [specify data type, e.g., public health/environmental/educational] data, as defined through participatory engagement with [Community Name]. Data may only be used for [describe purpose, e.g., monitor environmental health risks] and must not be used for any of the explicitly prohibited purposes listed below.
- **Scope**
  The scope of this agreement applies only to the data described in the [reference annex, e.g., Annex A: Dataset Datasheet]. Any reuse, processing, or disclosure beyond the defined purpose—including for commercial sale, surveillance, AI training, or political profiling—requires explicit approval confirmed through mutually accepted channels, including written, verbal (with recorded documentation), or certified digital communication. Such approval must be documented as an annex to this agreement and becoming part thereof.

WHAT: Determining Appropriate Data Assets

- **Data Categories**
  The covered data includes:
    - [List specific approved datasets, types, and sources].

  The following data types are explicitly excluded from this agreement:

    - [List excluded datasets, e.g., personally identifiable information, sensitive cultural heritage data, or any data that could enable re-identification when combined with other datasets].
- **Granularity**
  All data shared under this Agreement must be provided in anonymized or aggregated formats, consistent with Annex C: Privacy and Security Protocols. If individual-level data is required for research purposes, it must undergo a separate approval process from the [Community Representative/Oversight Body], with enhanced protections applied.

WHO: Identifying Actors, Roles, and Responsibilities

- **Defining the Community**
  "Community" refers to [describe the community, e.g., residents of [Community Name], represented by their elected leaders, cultural advisors, and designated governance body]. The community's representation and authority to enter into this agreement is documented in Annex B: Community Engagement Record.
- **Roles and Responsibilities**
    - **Data Steward (Party A)**: Responsible for secure data management, processing, and ensuring compliance with this Agreement. Party A maintains records of all access, processing, and outputs and provides these to Party B upon request.
    - **Community Representative/Oversight Body (Party B)**: Responsible for oversight, reviewing amendments, resolving disputes, and protecting community interests.
    - **Authorized Users**: Only [list approved individuals/entities, e.g., researchers at accredited institutions, approved public sector agencies, or nonprofit organizations directly engaged in the project] may access the data, and each must sign an acknowledgment of obligations before access.
- **Third-Party Use**
  No third party may access, receive, or use the data unless Party B provides explicit approval (confirmed through mutually accepted channels, as defined above). All third parties must:
    - Submit to the jurisdiction of [Community's chosen legal system or arbitration body] for any disputes related to data misuse, breach, or non-compliance.
    - Sign a binding agreement adopting terms at least as protective as this agreement, including clear indemnification of the community for any misuse, unauthorized disclosure, or breach by the third party.
- **Derivatives**
  For the purposes of this Agreement, "Derivatives" include:
    - Processed, modeled, or aggregated outputs that rely directly on the data covered by this agreement.
    - Any AI models, statistical analyses, visualizations, or synthetic data generated using this data.
    - Any outputs that would not exist but for the inclusion of this data, even if combined with external datasets.

  Derivatives may only be created if all of the following conditions are met:

    - Explicit prior approval has been granted by Party B (confirmed through mutually accepted channels).
    - Clear and visible attribution is given to [Community Name] in all outputs, publications, and products.

- Benefit-sharing provisions, as defined in the Benefit Sharing Mechanism, are applied to any commercial use or downstream monetization.
- Cultural or sacred data is excluded from derivative creation unless explicitly approved by Party B, with conditions for cultural safeguarding documented in Annex C: Privacy and Security Protocols.

- **Audit Rights**
  Party B has the right to conduct periodic audits of data use, derivative creation, and access logs. These audits may be:
  - Routine (announced with reasonable notice) or
  - Triggered (in response to suspected breach, with or without notice).

  Audit processes shall include access to:
  - Machine-readable licenses, technical access logs, and applied metadata controls.
  - Records demonstrating the application of technical safeguards as set out below and in Annex C: Privacy and Security Protocols.

HOW: Establishing Operational Strategy and Governance

- **Data Access**
  Access to the data is limited to the authorized users identified in this agreement. All access must occur through a secure data portal managed by Party A. Each access event must be logged, including the user, date/time, purpose, and data accessed. These logs must be retained for the duration of the agreement and made available to Party B upon request. Specific access conditions—such as permitted purposes, time limits, and any special safeguards—shall be jointly agreed and documented in the [Community Data Governance Record] maintained by both parties. All personnel handling the data must complete [training in cultural sensitivity, data protection] prior to receiving access.
- **Privacy and Security**
  All data must be [anonymized, pseudonymized, or aggregated] before sharing, using the methods set out in Annex C: Privacy and Security Protocols. Data storage, processing, and transmission must comply with [applicable laws, e.g. GDPR, Indigenous Data Governance Protocols, or national data protection laws].

  Party A shall apply all technical safeguards specified in Annex C, including watermarking, machine-readable metadata, encryption, privacy-enhancing technologies, and geofencing where relevant.
- **Transparency**
  Party A will maintain a [Data Use Register], updated at least [quarterly], recording:
  - Who accessed the data, when, and for what purpose.
  - What processing was performed and what derivatives were created (if any).
  - How governance conditions were applied during use (e.g., whether derivative creations received prior approval).

  This register will be accessible to Party B upon request and summarized in a community-facing report shared [annually] at [community meetings or via community-approved communication channels].

- **Participatory Methods**
  This agreement reflects community preferences gathered through structured engagement documented in Annex B: Community Engagement Record. This includes:
  - Dates, locations, and methods of engagement.
  - Names and affiliations of participants.
  - Key concerns raised and decisions reached.
- **Behavioral Use Restrictions**
  The data may not be used for:
  - [Surveillance, discriminatory profiling, or law enforcement targeting].
  - [Targeted advertising without explicit community-approved benefit-sharing].
- **Algorithmic Transparency**
  If the data is used to train or fine-tune AI models, Party A must provide Party B with:

- A plain-language description of the model's purpose and design.
- A summary of the data inputs used in training.
- A list of downstream uses for the model.
- Results of any bias audits conducted during model development.

Failure to provide or update this documentation shall result in immediate suspension of data access until full documentation is provided and reviewed by Party B.

In addition, Party A must ensure that all models trained using this data embed machine-readable metadata that tracks:

- The data's origin.
- The applicable licensing terms (including restrictions and attribution requirements).
- Any benefit-sharing obligations.

Technical non-compliance (e.g., failure to apply machine-readable licenses) shall be treated as a material breach of this agreement.

- **Bias Mitigation**
  Party A must conduct bias audits at agreed stages of model development, using methodologies agreed in advance and documented in Annex C. Results of these audits must be shared with Party B and included in the project's transparency reporting. If Party B identifies that the data has contributed to unacceptable bias, harm, or misrepresentation, Party B may:
  - Suspend data access until corrective actions are taken.
  - Require corrective actions such as retraining models, removing biased outputs, or ceasing certain uses altogether.
- **Data Rights**
  The community retains collective rights over the original data, in accordance with [applicable laws, Indigenous Data Sovereignty frameworks, and cultural governance norms], unless explicitly waived. These rights include:
  - The right to define and enforce conditions for access, sharing, and reuse.
  - The right to require attribution whenever data or derivatives are used, published, or commercialized.
  - The right to initiate audits and demand corrective action if the data is misused or governance conditions are breached.
- **Intellectual Property Rights on Outputs (Derivatives)**
  Any intellectual property created using the original data—including models, synthetic data, publications, software, or commercial products—must:
  - Clearly and visibly acknowledge the data's community origins.
  - Comply with the benefit-sharing terms defined in this agreement.
  - Avoid any implication that Party A (or any third party) holds exclusive rights to outputs derived from data provided under this agreement.

  Where national laws limit the recognition of collective intellectual property, the parties shall:
  - Record the community's contribution and governance conditions in all documentation (including licensing metadata embedded in derivatives).
  - Ensure that community representatives participate in co-authorship, patent filings, or licensing decisions where applicable.
- **Benefit-Sharing Mechanism**
  Any financial benefits, royalties, licensing fees, commercial revenues, or non-financial benefits arising from the use of the data or derivatives must be equitably shared with the community, in accordance with Annex D: Benefit Sharing and Reinvestment Plan. Funds should be reinvested into [community-defined priorities defined during the engagement process such as education, healthcare, infrastructure, cultural preservation, or economic development], with transparent reporting to the community at regular intervals on funds received, how they were allocated, and who made allocation decisions.
- **Dispute Resolution**
  Disputes arising under this agreement shall follow a tiered process. The process applies to all parties, including third parties who receive access to the data under this agreement.

Step 1: Community-Based Mediation
All disputes must first go to community-based mediation, facilitated by the Community Oversight Body (or a mutually agreed neutral facilitator).

- Mediation must begin within 30 days of written notice of a dispute.
- All parties must participate in good faith to seek a mutually agreeable resolution.

Step 2: Independent Arbitration
If mediation fails to resolve the dispute within 30 days, the dispute shall be referred to independent arbitration under the rules of [recognized arbitration body, e.g., ICC, UNCITRAL].

- Arbitration occurs in the language(s) of the community and considers the cultural and legal context documented in Annex B: Community Engagement Record.
- Arbitration decisions are final and binding on all parties, including third parties.

Step 3: Judicial Proceedings (if necessary)
If any party refuses arbitration or if an award requires judicial enforcement, the matter may be brought before the courts of [Community's chosen jurisdiction].

- As a condition of entering this agreement, all parties—including third parties—expressly agree to submit to the jurisdiction of [Community's chosen legal system] and waive any claims of forum non conveniens (inconvenient forum).

Special Provisions for Third Parties
Any third party accessing data under this agreement must:

- Formally agree to this dispute resolution process as a pre-condition to data access.
- Accept the jurisdiction of the community's legal system and the authority of the Community Oversight Body.
- Agree to participate in good faith mediation and binding arbitration before seeking judicial relief.

WHEN: Addressing Duration, Retention, and Review

- **Retention Period**
  Data shared under this agreement shall be retained for no longer than [specify period, e.g., five years] from the Effective Date, unless both parties formally agree to an extension through mutually accepted channels (see definition in Scope section). At the end of this period, all copies of the data—including backups and derivatives—must be either securely deleted or returned to the community, following procedures agreed upon in Annex C: Privacy and Security Protocols. If any portion of the data has been incorporated into a larger dataset, the terms of this agreement shall continue to govern the portions derived from the community's data, even after deletion of the standalone dataset.
- **Periodic Review**
  This agreement—and all ongoing uses of the original data—shall be formally reviewed at least annually to assess:
  - Whether the data's use still aligns with community preferences and stated purposes.
  - Whether new risks, technologies, or regulatory changes require adjustments.
  - Whether benefit-sharing and governance processes remain effective.

  The review process shall include [e.g. consultations with the community, data impact assessments, and publication of a publicly accessible review report].

- **Incident Response Timelines**
  In the event of a data breach, unauthorized use, or other material incident, Party A must notify Party B within [e.g., 72 hours] of discovery. Notification must include a description of the incident, the types and amount of data affected, immediate mitigation measures taken, and the steps planned to prevent recurrence. Party A must also notify any directly affected community members without undue delay and provide updates until the issue is fully resolved. Detailed procedures for incident response are outlined in [Annex E: Data Breach and Incident Response Plan].

- **Termination Clause**
  Either party may terminate this agreement with [insert notice period, e.g., 60 days] advance notice, unless termination is required sooner due to material breach. Upon termination, Party A must:
    - Cease all further use of the data.
    - Return or securely delete all copies and derivatives according to procedures defined in Annex C.
    - Provide Party B with a certification of deletion or return, confirmed through mutually accepted channels.

  Termination does not release either party from ongoing obligations related to confidentiality, attribution, or benefit sharing.

WHERE: Considering Contextual and Jurisdictional Factors

- **Jurisdiction Compliance**
  All data covered by this agreement shall be stored, processed, and governed within [jurisdiction(s) agreed by the parties, e.g., within the territory of the community or a nationally regulated data facility]. Data governance shall comply with [e.g. national data protection laws, Indigenous Data Sovereignty Frameworks, regional agreements such as GDPR]

  If a conflict arises between these laws and this agreement, the parties shall apply the legal framework that provides the strongest protection for community rights, as determined through mutual consultation or, if needed, by referral to a jointly designated legal expert.

- **Data Mobility**
  Data—including derivatives—may not be transferred, stored, or processed outside the agreed jurisdiction(s) unless Party B provides explicit approval confirmed through mutually accepted channels. Any approved cross-border transfer must:
    - Be governed by a binding legal agreement ensuring compliance with this agreement's terms.
    - Apply technical controls (e.g., geofencing, encryption, access controls) to prevent unauthorized access outside approved jurisdictions.
    - Include a disclosure process, ensuring the community understands the destination jurisdiction's data protection laws and any risks posed by weaker protections in the receiving jurisdiction.
    - Allow the community the right to object if protections are deemed insufficient.

Signatures

By signing this agreement, all parties affirm that they have the authority to act on behalf of their organizations and/or communities and agree to comply with its terms.

Party A:
[Name, Title, Date, Signature]

Party B:
[Name, Title, Date, Signature]

Final agreements should be accompanied by the following documents to support transparency, traceability, and compliance:

- A Dataset Datasheet (defining covered data).
- A Community Engagement Record (documenting the process that shaped preferences).
- A Privacy and Security Protocol (outlining technical safeguards).
- Agreed terms for benefit sharing and ongoing oversight.

Depending on the community's capacity and the scale of the project, these supporting documents may be consolidated into a single Community Data Governance Record to simplify administration.

Recommended Accompanying Documents
To ensure transparency, traceability, and clarity, the following annexes are recommended as core documentation:

- **Annex A: Dataset Datasheet** – A clear description of the data covered by the agreement, including source, format, sensitivity, any processing applied before sharing, and applicable anonymization, pseudonymization, or aggregation methods.
- **Annex B: Community Engagement Record** – Full documentation of the engagement process, including dates and locations of meetings; methods used; names and affiliations of participants; summary of key community preferences, concerns, and agreements reached; and the process by which community representatives were identified and selected.
- **Annex C: Privacy and Security Protocols** – Technical and procedural safeguards, including standards, technologies, and requirements for anonymization, aggregation, encryption, access and logging, and data retention and deletion.
- **Annex D: Benefit Sharing and Reinvestment Plan** – Detailed terms for benefit-sharing, including types of benefits covered; allocation percentages or mechanisms; community priorities for reinvestment; reporting obligations; and oversight mechanisms.
- **Annex E (optional): Data Breach and Incident Response Plan** – Pre-agreed steps to follow in the event of a breach, including notification timelines; roles and responsibilities during incident response; documentation and reporting requirements; and post-incident review and corrective action processes.

Note: Depending on the project's complexity, additional annexes may be developed as needed. For smaller-scale projects, or where the community has limited capacity, these annexes can be combined into a single Community Data Governance Record.

Appendix C: Scenarios

This appendix provides illustrative scenarios to demonstrate how organizations might pursue a social license for data reuse—particularly in the context of AI tools and applications. These examples offer practical guidance on how to apply the tools presented in this report, including the Social License Questionnaire (Appendix A) and Sample Clauses (Appendix B), across a range of real-world contexts.

These scenarios are designed to help practitioners understand how social licensing can be implemented across different contexts, from initial engagement to formal agreements. While fictional, they reflect common challenges and are intended as practical guides for applying the framework in real-world settings.

Scenario I: International Organization Engaging in Protection Monitoring

Context

- **SITUATION:** An international development organization, International Organization A, is looking at ways to conduct protection monitoring of internally displaced peoples. It needs information on protection risks, issues, incidents,and trends so that it can design effective programming to address the needs of these peoples (e.g. counseling, medical services, child care). There is also some interest in feeding the data into an organization-wide AI chatbot to help the organization's analysts answer basic questions about the context.

- **DATA:** The information needed for this work will be met through a combination of primary data—interviews with internally displaced peoples and hosts—and secondary data—statistical data supplied by government partners.

- **PURPOSE:** International Organization A has elected to pursue a social license because it believes it will:
    - *Improve Reputation:* Pursuing active engagement with local groups will help the organization's image across similar contexts, encouraging others to work with it.
    - *Adhere to its Principles:* Pursuing community support will allow the organization to live up to its values and social responsibility.

Phase I: Establishing Preferences

Phase I.I. Pre-Engagement Planning

- **DEFINING PURPOSE:** The international organization develops a succinct summary of what it is doing (protection monitoring), why it is doing it (program design and AI support), and the value proposition (improving service delivery for internally displaced people) for the community. This information is written as plainly and directly as possible, in a language that will be understandable for the audience.

- **IDENTIFYING STAKEHOLDERS:** International Organization A assesses that its work affects a specific settlement populated mostly by internally displaced people. The organization sends invitations to participate to:
    - ***Volunteer Representatives:*** The settlement already has several volunteer community leaders who speak on the community's interests and with whom the international organization has a pre-existing relationship.
    - ***Health Providers:*** As the data will be used to support service delivery related to physical and mental health, the international organization also seeks to secure support and insight from select health practitioners who serve the community. These individuals come from local NGOs and nonprofits.

Phase I.II. Community Engagement

- **APPROACH*:*** After receiving indication that these groups are interested in working with the International Organization, it considers the best means of engaging. The data needed is demographic data, has a direct impact on the health and well-being of the community, and carries substantial risks. However, the volunteer community leaders indicate that they do not have the interest or capacity for an extended level of engagement.

  Given these factors, the international organization opts to host a ***co-design workshop***, where community members and health providers can describe their attitudes on data use and impact.

- **TOOLS OF ENGAGEMENT:** The settlement that the International Organization is targeting has limited internet access and high levels of digital literacy are not guaranteed. As such, the international organization hosts conversations in person in the form of a structured conversation.

- **IMPLEMENTATION:** The international organization adapts the ***social license questionnaire*** to fit the context, adding detail to questions that are relevant for the settlement and removing content that is irrelevant. A single facilitator leads the participants in discussion, allowing group members to raise issues they consider important and asking them what their expectations would be to address those issues. A separate note-taker summarizes the discussion, noting that several issues surface repeatedly throughout the discussion.
  - *Purpose*: Given who the data will be collected from, representatives want to ensure that it is used purely for health program design and service delivery by the international organization. Members of the community mention they are particularly concerned about surveillance.
  - *Access Rights*: The community wants any information collected to be used solely by the international organization and not to be shared with anyone else without the community's explicit approval.
  - *Retention and Deletion*: Health providers want the data to only be held as long as it is necessary and deleted as soon as it ceases to be relevant for program design and service delivery to avoid potential misuse.

Phase II: Documenting Preferences

- **DRAFTING*:*** The international organization takes these inputs back and seeks to develop a contract that addresses the concerns outlined by the community representatives and health providers in their workshop. The international organization uses the **Sample Clauses for Social License-Compliant Data Sharing and Use Agreements** to guide its draft agreement. In addition to provisions on *Access Rights* and *Retention and Deletion*, the participants agree to include language on:
  - *Purpose*: The purpose of this agreement is to support the protection monitoring objectives identified in collaboration with the leaders of **Community B**. Specifically, the data will be used to design public health, mental health, and pediatric services and is prohibited from being used for surveillance.

- **VALIDATING*:*** The international organization reconnects with the participants of the prior co-design workshop. They explain the agreement that has been drafted and answer questions about its development. Community representatives and health providers are offered an opportunity to suggest additions and improvements. They suggest that while they agree with the agreement overall, they would like more language limiting the use of their data for any law enforcement or military purposes.

- **REVISION:** The international organization adapts the agreement to conform to the major suggested revision:
  - *Purpose*: The purpose of this agreement is to support the protection monitoring objectives identified in collaboration with the leaders of Community B. Specifically, the data will be used to design public health, mental health, and pediatric services and is prohibited from being used for surveillance *or by any law enforcement bodies.*

- **ENACTMENT:** The agreement is then circulated to relevant stakeholders to be signed by relevant representatives within the community who can monitor the data use and ensure that all parties abide by their contractual obligations.

Phase III: Enforcing Preferences

- **ENFORCEMENT MECHANISM:** The international organization examines the context to assess whether there are any accountability or enforcement mechanisms, beyond the data sharing agreement, that might be useful to ensure enforcement of the social license. It reaches several conclusions:
    - *National Laws and Policies:* While the international organization identifies that it may be useful to have several standards embedded in national law, it recognizes that it cannot influence the domestic political context.
    - *Local Governance Structures:* Instead of seeking to alter the legal context, the international organization can use its contacts to maintain contact with various community leaders. It forms a consultative group with internally displaced peoples and health providers to regularly consult them on their attitudes toward data collection processes.
- **DISPUTE RESOLUTION:** As the agreement goes into effect, the international organization continues communicating with community members to inform them of how data is being used to their benefit. If disputes or disagreements emerge about this data use or the degree to which stakeholders are conforming to expectations. If there are any disputes, they are adjudicated by the processes designated

Scenario II: Private Company Seeking Indigenous Language Data in Africa

Context

- **SITUATION:** A technology start-up, Company A, is developing AI-powered translation tools and chatbots. To improve its language models, it seeks to acquire text and oral data in an Indigenous language spoken in Community X, a rural region of an African country. This language is underrepresented in global datasets, but crucial for expanding the company's product to regional markets and improving inclusive service delivery. Much of the data exists in non-digital forms—within community libraries, oral traditions, schools, and cultural archives.
- **DATA:**
    - Oral recordings of storytelling, ceremonies, and cultural histories.
    - Written texts from community archives, schools, and libraries.
    - Linguistic data contributed by local researchers and cultural experts.
- **PURPOSE:** Company A pursues a social license to:
    - *Build Trust*: Given the history of extractive data practices, a social license demonstrates the company's commitment to ethical data governance.
    - *Respect Cultural Sensitivity*: The language holds deep cultural significance, requiring careful governance and safeguards.
    - *Prevent Reputational Risk*: Proactive engagement mitigates potential backlash and helps foster long-term partnerships with Indigenous communities.

Phase I: Establishing Preferences

Phase I.I. Pre-Engagement Planning

- **DEFINING PURPOSE:** Company A develops a clear explanation of:
    - What data it seeks and why (to improve Indigenous language support in AI translation tools).
    - How the data will be used (language model training and chatbot development).
    - Expected community benefits (better language representation in digital tools and support for language preservation).
- **IDENTIFYING STAKEHOLDERS:** Company A identifies and invites participation from:
    - Indigenous Elders and Knowledge Holders: Custodians of oral histories and cultural narratives.
    - Schools and Libraries: Institutions that safeguard written texts and historical documents.
    - Linguistic Researchers and Archivists: Experts involved in language preservation efforts.
    - Community-Based Organizations: Groups advocating for Indigenous rights and cultural preservation.

Phase I.II: Community Engagement

- **APPROACH:** Company A convenes a Community Data Committee made up of representatives from the identified stakeholder groups. This committee serves as the primary interface for the engagement process.
- **TOOLS OF ENGAGEMENT:**
    - Co-design workshops and focus groups, facilitated by a local expert in participatory data governance.
    - Use of the Social Licensing Questionnaire (Appendix A) to structure discussions and capture preferences.
- **IMPLEMENTATION:** Facilitators lead workshops and focus groups, inviting participants to define preferences across key areas, such as:
    - *Purpose*: The community emphasizes that data should be used primarily for language preservation and education, not solely for commercial gain. Any commercial use (e.g., paid translation services) should trigger benefit-sharing.
    - *Access and Custodianship*: Data should be stored locally in a community-controlled repository, with Company A only accessing data under agreed conditions.
    - *Derivative Works*: Community members express strong concerns about AI models generating synthetic Indigenous content (e.g., auto-generated text). These require prior review and explicit approval.
    - *Cultural Sensitivity*: Certain ceremonial songs, prayers, and stories are deemed off-limits for any data use, including AI training.
    - *Benefit Sharing*: Any profits linked to the data should directly fund community-led language preservation programs, including training youth in documentation and translation.

Phase II: Documenting Preferences

- **DRAFTING**: Company A, working with legal counsel and the Community Data Committee, drafts a Social License-Compliant Data Sharing and Use Agreement using the **Sample Clauses** as a guide. Key provisions include:
    - Explicit purpose limitations: Language preservation and education only, with restrictions on commercial use without benefit-sharing.
    - Data to be stored locally in the community's repository, with strict access logs.
    - Requirement for community oversight before any derivative works (like synthetic content) are created.
    - Establishment of a Benefit-Sharing Fund, with community oversight on its management.
    - Disputes to be resolved through community mediation first, with escalation to local arbitration if needed.
- **VALIDATING**: The draft agreement is presented to the community in a validation session, where community members review, discuss, and suggest changes.
- **REVISION:** Company A revises the agreement to incorporate these suggestions and formally signs it alongside authorized community representatives.

Phase III: Enforcing Preferences

- **ENFORCEMENT MECHANISM:** The Community Data Committee maintains ongoing oversight, with authority to:
    - Conduct annual audits of how data is used.
    - Require regular transparency reports from Company A, detailing how the data supports language tools and what benefits have been delivered to the community.
    - Approve or veto proposed derivative works before they are created or commercialized.
- **DISPUTE RESOLUTION:** Disputes are handled through a tiered process:
    - Community mediation facilitated by cultural elders.
    - If unresolved, escalation to national arbitration under a recognized arbitration body.
    - Only if these processes fail, parties may pursue formal legal action.

Scenario III: Public Health Data Sharing Between Government and Research Institutions

Context

- **SITUATION:** A national health ministry, National Health Ministry A, plans to collaborate with international research institutions to study environmental health risks—such as air and water pollution—affecting rural and Indigenous communities. This research will be used to fine-tune an existing AI model meant to support state administrative agents in answering questions about environmental policy. Officials intend the AI system, and the data fueling it, to inform public health programs, environmental policy, and climate adaptation strategies.

- **DATA:** The study will combine:
    - ***Health records*** from local clinics, documenting respiratory diseases, waterborne illnesses, and chronic conditions linked to environmental factors.
    - ***Environmental sensor data*** collected from air and water monitoring stations in the affected communities.
    - ***Community-reported health concerns*** collected through participatory surveys and focus groups.

- **PURPOSE:** The health ministry and its research partners elect to pursue a social license because they believe it will:
    - ***Improve Research Relevance***: Engaging communities will ensure research questions and data interpretations reflect lived experiences and local knowledge.
    - ***Build Trust for Data Access:*** Past extractive data collection projects have damaged trust, so transparent governance and community participation are needed to restore confidence.
    - ***Regulatory Compliance:*** Recent legislation passed by the national parliament requires national ministries to consult with communities prior to data collection.

Phase I: Establishing Preferences

Phase I.I. Pre-Engagement Planning

- **DEFINING PURPOSE:** The ministry and its research institution partner informs the regional government of its intention to collect data for public health, environmental policy, and climate adaptation policymaking. It further notes its intention to use AI tools to support this work. The ministry and regional partners announce the effort through their public communication channels (e.g. social media, townhall meetings, newsletters, posters and placards).

- **IDENTIFYING STAKEHOLDERS:** The ministry and the research institution assess whose expertise might be valuable and whose approval is necessary for the success of the effort. Going through their networks, they settle on several kinds of stakeholders, including:
    - ***Health System Patients and Providers:*** Individuals whose health records will be used to inform the data system as well as those organizations providing that care;
    - ***University Researchers:*** Academics who manage environmental sensors and produce analysis based on the data they produce;
    - ***Local Government Agencies:*** Community-level government institutions responsible for garbage collection and addressing pollution.

Phase I.II. Community Engagement

- **APPROACH:** As this work will inform public services, the ministry decides to two simultaneous participatory engagement approaches. First, it will host a general ***public meeting*** where any member of the public can express their concerns in a public setting. Second, it will form a series of ***focus groups*** of the previously identified stakeholders to highlight important issues that might be relevant for them to consider.

- **TOOLS OF ENGAGEMENT:** The ministry knows that many people in the community do not pay attention closely to public meetings so it seeks to promote its engagement with the public through social media and signage posted at bus stops, community centers, and other public settings. It further relies on its connections to regional and local government to support this effort of raising awareness.

  Meanwhile, the international research institution emails university researchers and health providers that it is already connected to. It seeks their help to identify health professionals, patients, and researchers who might be useful to serve in a focus group.

- **IMPLEMENTATION:** In public meetings, a member of the ministry presents on the proposed work to community members. Using the **Social Licensing Questionnaire** as guidance, they highlight what data will be used and how it will be used as well as when and where. Community members are then invited, in a Question and Answer format, to express their concerns.

  In the focus groups, the international research institution takes the lead in facilitation. Providing a series of scenarios to participants (each highlighting different potential situations in which the ministry might use the data), the facilitator seeks to understand what kinds of data usage the stakeholders are comfortable with and what they are not. The facilitator encourages the participants to explain *why* they feel a certain way and what factors may exacerbate or mitigate these feelings. Are there any guardrails they would like to see? Who do they not want involved?

  The combination of these engagement mechanisms surfaces several prominent concerns:
  - *Data Types:* The focus groups indicated that while they are supportive of most patient data informing these systems, they do not think any data about child patients should be used as they have few options if it is misused.
  - *Governance Mechanism:* The focus group further noted that it would like for researchers and patients to be regularly briefed on the applications of their data. They believe that a standing committee of both groups will allow for accountability.
  - *Data Storage:* Noting the involvement of the international research institution, the public meeting concludes with many participants expressing their desire for all environmental and health data to be stored locally and for the ministry (or another locally accountable body) to be charged with managing it.

Phase II: Documenting Preferences

- **DRAFTING:** Based on the preferences captured, the ministry outlines a framework for how it will interact with the community and the research institution using the **Sample Clauses**. Key provisions include:
  - Clear limits on the kinds of patients whose data will be integrated into the system.
  - A requirement that two bodies, one representing patient interests and the other representing researcher interests, be formed and meet regularly to discuss the data use and guide ministry policy.
  - Requirement that any and all data collected remain within the systems managed by the ministry and that the data only be accessible through portals that it manages.

- **VALIDATING:** The research institution and ministry share their work with the public and the focus group for review. The ministry receives additional comments during its public meeting. It concludes that there are no further revisions that it needs to make to the documents. It codifies them.

Phase III: Enforcing Preferences

- **ENFORCEMENT MECHANISM:** The committees of patients and researchers serves as a check on the ministry and its partner. Each month, the ministry must present on what it has done with the data and share a summary of the insights. If the ministry engages in any conduct that seems unethical or in violation of local expectations, the committee informs the ministry. It may also communicate with the public on current activities.

References

Adams R., Adeoye O., Osiruemu E., Owoyemi A., Nyakarundi A., Gitau S., Omulo G., Oduor M., Mwangi M., Karuri-Sebina G., Oduor C., Wafula M., and **Kaye S.** (**2023**), *A New Research Agenda for African Generative AI*, *Nature Human Behaviour*, vol. 7, p. 956–969, https://doi.org/10.1038/s41562-023-01735-1.

Aitken M., Toreini E., Carmichael P., Coopamootoo K., Elliott K., and **van Moorsel A.** (**2020**), *Establishing a Social Licence for Financial Technology: Reflections on the Role of the Private Sector in Pursuing Ethical Data Practices*, *Big Data & Society*, vol. 7, 4 March 2020, https://doi.org/10.1177/2053951720908892.

Ali S. H. (**2015**), *The Social License to Operate from Indigenous Communities*, presentation at the Extractive Industry Symposium, 2015, https://maramatanga.ac.nz/node/1827.

Alonso C., Kothari S., and **Rehman S.** (**2020**), *How Artificial Intelligence Could Widen the Gap Between Rich and Poor Nations*, *IMF Blog*, 2 December 2020, https://www.imf.org/en/Blogs/Articles/2020/12/02/blog-how-artificial-intelligence-could-widen-the-gap-between-rich-and-poor-nations.

Australian Research Data Commons (**2024**), *Language Data Commons of Australia Co-Design Workshop Report for the HASS and Indigenous Research Data Commons*, *Zenodo*, 12 March 2024, https://doi.org/10.5281/zenodo.10807355.

Benhamou Y. and **Dulong de Rosnay M.** (**2023**), *Open Data Commons Licenses (ODCL): Licensing Personal and Non-Personal Data Supporting the Commons and Privacy*, *SSRN*, 12 December 2023, https://doi.org/10.2139/ssrn.4662511.

Benjamin M., Hardt D., Hurwitz J., Perel M., Rodwin D., Smith G., and **Tannen J.** (**2019**), *Towards Standardization of Data Licenses: The Montreal Data License*, *arXiv*, 21 March 2019, https://arxiv.org/abs/1903.12262.

Benjamin R. (**2019**), *Race After Technology: Abolitionist Tools for the New Jim Code*, Polity.

Birhane A., Prabhu V., Katell M., Zanfir-Fortuna G., Marda V., Sambasivan N., and **Mohamed S.** (**2022**), *Power to the People? Opportunities and Challenges for Participatory AI*, *Proceedings of the 2022 ACM Conference on Equity and Access in Algorithms, Mechanisms, and Optimization (EAAMO '22)*, 6–9 October 2022, Arlington, VA, USA, https://doi.org/10.48550/arXiv.2209.07572.

Boutilier R. G. (**2014**), *Frequently Asked Questions About the Social Licence to Operate*, *Impact Assessment and Project Appraisal*, vol. 32, no. 4, p. 263–272, https://doi.org/10.1080/14615517.2014.941141.

Breakey H. (**2023**), *The Social Licence to Operate: Activist Weapon, Industry Shield, Empty Buzzword, or Vital Ethical Tool?*, in *Social Licence and Ethical Practice*, ed. Hugh Breakey, Emerald Publishing Limited, Leeds, p. 1–10, https://doi.org/10.1108/S1529-209620230000027001.

Bremmer I. and **Suleyman M.** (**2023**), *The AI Power Paradox: Can States Learn to Govern Artificial Intelligence—Before It's Too Late?*, *Foreign Affairs*, September/October 2023, https://www.foreignaffairs.com/world/artificial-intelligence-power-paradox.

Bühler M. M., Calzada I., Cane I., Jelinek T., Kapoor A., Mannan M., Mehta S., Micheli M., Mookerjee V., Nübel K., Pentland A., Scholz T. R., Siddarth D., Tait J., Vaitla B., and **Zhu J.** (**2023**), *Data Cooperatives as Catalysts for Collaboration, Data Sharing, and the (Trans)Formation of the Digital Commons*, 8 April 2023, https://ssrn.com/abstract=4413371.

Cardona-Rivera R. E., Alladin J. K., Litts B. K., and **Tehee M.** (**n.d.**), *Indigenous Futures in Generative Artificial Intelligence: The Paradox of Participation*, in *Teaching and Generative AI*, https://uen.pressbooks.pub/teachingandgenerativeai/chapter/indigenous-futures-in-generative-artificial-intelligence-the-paradox-of-participation/.

Carroll M. W. (**2015**), *Sharing Research Data and Intellectual Property Law: A Primer*, *PLOS Biology*, vol. 13, no. 8, e1002235, 27 August 2015, https://doi.org/10.1371/journal.pbio.1002235.

Casal–Ribeiro M., Peixoto U. I., Medeiros–Leal W., and **Santos R.** (**2024**), *Research Fatigue's Impact*

on Small-Scale Fishers' Engagement: A Case-Study from Azores Fisheries, *Marine Policy*, vol. 170, December 2024, article 106404, https://doi.org/10.1016/j.marpol.2024.106404.

Centre for Information Policy Leadership and **BKL** (**2024**), *The Limitations of Consent as a Legal Basis for Data Processing in the Digital Society*, 6 December 2024, https://www.informationpolicycentre.com/uploads/5/7/1/0/57104281/cipl_bkl_limitations_of_consent_legal_basis_data_processing_dec24.pdf.

Cheesman M. (**2024**), *Conjuring a Blockchain Pilot: Ignorance and Innovation in Humanitarian Aid*, *Geopolitics*, 14 August 2024, https://doi.org/10.1080/14650045.2024.2389284.

Contractor D. and **Muñoz Ferrandis C.** (**2022**), *The BigScience RAIL License*, *Hugging Face*, 19 May 2022, https://bigscience.huggingface.co/blog/the-bigscience-rail-license.

Corbett E., Denton E., and **Erete S.** (**2023**), *Power and Public Participation in AI*, *Proceedings of the 2023 ACM Conference on Equity and Access in Algorithms, Mechanisms, and Optimization (EAAMO '23)*, October 2023, https://doi.org/10.1145/3617694.3623228.

De La Cruz G. (**2017**), *Look Back: The 1996 Marcopper Mining Disaster*, *Rappler*, 24 March 2017, https://www.rappler.com/environment/disasters/165051-look-back-1996-marcopper-mining-disaster/.

Demajorovic J., Lopes J. C., and **Santiago A. L. F.** (**2019**), *The Samarco Dam Disaster: A Grave Challenge to Social License to Operate Discourse*, *Resources Policy*, vol. 61, June 2019, p. 273–282, https://doi.org/10.1016/j.resourpol.2019.01.017.

Ding J., Jäger E., Ivanova V., and **Bunz M.** (**2024**), *My Voice, Your Voice, Our Voice: Attitudes Towards Collective Governance of a Choral AI Dataset*, *arXiv preprint* arXiv:2412.01433, https://arxiv.org/abs/2412.01433.

Douglas A. (**2022**), *The Future is Data (Cooperatives)*, *Karya*, 22 November 2022, https://karya.in/resources/blog/the-future-is-data/.

Dumas J., Nam C., Smith W., Atkinson D., and **DeCario N.** (**2023**), *The AI2 ImpACT License: A New Way to Think About AI Licensing*, *AI2 Blog*, 27 July 2023, https://medium.com/ai2-blog/the-ai2-impact-license-a-new-way-to-think-about-ai-licensing-bc90ff26a9ee.

European Commission (**n.d.**), *EU Register of Data Intermediation Services*, *Shaping Europe's Digital Future*, https://digital-strategy.ec.europa.eu/en/policies/data-intermediary-services.

European Union (**2023**), *Regulation (EU) 2023/2854 of the European Parliament and of the Council of 13 December 2023 on Harmonised Rules on Fair Access to and Use of Data and Amending Regulation (EU) 2017/2394 and Directive (EU) 2020/1828 (Data Act)*, *Official Journal of the European Union*, 22 December 2023, https://eur-lex.europa.eu/eli/reg/2023/2854/oj/eng.

Fairly Trained (n.d.), *About*, https://www.fairlytrained.org/about.

Fan Q. and **Qiang C. Z.** (**2024**), *Tipping the Scales: AI's Dual Impact on Developing Nations*, *World Bank Blogs*, 3 June 2024, https://blogs.worldbank.org/en/digital-development/tipping-the-scales--ai-s-dual-impact-on-developing-nations.

Frechtling D. (**2023**), *The Mismanagement of User Consent Data and Its Consequences*, *IAPP News*, 9 March 2023, https://iapp.org/news/a/the-mismanagement-of-user-consent-data-and-its-consequences/.

Fruchterman J., McKinney-Bock K., and **Francis S.** (**2024**), *A Better Deal for Data: The Case for Model Data Commitments and Agreements*, *Tech Matters*, 2 April 2024, https://bd4d.org/resources/24_04-a_better_deal_for_data.pdf.

Fung A. (**2006**), *Varieties of Participation in Complex Governance*, *Public Administration Review*, vol. 66, no. s1, p. 66–75, https://doi.org/10.1111/j.1540-6210.2006.00667.x.

Gehman J., Lefsrud L. M., and **Fast S.** (**2017**), *Social License to Operate: Legitimacy by Another Name?*, *Canadian Public Administration*, vol. 60, no. 2, p. 293–317, https://doi.org/10.1111/capa.12218.

Ghosh S., Venkit P. N., Gautam S., Wilson S., and **Caliskan A.** (**2024**), *Do Generative AI Models Output Harm while Representing Non-Western Cultures: Evidence from A Community-Centered Approach*, *arXiv*, 20 July 2024, https://arxiv.org/abs/2407.14779.

Global Partnership on AI (**2022**), *Intellectual Property Expert*

Preliminary Report on Data and AI Model Licensing, November 2022, https://gpai.ai/projects/innovation-and-commercialization/intellectual-property-expert-preliminary-report-on-data-and-AI-model-licensing.pdf.

Hamadeh N., Van Rompaey C., and **Metreau E.** (**2023**), World Bank Group Country Classifications by Income Level for FY24 (July 1, 2023–June 30, 2024), *World Bank Blogs*, 30 June 2023, https://blogs.worldbank.org/en/opendata/new-world-bank-group-country-classifications-income-level-fy24.

IEEE Public Safety (**2024**), Predictive Analytics in Disaster Prevention: Machine Learning Models for Early Warning Systems, https://publicsafety.ieee.org/topics/predictive-analytics-in-disaster-prevention-machine-learning-models-for-early-warning-systems.

Ihlen Ø. and **Raknes K.** (**2020**), *Appeals to 'the Public Interest': How Public Relations and Lobbying Create a Social License to Operate*, *Public Relations Review*, vol. 46, no. 5, December 2020, article 101976, https://doi.org/10.1016/j.pubrev.2020.101976.

Information Commissioner's Office (**n.d.**), *Why Is Consent Important?*, https://ico.org.uk/for-organisations/uk-gdpr-guidance-and-resources/lawful-basis/consent/why-is-consent-important/.

International Network on Digital Self-Determination (**2023–2025**), *From Principle to Practice*, https://cms.thegovlab.com/assets/7b226f48-74a1-4a08-aecf-bbe64652f298/.

Kenton W. (**2024**), Social License to Operate (SLO): Definition and Standards, *Investopedia*, 31 August 2024, https://www.investopedia.com/terms/s/social-license-slo.asp.

Kim D. (**2020**), '*AI-Generated Inventions': Time to Get the Record Straight?*, *GRUR International*, vol. 69, no. 5, p. 443–456, https://academic.oup.com/grurint/article/69/5/443/5854752.

Licensing African Datasets (**n.d.**), *Licensing African Datasets*, https://licensingafricandatasets.com/.

Longpre S., Mahari R., Obeng-Marnu N., Brannon W., South T., Kabbara J., and **Pentland S. (2024)**, Data Authenticity, Consent, and Provenance for AI Are All Broken: What Will It Take to Fix Them?, An MIT Exploration of Generative AI, 27 March 2024, https://mit-genai.pubpub.org/pub/uk7op8zs/release/2.

Mann M., Mitchell P., Foth M., and **Anastasiu I.** (**2020**), *#BlockSidewalk to Barcelona: Technological Sovereignty and the Social License to Operate Smart Cities*, *Journal of the Association for Information Science and Technology*, vol. 71, no. 9, September 2020, p. 1103–1115, https://doi.org/10.1002/asi.24387.

Meesters M., Wostyn P., van Leeuwen J., Behagel J. H., and **Turnhout E.** (**2021**), *The Social Licence to Operate and the Legitimacy of Resource Extraction*, *Current Opinion in Environmental Sustainability*, vol. 49, April 2021, p. 7–11, https://doi.org/10.1016/j.cosust.2020.11.002.

Mercer-Mapstone L., Rifkin W., Moffat K., and **Louis W.** (**2017**), *Conceptualising the Role of Dialogue in Social Licence to Operate*, *Resources Policy*, vol. 54, December 2017, p. 137–146, https://doi.org/10.1016/j.resourpol.2017.08.001.

Moffat K., Lacey J., Zhang A., and **Leipold S.** (**2016**), *The Social Licence to Operate: A Critical Review*, *Forestry: An International Journal of Forest Research*, vol. 89, no. 5, p. 477–488, https://doi.org/10.1093/forestry/cpv044.

Mucci T. (**2024**), What is Data Sharing?, *IBM*, 30 August 2024, https://www.ibm.com/think/topics/data-sharing.

Mulligan C. E. A. and **Godsiff P.** (**2023**), *Datalism and Data Monopolies in the Era of A.I.: A Research Agenda*, *arXiv*, 16 July 2023, https://arxiv.org/abs/2307.08049.

National Network of Libraries of Medicine (**n.d.**), *Data Provenance*, https://www.nnlm.gov/guides/data-glossary/data-provenance.

Nayebare M., Mushema A., Nsabimana E., Namugerwa R., Nabaggala M., Kyomuhangi T., Ninsiima J., and **Aduwo J.** (**2023**), *Interim Report for Ubuntu-AI: A Bottom-up Approach to More Democratic and Equitable Training and Outcomes for Machine Learning*, *Democratic Inputs for AI*, OpenAI Foundation, San Francisco, 29 September 2023, https://par.nsf.gov/servlets/purl/10543852.

Organisation for Economic Co-operation and Development (OECD) (**2022**), *OECD Guidelines for Citizen Participation Processes*, OECD, Paris, September 2022, https://doi.org/10.1787/63b34541-en.

Osseiran N., Asher-Schapiro A., and **Farouk M.** (**2023**), *In Middle East, Poor Miss Out as 'Faulty' Algorithms Target Aid*, *Context*, 4 October 2023, https://www.context.news/ai/in-middle-east-poor-miss-out-as-faulty-algorithms-target-aid.

Owen J. R. and **Kemp D.** (**2012**), *Social Licence and Mining: A Critical Perspective*, *Resources Policy*, vol. 37, no. 3, p. 346–357, https://doi.org/10.1016/j.resourpol.2012.06.016.

Partnership on AI (**2021**), *Responsible Sourcing of Data Enrichment Services*, 17 June 2021, https://partnershiponai.org/wp-content/uploads/2021/08/PAI-Responsible-Sourcing-of-Data-Enrichment-Services.pdf.

Pushkarna M. and **Zaldivar A.** (**2022**), *The Data Cards Playbook: A Toolkit for Transparency in Dataset Documentation*, *Google Research Blog*, 17 November 2022, https://research.google/blog/the-data-cards-playbook-a-toolkit-for-transparency-in-dataset-documentation

Revel M. and **Penigaud T.** (**2025**), *AI-Facilitated Collective Judgments*, *arXiv*, 6 March 2025, https://doi.org/10.48550/arXiv.2503.05830.

Shakir U. (**2024**), *Uber for AI Labeling*, *The Verge*, 27 November 2024, https://www.theverge.com/2024/11/27/24307360/uber-scaled-solutions-ai-labeling-workforce.

Sharma A., Ajadi S., and **Beavor A.** (**2020**), *Artificial Intelligence and Start-Ups in Low- and Middle-Income Countries: Progress, Promises and Perils*, *GSMA*, 2020, https://www.gsma.com/solutions-and-impact/connectivity-for-good/mobile-for-development/wp-content/uploads/2020/10/Artificial-Intelligence-and-Start-Ups-in-Low-and-Middle-Income-Countries-Progress-Promises-Perils-Final.pdf.

Solove D. J. (**2024**), *Murky Consent: An Approach to the Fictions of Consent in Privacy Law*, *Boston University Law Review*, vol. 104, p. 593–639, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4333743.

Stephens M., Aziz M. H., and **Lopez M.** (**2025**), *Bridging the AI Divide*, *Davos 2025 | Policy Brief*, The Digital Economist, https://www.thedigitaleconomist.com/_files/ugd/92dfa2_7aafc865e9ac4dbd958e0239599deb8f.pdf.

Stihler C. (**2023**), *Exploring Preference Signals for AI Training*, *Creative Commons*, 31 August 2023, https://creativecommons.org/2023/08/31/exploring-preference-signals-for-ai-training/

The Alan Turing Institute (**n.d.**), Algorithmic Allocation of Public Resources, https://www.turing.ac.uk/research/research-projects/algorithmic-allocation-public-resources

The GovLab (**n.d.**), Contracts for Data Collaboration, https://contractsfordatacollaboration.org/

Thomson I. (**1999**), *A Social License to Operate: Essential for Success in International Exploration*, keynote address at the Cordilleran Roundup, Vancouver, January 1999, https://www.socialicense.com/publications/ROUNDUP1999%20SLO.pdf.

Verhulst S. (**2025**), *Data Stewardship Decoded: Mapping Its Diverse Manifestations and Emerging Relevance at a Time of AI*, 4 February 2025, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=5124555.

Verhulst S. and **Young A.** (**2022**), *Identifying and Addressing Data Asymmetries So as to Enable (Better) Science*, *Frontiers*, 18 July 2022, https://www.frontiersin.org/journals/big-data/articles/10.3389/fdata.2022.888384/full.

Verhulst S. G. (**2023**), Operationalizing Digital Self-Determination, *Data & Policy*, vol. 5, e14, https://doi.org/10.1017/dap.2023.11.

Verhulst S. G. and **Saxena S.** (**2020**), *The Need for New Methods to Establish the Social License for Data Reuse*, *Data & Policy Blog, Medium*, 31 August 2020, https://medium.com/data-policy/the-need-for-new-methods-to-establish-the-social-license-for-data-reuse-e7c67bdc4aff.

Verhulst S. G. and **Schüür F.** (**2023**), *Interwoven Realms: Data Governance as the Bedrock for AI Governance*, *Data & Policy Blog, Medium*, 20 November 2023, https://medium.com/data-policy/interwoven-realms-data-governance-as-the-bedrock-for-ai-governance-ffd56a6a4543.

Verhulst S. G., Chafetz H., and **Zahuranec A. J.** (**2024a**), *Data Commons: Under Threat by or the Solution for a Generative AI Era? Rethinking Data Access and Re-Use*, *SSRN*, 21 May 2024, https://doi.org/10.2139/ssrn.4836354.

Verhulst S. G., Sandor L., Mejia Pardo N., Murray E., and **Addo P.** (**2024b**), *Responsible Data Re-use in Developing Countries: Social Licence through Public Engagement*, *Technical Reports No. 76*, Agence Française de Développement, March 2024, https://www.afd.fr/en/ressources/responsible-data-re-use-developing-countries-social-licence-through-public-engagement.

Verhulst S. G., Sandor L., and **Stamm J.** (**2023**), *The Urgent Need to Reimagine Data Consent*, *Stanford Social Innovation Review*, 26 July 2023, https://ssir.org/articles/entry/the_urgent_need_to_reimagine_data_consent.

Wagner I. (**2022**), *Privacy Policies Across the Ages: Content and Readability of Privacy Policies 1996–2021*, *arXiv*, 21 January 2022, https://arxiv.org/abs/2201.08739.

Waithira N., Mukaka M., Kestelyn E., Chotthanawathit K., Nguyen Thi Phuong D., Nguyen Thanh H., Osterrieder A., Lang T., and **Cheah P. Y.** (**2024**), *Data Sharing and Reuse in Clinical Research: Are We There Yet? A Cross-Sectional Study on Progress, Challenges and Opportunities in LMICs*, *PLOS Global Public Health*, vol. 4, no. 11, e0003392, https://doi.org/10.1371/journal.pgph.0003392.

World Economic Forum (**2022**), *Advancing Digital Agency: The Power of Data Intermediaries*, February 2022, https://www3.weforum.org/docs/WEF_Advancing_towards_Digital_Agency_2022.pdf.

Young A., Verhulst S. G., Safonova N., and **Zahuranec A. J.** (**2020**), *The Data Assembly: Responsible Data Re-Use Framework*, *The Data Assembly*, November 2020, https://thedataassembly.org/files/nyc-data-assembly-report.pdf.

Agence française
de développement
5, rue Roland Barthes
75012 Paris I France
www.afd.fr

Éditions Agence française de développement (AFD) publishes analysis and research on sustainable development issues.

Conducted with numerous partners in the Global North and South, these publications contribute to a better understanding of the challenges faced by our planet and to the implementation of concerted actions within the framework of the Sustainable Development Goals.

With a catalogue of more than 1,000 titles and an average of 80 new publications published every year, Éditions Agence française de développement promotes the dissemination of knowledge and expertise, both in AFD's own publications and through key partnerships.

Discover all our publications in open access at editions.afd.fr.

Towards a world in common.

Publishing Director Rémy Rioux
Editor-in-Chief Thomas Melonio
Graphic creation MeMo, Juliegilles, D. Cazeils
Design and production AFD

Legal deposit 2nd quarter 2025
ISSN 2492-2838
Printed by the AFD reprographics department

To see other publications
in the **Technical Reports** series, visit:
https://www.afd.fr/en/collection/technical-reports