Responsible Data Re-use in Developing Countries: Social Licence through Public Engagement

March 2024
No. 76

Technical Reports

Authors Stefaan G. VERHULST (The Data Tank)
Laura SANDOR (The Data Tank)
Natalia MEJIA PARDO (The Data Tank)
Elena MURRAY (The Data Tank)
Peter ADDO (Agence française de développement)

The Data Tank
AFD éditions
AGENCE FRANÇAISE DE DÉVELOPPEMENT

Agence française de développement

Rapports techniques

Les nombreux rapports, études de faisabilités, analyses de cas et enquêtes de terrain produits par l'AFD contiennent des informations très utiles, en particulier pour les praticiens du développement. L'objectif de cette série est de partager des informations techniques, géographiques et sectorielles sur une dimension du développement et d'en faire un retour d'expérience.

Les opinions exprimées dans ce papier sont celles de son (ses) auteur(s) et ne reflètent pas nécessairement celles de l'AFD. Ce document est publié sous l'entière responsabilité de son (ses) auteur(s) ou des institutions partenaires.

AFD Technical reports

The various reports produced by AFD (feasibility, case studies and field surveys) contain very useful informations, especially for development practitioners. This series aims to provide technical, geographic and sectoral informations on development issues and to share experiences.

The opinions expressed in this paper are those of the author(s) and do not necessarily reflect the position of AFD. It is therefore published under the sole responsibility of its author(s) or its partner institutions.

Responsible Data Re-use in Developing Countries: Social Licence through Public Engagement

Authors

Stefaan G. VERHULST
The Data Tank

Laura SANDOR
The Data Tank

Natalia MEJIA PARDO
The Data Tank

Elena MURRAY
The Data Tank

Peter ADDO
Agence française
de développement

Abstract
The datafication era has transformed the technological landscape, digitizing multiple areas of human life and offering opportunities for societal progress through the re-use of digital data. Developing countries stand to benefit from datafication but are faced with challenges like insufficient data quality and limited infrastructure. One of the primary obstacles to unlocking data re-use lies in agency asymmetries—disparities in decision-making authority among stakeholders—which fuel public distrust. Existing consent frameworks amplify the challenge as they are individual-focused, lack information, and fail to address the nuances of data re-use.
To address these limitations, a Social License for re-use becomes imperative—a community-focused approach that fosters responsible data practices and benefits all stakeholders. This shift is crucial for establishing trust and collaboration, and bridging the gap between institutions, governments, and citizens.

This paper seeks to understand the different components when considering establishing a Social License framework in developing countries. These components include determining which and when actors should be engaged, approaches for public engagement, and various influences of contextual factors on engagement. The paper considers possible outcomes.

The paper examines three developing countries' use cases in Colombia, Senegal, and Nepal, showcasing how components influence the establishment of a data re-use project. The paper analyzes interviews conducted with use case stakeholders and domain experts to further establish key components of implementing a Social License. Success factors include engagement, understanding local contexts, and stakeholder collaboration. Overcoming obstacles requires innovative strategies, including leveraging local networks and developing user-friendly interfaces.
The projects gained trust through clear communication, the involvement of local networks, and ensuring accountability.
Recommendations and a toolkit guide future data re-use projects in developing countries, emphasizing citizen engagement methods for establishing a Social License and ensuring responsible data practices.

Keywords:
Agency, Asymmetries, Citizen Participation, Colombia, Consent, Data Reuse, Data re-use, Development, Low-income Countries, Middle-income Countries, Nepal, Open Data, Participatory Framework, Public Engagement, Senegal, Social License, Social License Framework, Use Cases

Areas:
Colombia, Senegal, Nepal

Acknowledgments

We would like to acknowledge those who contributed to the research of this paper by participating in the interview process: Igbal Safarov, Sille Sepp, Tiago Peixoto, Shashish Maharjan, Ronda Železný-Green, Isabel Cardenas, Luis Armando Muñoz, Steven Sotelo, Daniel Jimenez, Christopher Wilson, and Natalia Carfi. 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 Peter Addo, Claire Lavielle, Mirka Snopcokova, Amandine Kashani-Poor, Raphaëlle Sardier, Cecilia Poggi, Alexis Frémeaux, Ben-Hur Kabengele, Samantha Richard, Pierre-Arnaud Barthel, Eric Beugnot, Bertrand Savoye, Etienne Charrière, and Adrien Lorenceau.

We are also grateful to the peer reviewers who provided invaluable input on a semi-finalized draft, including Jerome Atutornu, Camilo Cetina, Anastesia Taieb and Alek Tarkowski.

Your valuable participation and insights have been instrumental in shaping the findings and conclusion, and we express our sincere gratitude for your assistance and invaluable contributions.

Executive Summary

The current era of datafication, which transforms multiple aspects of life into digital data, offers significant opportunities for sustainable development, especially in developing countries. Re-using existing data for societal benefits is particularly promising for developing countries given existing resource constraints, encouraging:

- Enhanced Governance and Policy-Making: Effective and responsible re-use of data can lead to better policy decisions in healthcare, education, and economic development, helping to address specific sectoral need,
- Economic Development: Responsible data re-use practices can drive economic growth by aiding informed decision-making in businesses, creating jobs, and attracting investment; i.e. re-using satellite data can improve agricultural efficiency and productivity,
- Social Equity and Inclusion: Responsible re-use of data can help identify and address social disparities, allowing developing countries to distribute resources more effectively and reduce inequality,
- Crisis Management and Response: During emergencies such as natural disasters or health crises, access to accurate and timely data is vital for effective response and mitigation.

Nevertheless, territories and regions may encounter distinctive challenges that impede the full realization of these opportunities, such as poor data quality, underdeveloped technological infrastructure, and limited institutional capacities.

In the context of developing countries, a critical yet often overlooked challenge in maximizing the benefits of data re-utilization is the existence of agency asymmetries. These asymmetries, referring to imbalances in power and influence among different stakeholders, significantly impact the manner in which data is managed and utilized. This issue gains heightened importance in scenarios where data, originally collected for specific purposes, is repurposed for other uses. In such cases, the original data subjects, particularly those from marginalized communities, frequently lack the ability to influence or even be aware of these secondary uses. Consequently, data is at risk of being utilized in ways that disproportionately benefit a select few, potentially leading to exploitation and the extraction of value at the expense of the broader community.

These agency asymmetries not only pose a threat to the equitable re-utilization of data but also severely undermine public trust in the institutions responsible for data governance. When people perceive they are in a situation where their data is being used in ways that are unfair, non-transparent, or do not include their participation or benefit, this situation cultivates skepticism and resistance towards data re-use. This erosion of trust is a significant barrier to the effective and ethical re-utilization of data for purposes of sustainable development in developing countries. Therefore, addressing these asymmetries is crucial in ensuring that data re-utilization aligns with principles of equity, inclusivity, and serves the broader goals of societal progress and development.

Today's dominant data governance approach to managing agency asymmetries is linked to the principle of individual consent. This method, however, is increasingly being questioned for its efficacy and appropriateness. One key limitation is its focus solely on individual preferences, often overlooking the broader implications of data re-use for community welfare and collective interests. This oversight is particularly problematic in societies with a collective orientation, where prioritizing individual choices may disrupt the societal fabric. Additionally, present consent models typically offer limited information into how data is collected and repurposed and reduce complex decision-making to a binary yes-or-no choice. Such an approach neglects the wider consequences and potential advantages of data re-use, especially in scenarios that transcend individual interests and impact communities or society at large.

To overcome these shortcomings, this position paper proposes the development of a 'Social License' model for data re-use, with a specific focus on its application in developing countries. Drawing inspiration from practices in sectors like the extractive industries, the Social License concept revolves around fostering sustained public engagement and building trust to legitimize data re-use practices.

Social License approach is committed to ensuring that data use adheres to democratic principles, maintains transparency, and reflects community preferences. It aims to establish more equitable and legitimate data re-use practices that can effectively contribute to development across diverse sectors.

At the heart of the Social License framework is the shift from relying solely on individual consent to complementing it with community engagement and consensus building. This significant shift towards collective decision-making is essential for fostering responsible and equitable data re-use practices. Emphasizing community self-determination, the model seeks to address and bridge the existing trust deficit between institutions, governments, and the public. This change also represents a more strategic and targeted approach to harness the potential of data re-use for sustainable development and societal advancement.

This position paper proposes an analytical and actionable framework for establishing a Social License in data re-use. The framework is structured around five critical components that need to be considered when seeking to establish a Social License:

- **WHO:** Stakeholder Identification: identifying relevant stakeholders based on the nature and expected outcomes of re-using data. This may include data subjects, community representatives, experts, public officials, and data holders,
- **HOW:** Engagement Methods: determining the participatory approach that may be appropriate, including Information Dissemination, Consultation, Active Discussion, and Co-decision processes,
- **WHEN:** Timing in Data Life Cycle: determining when to involve stakeholders at various stages of the data life cycle such as collection, processing, analyzing, sharing and/or using of data,
- **WHERE:** Contextual Factors: understanding the demographic, socio-economic, and socio-political contexts, along with technological accessibility and existing levels of trust in institutions,
- **WHAT:** Format of Social License: deciding what format the Social License should take, such as a formal contract/document or an informal agreement. This clarity helps set clear expectations and boundaries for parties involved.

The paper further discusses success factors and challenges, drawing insights from case studies in Colombia, Senegal, and Nepal. These examples illustrate the effective use of stakeholder engagement in data re-use projects, including agricultural data for weather prediction in Colombia, census and meteorological data for climate warning systems in Senegal, and geospatial data for disaster management in Nepal. The cases underscore the significance of:

- **Tailored engagement:** so that stakeholders are effectively engaged throughout key stages of data re-use projects, the method of citizen engagement employed should be tailored to each project,
- **Understanding of local context:** knowledge of contextual indicators and local networks allows project leaders to implement the required tools and human infrastructure to ensure the long-term success of a project,
- **A collaborative stakeholder approach:** taking the expectations of relevant stakeholders into consideration was key to the successful implementation of the projects.

However, these initiatives also faced obstacles such as:

- **Limited outreach:** factors that impair project implementation and outreach, such as a lack of adequate internet access for participants, may be present in developing countries. Alternative methods of information dissemination may be utilized to overcome this obstacle.
- **Inadequate infrastructure:** less developed technological and human infrastructure in developing countries can also affect the success of data re-use projects. Strategies such as building stakeholder capacity and the development of interfaces that are adapted to local digital literacy skills can allow for this barrier to be overcome.
- **Challenges in building trust:** communities may be distrustful or hesitant to be involved in data re-use projects, particularly in rural areas. Building trust between participants when establishing citizen engagement is crucial for successful implementation.

To address these challenges and advance the implementation of a Social License in data re-use, the paper suggests the following:

- **Encompass context-specific approaches:** Tailor citizen engagement strategies to the unique social, cultural, political, and economic contexts of each project, leveraging local infrastructure and social networks,
- **Implement discussion and/or co-decision engagement participation:** Actively involve local stakeholders in discussions, design, and/or implementation of data projects, using participatory approaches like Deliberative Discussions and Citizen Assemblies,
- **Incorporate alternative communication channels:** For areas with limited digital access, use alternative methods like radio, local meetings, and printed materials for information dissemination and feedback,
- **Build and maintain trust:** Engage local leaders in the process, addressing community concerns transparently, and communicate the benefits of data sharing and re-use. Continuously maintain this trust through open communication and feedback mechanisms,
- **Strengthen local capacity and infrastructure:** Improve local data literacy and technical skills by involving community members, such as university students and first responders, in data-related activities,
- **Create collaboration and partnership:** Establish cross-sector partnerships to share expertise, best practices, and ensure sustainable project execution, emphasizing local leadership and existing networks,
- **Focus on developing deliverables from public engagement:** Ensure that public engagement results in concrete deliverables, such as frameworks or licensing conditions, outlining stakeholder preferences and expectations about data usage,
- **Consider artificial intelligence (AI):** Explore how public engagement can be used to establish a Social License for AI, particularly in the context of training data for Large Language Models (LLMs).

Finally, the paper embeds the framework, lessons learned and recommendations into a Social License Tool Public Engagement Toolkit, which seeks to guide policymakers, development agencies, and other stakeholders. The Toolkit is an invaluable tool in effectively implementing citizen engagement in data re-use for development purposes in developing countries, addressing agency asymmetries and fostering a more equitable data ecosystem.

1. Setting the Stage: Data, Development, and Agency

1.1. The Potential of Data Re-use for Development

Over the course of the past few decades, we have entered an era of datafication, where human behavior and activity is quantified, digitized, and measured as a direct result of technological advancements. It is estimated that around 328.77 million terabytes of data are created every single day, generating about 120 zettabytes a year (Duarte, 2023). This era has given rise to a number of opportunities and developments around the globe. Re-using existing data for societal benefits is particularly promising for developing countries given existing resource constraints, encouraging:

- Enhanced Governance and Policy-Making: Effective and responsible re-use of data can lead to better policy decisions in healthcare, education, and economic development, helping to address specific sectoral need,
- Economic Development: Responsible data re-use practices can drive economic growth by aiding informed decision-making in businesses, creating jobs, and attracting investment; i.e. re-using satellite data can improve agricultural efficiency and productivity,
- Social Equity and Inclusion: Responsible re-use of data can help identify and address social disparities, allowing developing countries to distribute resources more effectively and reduce inequality,
- Crisis Management and Response: During emergencies such as natural disasters or health crises, access to accurate and timely data is vital for effective response and mitigation.

The potential benefits – and some of the risks – of data re-use have become a subject of increasing interest in recent years. Among other positive outcomes, data re-use can lead to cost savings, improved decision-making, greater innovation (i.e., in research), and cross-sectoral or cross-organizational insights. Sille Sepp from MyData Global commented "The more we enable data sharing, especially across sectors, the more meaningful it becomes by fostering more informed decisions and policy-making. It is equally crucial to make data available for citizens to empower them with information and insights, and provide them with necessary tools to control data about themselves."

An example of these benefits can be found in Namibia, where researchers re-used satellite and cell phone data to develop a map that showed the spread of malaria and the locations of where elimination efforts were most effective. By using the shared anonymized cell phone records of 1.2 million users from Mobile Telecommunications Limited, the researchers were able to create detailed maps of the movements of its citizens. This effort allowed for the identification and provision of bed nets to 80,000 people at the highest risk of being exposed - and becoming transmitters - during the malaria transmission cycle in 2013 (Guay, 2017). The result of this project was that fewer people were infected with malaria.

Although there are several benefits associated with re-using data, there are also risks and challenges, including possible violations of privacy, harm resulting from data breaches, inaccurate or outdated information, and potential mis-use or misinterpretation of data. In developing countries, these associated risks and challenges are not only present but may also be more pronounced due to other limitations such as infrastructure limitations, inadequate regulatory frameworks, low digital literacy, and variable data quality. Economic constraints further hinder investment in data management and security, while cultural and social norms may conflict with global data protection standards. Additionally, reliance on external data sources and potentially unstable political environments can compromise data sovereignty.

According to Ronda Železný-Green from Panoply Digital: "So much has been extracted from there that the resources to be able to do a lot of work like this (since data by its nature is resource-intensive), are not available. And so what this causes is a lot of the governments in developing countries are not able to use or engage in evidence-based decision making because they were robbed of the necessary inputs." Under these conditions, i.e. lack of infrastructure, resources, and human capacity, implementing data re-use projects and initiatives can result in more risks and challenges than benefits to a community or country.

1.2. Asymmetries in the Data Ecosystem and Their Implications

Despite the abundance of data and the clear potential benefits in re-using it, the data ecosystem remains hampered by a number of inequalities and asymmetries, many of which exacerbate existing socio-economic divisions. These asymmetries severely limit the potential of re-using data for the public good. Mitigating these asymmetries is therefore crucial to deploying data more generally, but perhaps especially in a development context.

Data asymmetries can be defined as the divide or disproportion of access to data (Verhulst & Young, 2022). The development of this divide is influenced by the interaction between those providing data and those on the receiving end, which look to benefit from accessing, using, and re-using data. There are five different ways in which data asymmetries can manifest (Verhulst & Young, 2022):

1. Business-to-consumer (B2C): B2C data asymmetries arise when organizations or companies gather an extensive amount of user data during the interaction of services or sale of goods. This data include, transaction history, browsing activities, and other important socio-demographic details. In some cases, users are not fully aware of the extent in which their data is being shared.
2. Business-to-business (B2B): Data monopolies, known as B2B data asymmetries, have emerged, overriding companies, sectors, and the broader economy. These companies have access to extensive data originating from various sources and use it to create barriers to entry through data combination and machine learning. As a result, there is suppressed innovation, reduced competition and potential harm to consumer rights. For this reason, there have been calls for increased regulation and stricter enforcement of antitrust laws.
3. Business-to-government (B2G): B2G data asymmetry refers to the extent in which governments are able to access important datasets that are being held by companies. B2G data asymmetries have increasingly become a focus point of policymakers due to the lack of access to data and insights that are currently being held in the private sector. Their aim is to enable better government decision-making and service delivery using this information.
4. Government-to-citizen (G2C): G2C data asymmetries refer to siloed data collected by the government, reducing transparency as well as limiting the ability for citizens to gain insights and value from it.
5. Business-to-science (B2S): B2S data asymmetries occur when privately held data is unavailable to scientific research. This is due to information remaining siloed over business' concerns with regard to competitive advantage, privacy harms and more.

What can be found in these data asymmetries, is a disproportionate amount of agency between stakeholders. Agency asymmetries can be defined as imbalances in power and influence among different stakeholders, significantly impacting the manner in which data is managed and utilized. This occurrence is widespread and is recognized as a key feature of the data economy, where network effects can quickly give rise to power asymmetries concerning data. This issue gains heightened importance in scenarios where data, originally collected for specific purposes, is repurposed, and re-used, for other uses. In such cases, the original data subjects, particularly those from marginalized communities or developing countries, frequently lack the ability to influence or even be aware of these

secondary uses. Consequently, data is at risk of being utilized in ways that disproportionately benefit a select few, potentially leading to exploitation and the extraction of value at the expense of the broader community.

These agency asymmetries not only pose a threat to the equitable re-utilization of data but also severely undermine public trust in the institutions responsible for data governance. When people perceive they are in a situation where their data is being used in ways that are unfair, non-transparent, or do not include their participation or benefit, this situation cultivates skepticism and resistance towards data re-use. This erosion of trust is a significant barrier to the effective and ethical re-utilization of data for purposes of sustainable development in developing countries. Therefore, addressing these asymmetries is crucial in ensuring that data re-utilization aligns with principles of equity, inclusivity, and serves the broader goals of societal progress and development.

In developing countries, agency asymmetries can manifest in various forms, most notably when private companies, normally based in developed countries, benefit from data collection in developing countries without any accountability or consequences. This type of practice resembles past colonial resource extraction patterns, where the data holder is unable to obtain any type of benefit from sharing their data (Muyoga, et. al, 2022). In these circumstances, there is a need to develop new approaches to address data asymmetries to promote more equitable data practices, and fully realize the potential benefits of data re-use for the common good.

1.3. New Approaches to Tackling Agency Asymmetries: From Consent to a Social License

How can agency asymmetries be solved, or at least mitigated? Current approaches focus on the notion of informed consent as a solution to the problem of agency asymmetries. Informed consent is usually presented in a “tick box” format with an extended Privacy Policy document to accompany it. In many cases, this attempt to provide informed consent is seen as sufficient as most of the information is presented to the user. However, there are drawbacks to this approach, which range from a lack of user data literacy to users only being provided two options with regard to sharing data: users either opt-in, or opt-out.

Despite their widespread deployment, current theories and practices of consent fail to adequately address the problem of agency asymmetries. A large part of the problem stems from the fact that current practices are narrowly focused on concerns such as privacy and commercial re-use, failing to take into account the complex web of causes and associations that contribute to asymmetries. Consent models fail to provide the contextual information necessary for individuals and communities to understand how their data will be used and re-used. These agreements are often binary, with individuals agreeing to terms without fully comprehending the conditions under which their data is being stored, distributed, used, and re-used, and for what purpose (Lindegren et al., 2019). As an example, consider standard toolkits of consent procedures when registering with a service provider. These frameworks typically use methods like checkboxes and privacy notices. However, this existing design is limited, offering users only two choices: either “accept all conditions" to access the service or “do not accept”, resulting in the inability to use the service. This setup forces users to consent to the collection of all their data without providing nuanced choices (Lindegren et al., 2019).

Furthermore, consent is typically given by individuals, overlooking the ethical and policy-related debates around data repurposing, and the broader impact the re-use of data can have on groups and communities. For example, refugee data that is collected based on an individual consent model may then be used for decision-making that impacts whole groups or populations of refugees - including those that did not consent to the initial data collection process (Verhulst et. al, 2023). The lack of more nuanced consent methods, presented along with granular and contextual information, is critical in perpetuating and often exacerbating agency asymmetries (Verhulst et. al, 2023). Such problems may be especially relevant in developing countries, where there may be a digital divide in access to technology and data literacy as well as varying views on data being collective rather than individualistic (Boscarino et al., 2022).

In order to overcome the problems posed by agency asymmetries, it is essential to move beyond only using simplistic, binary, and individualized notions of consent. In this context, the concept of a Social License framework emerges as a tool that can contribute and complement to building a common, responsible framework for informed agreement between data subjects, data controllers, and data processors. A Social License framework allows for the strengthening of trust between key stakeholders and, over time, helping mitigate agency asymmetries.

1.4. A Social License Framework in a Data Context

The concept of a Social License, sometimes referred to as a Social License to Operate (SLO), emerged in the 1990s to describe the processes and authorizations necessary for companies in the extractive industries, such as mining, to commence their operations, especially in developing countries (McConaughey et al., 2023). The acquisition of these licenses require buy-in and engagement with a range of relevant stakeholders, including local residents, influential community members, organized interest groups, and governmental actors, all of whom might be impacted by the project's operations. SLOs were deemed to be important for these extractive industries due to the immense impact they would have on the environment, for example through waste disposal or dust and noise pollution. Therefore, in order to reduce pushback against these projects, it was seen as essential to involve relevant community and governmental stakeholders in conversations about how to approach the projects.

The concept of a Social License can be applied in the context of data as a way of building a more responsible framework for data re-use, helping mitigate some of the typically resulting agency asymmetries. The definition that has been offered for a Social License for Data Re-use *is the process of building trust and legitimacy from ongoing (i.e., constantly renewed) citizen engagement and acceptance of how data is being accessed and re-used.* A Social License framework focuses on ensuring that data is re-used ethically, transparently, and in ways that align with the preferences and the priorities of the public. While the Social License has been applied in other contexts, its relevance is only now emerging in the area of data re-use. The unique characteristics of the Social License holds the potential to successfully implement data re-use projects which aim to address current societal challenges, especially in developing countries. However, since the Social License is relatively new to the area of data re-use, there are limited instances showcasing its capability.

Social License in Data Re-Use

A Social License is an implicit societal acceptance of an activity based on trust, legitimacy, and perceived alignment with community values. It differs from open data licenses[1], as it is not legal in nature but rather is an agreement or plan that is developed through stakeholder engagement and collaboration. In the context of data, it captures collective consent by focusing on the tacit endorsement of individuals and communities for the collection, use, and re-use of their data. As such, a Social License expresses preferences or expectations and can support and accelerate the creation of more formalized contractual or licensing approaches.

Relying solely on individual consent for data re-use can leave individuals and communities vulnerable to exploitation and disempowerment, particularly in situations where power asymmetries exist (i.e., corporations vs. users). This underlines the critical role of Social License as a complementary approach. By fostering collective understanding, trust, and shared values regarding data re-use practices, a Social License empowers communities to hold data actors accountable and ensure ethical data use (Verhulst et al., 2023).

[1] Open data licensing (or data licensing) refers to the act of granting or acquiring the legal authorization to use or access data. This process often involves a contractual agreement between the data provider and the data user - with the purpose of determining the conditions in which data can be utilized, accessed, and shared.

1. A Social License is a Dynamic and Multifaceted Construct:

- Beyond Permission: Unlike consent, which focuses on individual authorization, Social License is a broad, dynamic concept encompassing societal acceptance and implicit approval of data practices. It's not a static permission but a constantly evolving understanding shaped by ongoing dialogue and trust dynamics between individuals, communities, and data actors (Coleman et al., 2022),
- More Than Just Legality: While legal frameworks provide baseline regulations, Social License goes beyond legal compliance. It's about legitimacy in the eyes of society, built upon trust, fairness, and alignment with shared values. For example, even if data collection is legally permissible, it might lack Social License if perceived as unfair or violating cultural norms (Allen & Wilson, 2008),
- Contextually Dependent: Social License is not a universal concept but varies across cultures, communities, and data uses. What constitutes acceptable data practices in one context might not in another. Understanding these nuances is crucial for building trust and achieving data justice (Jobin & Gagnon, 2014).

2. Key Pillars of a Social License for Data:

- Fairness and Equity: Data practices should ensure equitable distribution of benefits and minimize potential harms, particularly for marginalized communities. This might involve data-driven solutions addressing inequalities in healthcare, education, or financial inclusion (OECD, 2020),
- Transparency and Accountability: Data flows, decision-making processes, and potential risks should be transparent and open to public scrutiny. Robust mechanisms for accountability, including redress for data misuse, are essential (European Commission, 2016),
- Sustainability and Responsibility: Data use should be sustainable, considering long-term impacts on individuals, communities, and the environment. Ethical considerations regarding privacy, security, and data ownership must be prioritized (Data & Trust Alliance, 2020).

3. Examples of Building a Social License for Data Re-use:

- The Data Assembly: An initiative by the GovLab, The Data Assembly seeks to utilize public input on data re-use for crisis response in the United States. In 2020, the format was used to create recommendations for guiding a response to the Covid-19 pandemic in New York City, based on the inputs of three 'mini-publics' comprising data holders, policymakers, advocacy and civil rights representatives, and local residents (Young et al., 2020),
- The Use of Linked Administrative Health Data Sets in Canada: Paprica et al.'s 2019 study found that although members of the public in Ontario generally support research based on linked administrative health data, there was no blanket approval. The study demonstrates that researchers and organizations that hold and use health data should do so in alignment with a Social License, and engage with members of the public to address their concerns around privacy and security and increase public trust and support,
- Linking Patient Health Datasets in the UK: A 2023 study by Ford et al. examined how anonymised patient health data sets from distinct medical services in Southeast England could be linked and used by public health experts for research and planning. The researchers found that citizens largely supported the use of linked data sets as long as their expectations about transparency and public engagement were met, which could be successfully mitigated by the use of a Social License.

Citizen engagement is an essential component of a Social License. Meaningful involvement of the various stakeholders in an initiative, including data holders and data subjects, can help address asymmetries in a data ecology, for instance, by giving stakeholders an opportunity to share information, understand each others' concerns and imperatives, and overall help build trust, transparency and accountability. Citizen engagement also allows for greater explainability, which is necessary to prevent asymmetries in information and understanding between data controllers, processors, and the general public (Aitken et al., 2020; Muller et al., 2021; Choo et al., 2021). At a broader level, the two main advantages of citizen engagement are: 1) redistributing the costs and benefits of data re-use in a fairer manner, allowing currently disadvantaged stakeholders to bear less costs and benefit more, and 2) redirecting some of the attention from individual consent to a more deliberative, informed, and group-based decision-making process.

Establishing a Social License in developing countries is important for the following reasons.

1. Vulnerable Populations: developing countries often face higher data vulnerabilities due to limited resources, weak legal frameworks, and digital literacy gaps. A Social License approach can empower communities to have a say in data use and protect themselves from exploitation (Jobin & Gagnon, 2014),
2. Locally Tailored Data Governance: Imposing external models without considering local contexts can exacerbate inequalities. Building a Social License through participatory processes ensures data governance frameworks reflect local values and needs (Degeling et al., 2020),
3. Addressing Power Asymmetries: Communities in developing countries often lack the same bargaining power as data-collecting entities. A Social License approach can shift power dynamics by promoting transparency, accountability, and collective decision-making (Coleman et al., 2022).

A Social License is a crucial concept in navigating the complex landscape of data in the digital age. By prioritizing transparency, fairness, and community engagement, individuals and communities have more agency over their data and reap the benefits of data-driven solutions. Without it, developing countries would struggle to ensure ethical and inclusive data governance, hindering their ability to achieve digital justice and sustainable development.

1.5. Scope of Position Paper

Given the growing relevance of datafication, as well as the potential benefits and risks of data re-use for achieving development goals, the primary purpose of this paper is in assisting policymakers and development agencies to better understand–and implement–the potential of data re-use projects in developing countries. To that end, the aim is to examine prevailing conditions and provide informed recommendations to help foster Social Licenses; in particular, by fostering public engagement through various participatory methods. These conditions include: the different actors that need to be engaged, the different approaches for public participation, the different moments of engagement in The Data Life Cycle, the different influences of contextual factors, and the different agreement outcomes of participation.

Although agency asymmetries are a concern shared by many countries, data practices and ecologies vary across countries, and some conditions are specific to developing countries. Developing countries face concerns over a lack of quality data, less robust infrastructure, and reduced resource and institutional capacity. Cultural and social norms and contexts also vary, both between developing countries and developed countries, and among developing countries themselves. As such, this paper examines how the contextual factors and conditions unique to developing countries should be considered in establishing a Social License for data re-use.

2. Understanding a Social License Framework

In the context of Europe's General Data Protection Regime (GDPR) and other data governance frameworks, a Social License would involve more active and meaningful participation by citizens in the decision-making processes about how their data is used and re-used beyond consent. This can take the form of public consultations, user feedback mechanisms, and transparent reporting practices. By doing so, organizations can not only comply with the legal requirements of frameworks like GDPR but also gain public trust and legitimacy. The application of a Social License in data governance is about going beyond the legal compliance of frameworks like GDPR and engaging with the public in a way that upholds the spirit of these laws and governance frameworks. By doing so, organizations can achieve a higher level of trust and legitimacy, which is beneficial not just for regulatory compliance but also for building a more ethical and responsible, data-driven society.

Social License and Legal Developments in Europe

Recently in Europe, several expert groups have investigated ways to advance the re-use of data, specifically privately held data, for public interest purposes. Their aim is to prepare legislation to streamline access for re-use. A central recommendation in this approach was to establish a Social License, focusing on public engagement.

The High Level Expert Group on B2G Data Sharing (European Commission, 2018) was a group established by the European Commission in 2018. The group focused on facilitating access to and re-use of private sector data for public interest purposes. Its objectives included identifying good practices in B2G data sharing, assessing legal, economic, and technical obstacles to such sharing, and providing recommendations to the Commission on further developing its B2G data sharing policy. The group's work acknowledged the relevance of data held by various private entities, like telecom operators, online platforms, and retailers, which can be integral in addressing public challenges such as epidemics, urban planning, and environmental protection. Some of its recommendations formed the basis for the Data Act (European Commission, 2024), which was adopted on 13 December 2023 and published in the Official Journal of the European Union on 22 December 2023 (although the scope of the data access recommendations were heavily watered down in the final version). The Data Act complements the Data Governance Act (European Commission, 2023), which focuses on the transfer of non-personal data and rules around the reuse of public sector data. While the Data Governance Act provides a legal framework for data sharing, the Data Act clarifies who can create value from data and under what conditions. A key finding of the High Level Expert group was that, "Citizen engagement in B2G data-sharing relations is underdeveloped" (European Commission , 2020).

Another group that has been instrumental in socializing the need for a social license involves the Expert Group on "Facilitating the use of new data sources for official statistics", established by Eurostat. Its directive was to provide advice and assistance to Eurostat on issues related to access to privately held data, particularly in the context of the new roles of statistical offices in a data-driven society and economy. The group's work included reviewing implications for official statistics of general policy developments regarding access to privately held data, reflecting on how these policy developments could be extended or complemented for statistical purposes, examining specific data sources and issues, and producing recommendations for policy initiatives specific to the statistical domain. These recommendations have since been integrated in newly proposed general statistical legislation (Eurostat, 2023). The group also provides operational guidance on applying or adapting principles on B2G data sharing when data are used for statistical purposes. In its report, the Expert Group emphasized the importance of a social license for data re-use (Eurostat, 2022).

2.1. WHO: Different Actors that need to be Engaged

Citizen engagement is the foundation for any Social License. Involving communities and citizens allows policymakers to make more legitimate and equitable decisions by providing better and more effective data (UN, 2013). However, it is important to note that depending on the nature of the project, initiative, or policy, it is not always necessary to call for the entire community or all local residents to be involved. Instead, certain stakeholders, professionals, experts, or other knowledgeable groups within a specific field relevant to the project might be more suitable in discussions or decision-making processes. For instance, consider a project centered around agriculture development. In such cases, seeking input from farmers would be the most valuable and pertinent, as they not only possess expertise and practical knowledge in the subject but are also the most affected public stakeholder group. It's important to recognize that expertise goes beyond formal education to include the lived experiences of individuals, offering unique insights into complex issues.

2.2. HOW: Different Approaches for Public Engagement

Effective engagement methods are those that include all affected communities, including marginalized groups and that avoid implementing top-down approaches that can perpetuate inequalities (Boersma, 2020). According to the OECD (2022), citizen participation has key benefits, including: (1) increasing transparency, inclusivity, legitimacy, accountable policy-making, and increased public trust in an institution; (2) increasing up-to-date solutions by tapping into collective knowledge and citizen information (3) increased inclusivity and diversity policy-making by strengthening the representation of minorities, and (4) enhanced legitimacy and social support for change. Applied to data re-use, by enhancing citizen engagement organizations can better adhere to the values of transparency, accountability, and equity, which are key to effective data governance but are often inadequately implemented.

Participatory engagement can take many forms, showing wide variabilities depending on the use case and the socioeconomic or cultural context (among other variables). Figure 1 illustrates a taxonomy of participatory approaches for public engagement. This figure illustrates that the extent of citizen involvement ("Minimal," "Intermediate," "High") is one of the most important axes of variability. Within these broad variables, participatory approaches can further be classified into four categories: Information, Consultation/Feedback, Discussion, and Co-decision.

- **Information** refers to the process of bringing the public's attention to the current data and decisions around a project. This can be done via press articles, websites, letters, brochures, public meetings, etc. The aim of the information dissemination is to provide a clear, complete, and comprehensive understanding of the project at hand to the public,
- A **Consultation/Feedback** participatory approach focuses on interactions that allow stakeholders to gain insights into a project, initiative, or policy by providing a platform for the public to voice their opinions on the matter at hand. This can be done at the beginning of an initiative (an informative approach) to understand the public's position, and can also be done throughout the initiative (a feedback approach) in order to receive continuous public reactions and assessments. The goal of an information/feedback approach is to increase transparency in decisions that may affect the public. In general, this participatory approach is considered "minimal" as it does not provide the ability for the community or citizens to engage with other stakeholders and does not provide the ability to be part of the decision-making process. Some examples include: deliberative polling, surveys, public meetings, and briefings,
- **Discussion** can be regarded as an "intermediate" level of engagement. Discussion methods provide communities and citizens with the opportunity to be part of the early decision-making process. The approach allows for the collection of information and the consultation of citizens on certain subjects, such as data re-use. This method permits for diverse perspectives, expertise, and insights to be heard and discussed to shape the direction of projects or policies. Examples include: focus groups, juries/panels, deliberative discussions, and citizen's assemblies,

- **Co-decision** is a process where the authorities and citizens make joint decisions and ments through the mechanisms of direct democracy. Co-decision can be accomplished through two avenues:

  1. Co-production or partnership, is the joint elaboration of a project, where citizens take part in the realization of the project with technicians and work out solutions together,
  2. Delegation is a process where the public authority delegates part of its power to citizens and accepts to be involved in the decisions taken by them.

As Co-decision values consist of insight and input, it is considered to be on the higher side of the scale of participatory engagement. Examples include committees, official advisory groups, and community participation.

Figure 1 – Taxonomy of Participatory Approaches for Public Engagement

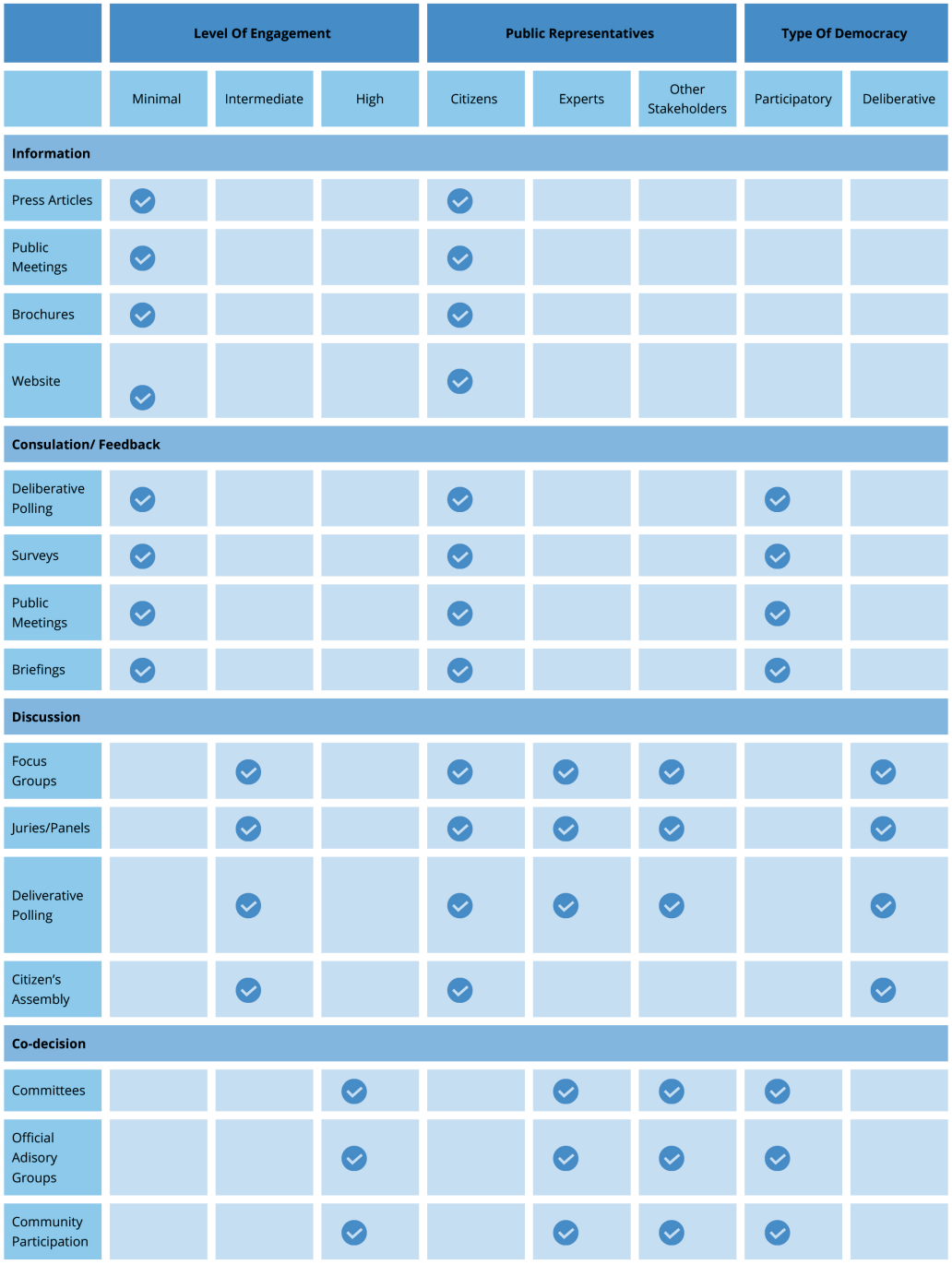


Figure 1 was created to visually demonstrate the classification of the different participatory approaches for citizen engagement.

2.3. WHEN: Different Moments of Engagement Along The Data Life Cycle

When re-using data for a different purpose than what it was originally collected for in a new project, the data is subject to go through The Data Life Cycle again. In this process, data is engaged and interacted with differently at each stage. These stages are:

1. **Planning Research:** refers to the action of identifying the data that is needed for a purpose, project or initiative,
2. **Collecting Data:** involves instruments and methods for experiments, observations, surveys, etc. to gather the desired data,
3. **Processing and Analyzing Data:** require the data to first be cleaned, noise eliminated or transformed and then to be interpreted to find results,
4. **Sharing:** includes a data citation or indication stating where and under what conditions the data can be accessed,
5. **Preserving Data:** is a process in which data is archived in a suitable location (locally or non-public) like a data center or repository. This process additionally can require data to be involved in quality assurance, format conversion, required access control, etc.,
6. **Re-using Data:** is a process in which data is made accessible so that it can be used again for other purposes to generate new insights, decision-making around policy or commercial use.

Depending on the project, citizens and stakeholders can be engaged at various stages of The Data Life Cycle. Involving different actors at specific periods of the cycle facilitates meaningful contributions in shaping decisions and outcomes around the data being re-used. This process enhances the confidence in meeting the expectations of the different stakeholders by developing trust through increased transparency and accountability in the process. Determining when stakeholders should be involved in the Data Life Cycle depends on many variables such as the aim of the project, the types of actors involved, and the contextual factors that influence participation.

2.4. WHERE: Different Influences of Contextual Factors

Understanding the influence of contextual factors on participatory approaches is vital for the success of citizen engagement. Contextual factors can include factors such as demographics, cultural norms, socio-economic backgrounds, past and current political dynamics, historical injustices, use of technological tools, current events, or past events that have fueled mistrust. By recognizing these influences, especially in developing countries, practitioners and policymakers can tailor citizen engagement methods to be more culturally sensitive, inclusive, and relevant to the needs of the communities involved. In addition, a contextual analysis based on these factors allows for the identification of potential barriers to participation and the design of strategies that allow for maximum involvement and empowerment of communities.

Socio-demographic factors include variables such as age, gender, population distribution, and education level, each of which significantly influence citizen participation. For example, studies have found that male citizens are more likely to engage with Open Government Data (OGD) and concepts of data re-use due to higher levels of education and generally more active participation in political and public spaces (Saxena & Janssen, 2017). Citizen engagement is also strongly influenced by the capability (Purwanto et al., 2019; Smith & Sandberg, 2018) and competence (Wirtzetal, 2018) citizens have in using technology.

Inequality levels and Data Literacy are two variables within the Socio-economic factor that influence citizen's willingness to participate in. In developing countries or countries with considerable disparity between economic statuses, citizens are less likely to engage due to factors such as lack of trust, perceived power imbalances, difference in priorities, unequal representation, and limited access to

resources (Purwanto, et al., 2020). In the context of data re-use, knowing the data literacy rate of a country is important as it helps further realize the impact of factors like limited access to resources and education around technology, data, etc. Data literacy can be measured through different variables such as access to internet connection and use of computers every day, these variables account for the implementation of technology such as the use of computers in the daily life of citizens and the possibility of using them through access to the internet. If there is a high internet coverage and a constant use of technological devices, it is possible to assume that citizens have a greater capacity to understand and get involved in debates about data, its re-use, and its benefits as a result of their environment.

Socio-political factors consist of three variables: corruption levels, political participation, and trust in institutions. Similar to the variables in the Socio-economic indicator, these variables can indicate to what extent citizens are willing to participate and engage in projects or policy-making initiatives. The higher the level of corruption and distrust in institutions, the less likely citizens are willing to participate. In addition to these factors, it is important to note the timing of the data re-use project research/ implementation. Current events that influence the country or community in which the data re-use project takes place, can be heavily impacted, influencing the willingness of citizens and other actors to engage.

2.5. WHAT: Different Social License Outcomes

Establishing a Social License can be accomplished in many ways. A Social License can either be developed as an official contract that details the needs, expectations, and conditions of parties involved, or an informal agreement that captures the collective consent of the stakeholders. As Social Licenses in the context of data re-use are an emerging practice, current agreements have been mainly informal in nature, especially in developing countries as data re-use projects and citizen participation are becoming more established. There's increasing awareness of the need for formal contracts to express a Social License to enhance structure, transparency, and trust, thereby preventing greenwashing— misleading claims of acting in the best interest. Formalizing Social License standards will be key to ensuring the integrity of data re-use projects.

3. Use Cases: Citizen Engagement in Data Re-use Projects

Implementing a Social License in data re-use projects or initiatives is a relatively novel concept. Nevertheless, there are a few projects that have sought to establish trust through participation. In the below, we explore a few case studies that demonstrate the variety in practice. As such we seek to gain an initial understanding of how citizens were engaged, when they were engaged, and which methods were used. Selecting the uses cases during the extensive literature review revolved around three primary points:

1. The project must include the re-use of data,
2. The project must be geographically located in a developing country,
3. The project must demonstrate effective and innovative approaches to engaging citizens or stakeholders.

While acknowledging that there could be many other possible cases, a total of three were found which matched the specified criteria. They involve:

1. Pronòsticos Aclímate Colombia: Re-use of Agricultural Data to develop a Climate Service Platform
2. Kaffrine, Senegal: Re-use of Census Data to Establish a Climate Warning System
3. Pokhara, Nepal: Data Re-use to Develop an Open Street Map for Disaster Management

3.1. Pronósticos Aclímate Colombia: Re-use of Agricultural Data to Develop a Climate Service Platform

- ➔ *WHO: Ministry of Agriculture and Rural Development, the National Institute of Hydrology, Meteorology, and Environmental Studies, Center for Tropical Agriculture-CIAT, and the Farmers' Associations (i.e. Fedearroz and the National Coffee Federation),*
- ➔ *HOW: Co-decision (Round tables and expert engagement),*

- ➔ *WHEN: Citizens were engaged and consulted at the stages of planning, processing and analyzing the data, publishing and sharing the data, and re-using the data. Citizens were also involved in the data collection stage, prior to the start of the project,*
- ➔ *WHERE: The project took place in Colombia, where data literacy levels are approximately 70% for access to the internet, and 45% for daily use of a computer. 42% of the population has reached a high-school level of education,*
- ➔ *WHAT: The project re-used agricultural data from different private and public stakeholders to develop a meteorological predictive system. The Social License that was established was an informal agreement between the project participants.*

The Aclímate Colombia project was a cross-sector initiative within Colombia that aimed to develop a digital agro-climatic forecast system. Its objective was to aid local farmers improve their crop yields and mitigate losses caused by the impacts of climate change by providing them with insights to make better planting decisions. The long-term goal of this initiative was to ensure food security in Colombia by using predictive data on weather patterns in different regions.

The project, led by the International Center for Tropical Agriculture (CIAT), re-used open government datasets from National Institute of Hydrology, Meteorology, and Environmental Studies (IDEAM) and data from private sector farmer associations to develop the system. The data originating from private sector farmer associations included the past 20 years of harvest monitoring, annual surveys, and records of cropping events (CCAFS and CGIAR, 2015).

The different stakeholders in this project included, the Ministry of Agriculture and Rural Development, IDEAM, the International Center for Tropical Agriculture (CIAT), and various farmers' associations, such as Fedearroz, the national rice farmer's association, and the National Coffee Federation. The methodology of CIAT included developing round tables between scientists, experts and farmers, using a co-decision engagement approach to understand how to successfully implement the insights provided by the predictive agro-climatic system. Additionally, the round tables served to raise awareness about the potential uses of data sets, and to improve data skills and literacy for the members of the associations. These engagements took place over a period of three to four months, and were key to ensuring the sustainability of the project by teaching critical skills within the associations (CCAFS and CGIAR, 2015).

Initially, the project was implemented in six rural localities across four departments in Colombia and then scaled to 34 rural localities across nine departments (Sotelo, et al., 2020). In the following year of the initial implementation of the initiative, farmers' decision-making led to the savings of about $3.6m (Young & Verhulst, 2017).

3.2. Kaffrine, Senegal: Re-use of Census Data to Establish a Climate Warning System

- ➔ *WHO: Agriculture and Food Security (CCAFS), National Civil Aviation and Meteorology Agency of Senegal (ANACIM), farmers, climatologists, agricultural scientists, NGOs and media,*
- ➔ *HOW: Discussion (workshops),*
- ➔ *WHEN: Citizens were engaged during the stages of planning, processing and analyzing data, publishing and sharing data, and re-using data,*
- ➔ *WHERE: Kaffrine, the main agricultural region of Senegal. In Senegal, approximately 42% of the population has access to the internet, and 49% of the population has a high-school level of education,*
- ➔ *WHAT: The project re-used census and meteorological data to create a climate warning system for farmers. An informal agreement was used as the basis for a Social License.*

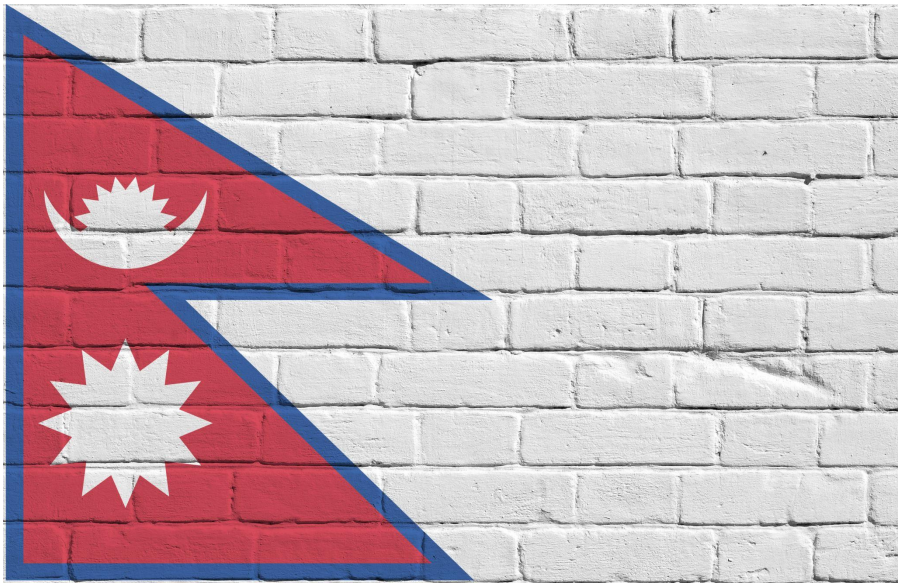
Over the course of the past few decades, Senegal has been heavily affected by unpredictable and irregular rainfall due to climate change (National Agency for Civil Aviation and Meteorology of Senegal et al., (n.d)). As a result, workers in the farming industry – the main source of income in the rural economy – experienced hardships. To address this issue, the Climate Change, Agriculture and Food Security (CCAFS)[2] and the National Civil Aviation and Meteorology Agency of Senegal (ANACIM)[3] joined forces to develop and establish a warning system for rural farmers. In 2011, a pilot project was initially implemented in the region of Kaffrine and then later expanded to regions of Diourbel, Fatick, Louga, and Thies. Between 2011 to 2014, the study concluded that 3.9 million rural people were potentially being reached through different channels. By 2015, the project extended to the rest of the country through an additional partnership with the Union des Radios Associatives et Communautaires du Sénégal (URAC)[4].

The objective of the initial project in Kaffrine was to re-use combined census-gathered farmer statistical information and meteorological data to establish a warning system for rural farmers, who relied on traditional methods. ANACIM and CCAFS conducted a survey to understand the methods that rural farmers used in Kaffrine, and the results indicated that CCAFS would have to build trust with the farming communities in order for the project to be successful (CCAFS & CGIAR, 2016). The project implemented a multidisciplinary group model allowing stakeholders, including farmers, climatologists, agricultural scientists, NGOs and media to work together in teams to develop solutions. The main aims of the engagements were: 1) providing local knowledge to scientists from traditional and indigenous stakeholders to fill in knowledge gaps regarding forecasting; 2) empowering local stakeholders to access climate information in decision making, and; 3) enabling local stakeholders access to weather forecasting and concepts used by forecasters.

[2] https://ccafs.cgiar.org/
[3] https://www.anacim.sn/
[4] https://uracsenegal.sn/page_accueil/

Challenges identified throughout the project included hesitation and pushback to using climate information as a means for forecasting by rural farmers. However, through stakeholder workshops, farmers were able to express their knowledge and provide feedback, helping scientists tailor the project to fit farmer needs while introducing climate information methodology. The expanded pilot project reached an overall estimated 7.4 million people by August 2015 nationwide, using a total of 82 rural community radio stations (CCAFS & CGIAR, 2016).

3.3. Pokhara, Nepal: Data Re-use to Develop an Open Street Map for Disaster Management



- ➔ *WHO: Kathmandu Living Labs, Local Government, Local Schools and Universities, Local Communities, Non-Government Organizations, the American Association of Geographers, and the United States Department of State,*
- ➔ *HOW: Discussion (workshops),*
- ➔ *WHEN: Citizens were engaged in the stages of data collection, data processing and analyzing, data sharing and publishing, and data re-use,*
- ➔ *WHERE: Pokhara, a metropolitan city in central Nepal. 91% of the population in Nepal has access to the internet, and 86% of the population has a high-school level of education,*
- ➔ *WHAT: The project re-used spatial data to create an Open Street Map for disaster management. The introduction of workshops on the use of the platform in schools and universities was key to increase data literacy and long-term adoption of the platform in future emergency situations. An informal agreement between project participants formed the basis of the Social License for this project.*

Pokhara Metropolitan City, located in western Nepal, is ranked as a natural disaster-prone area (Observatory for Public Innovation, 2019). After a 2015 earthquake in Nepal, it became evident that there was a need for creating resilient and planned cities with tools for effective response to natural disasters. In response to this need, Kathmandu Living Labs implemented a project to create a robust OpenStreetMap – an online platform – of the city for disaster resilience and emergency management in Pokhara[5].

[5] https://preparepokhara.org/#/amenities/hospital

The pilot project was funded by the American Association of Geographers and Human Information Unit, and involved the Humanitarian Information Unit from the US Department of State, the Ministry of Federal Affairs and Local Development of Nepal, the local government of Pokhara, and Kathmandu Living Labs. The local government of Pokhara managed the database, and local citizens were engaged to map their neighborhoods with instructions on how to use spatial data (Observatory for Public Innovation, 2019).

Kathmandu Living Labs held different workshops and training sessions to engage citizens and promote the use of tools by local schools, universities, and community partners. The project trained nearly 600 local people in the collection, usage, and dissemination of map data, disaster management, and open mapping. It also raised awareness of the value of open geospatial data, highlighting working on the ground with communities to generate more accurate geospatial datasets. By 2018, the project became a reference in the region and a hub of geospatial expertise, providing support to other cities (Observatory for Public Innovation, 2019).

4. Use Cases: Analysis

4.1. Pronósticos Aclímate Colombia: Re-use of Agricultural Data to Develop a Climate Service Platform

Findings:

➔ The Co-decision participatory approach and successful implementation of the project

This project implemented a *Co-decision Participatory Approach*, characterized by a high level of engagement with various public representatives. It involved a number of different stakeholders throughout the project lifecycle, from its inception to data collection, including the Ministry of Agriculture, IDEAM, local governments, farmer associations, and CIAT. The development of the digital platform was facilitated by workshops to explain the utility and benefits of climate predictive data. Serving as the project lead, CIAT involved farmers, in collaboration with the farmer associations, to conduct studies on the efficacy and uptake of the predictive model. In addition, CIAT worked directly with the associations to use their databases to improve the model. Importantly, CIAT also engaged the farmer associations in round table discussions to analyze and interpret the outputs of the prediction model. Daniel Jimenez from CIAT commented, "The interpretation of data, in collaboration with the associations, was key. It wasn't just about having people with digital and technical expertise analyze the information; it was about involving members from the associations to guide the process, as they are the ones who deeply understand the problems."

It was a priority for CIAT to develop user-centered digital tools that were intuitive and accessible to users, and they achieved this by evaluating the digital capacity of technicians and farmers when creating the Aclímate platform. Moreover, CIAT created a knowledge management unit to bring stakeholders together and create specific strategies to embed the processes in the association's structures. Luis Armando Muñoz from CIAT said, "A knowledge management unit was created in CIAT to help integrate all the stakeholders of the project. The purpose was beyond holding workshops but more about changing institutional structures and strengthening institutional capacities."

The primary outcome of this participatory approach was the successful implementation and ownership of the platform. Farmer associations embraced methods taught by CIAT to translate climate predictive data into agricultural decisions. There is suggestive evidence indicating the potential sustainability of this effort. Between 2015 and 2018, when project funding ceased and the Aclímate platform remained unupdated, farmer associations continued to employ the methodology and shared results on their respective websites. This suggests the project's potential for enduring impact beyond its initial duration and underscores the adoption of data-driven decision-making methodologies. Luis Armando Muñoz from CIAT said, "After training the associations on the use of the data analysis tools, they started evaluating their employees on the use of these new knowledge and skills as part of their performance. So, that is a change in the institutional policy that permeates even the human resources unit, because it includes in their performance evaluation the use of the tools that we were teaching them."

➔ Building trust by demonstrating clear benefits from sharing and re-using data

When CIAT approached the farmer associations about the possibility of data collaboration, farmer associations were hesitant to open databases collected over 15+ years. Nonetheless, CIAT presented the benefits of re-using the agricultural data to train a predictive model. Precisely understanding regional climate variations and the probability of rain or drought could contribute to making informed decisions. Daniel Jimenez from CIAT commented, "The associations initially shared some of their data. By leveraging this information, we demonstrated the power of data-driven insights. This significantly accelerated our collaborative efforts, leading to the achievement of even greater results."

→ Outreach challenges were overcome by using alternative tools and building on existing local networks

The project managed to overcome infrastructure challenges such as internet connectivity, digital literacy, and computer usage by adapting platform workshops based on the specific capabilities of each actor. The workshops built on existing networks in the Colombian agriculture sector, and used alternative tools to disseminate climate-predictive information.

CIAT adapted training based on the digital literacy and capabilities of the stakeholders by filling the knowledge gaps. Moreover, alternative dissemination tools were used, such as print, radio, WhatsApp groups, and meetings with local actors so technical information could reach farmers in the field. These mechanisms assisted farmers who could not use the Aclímate platform in possibly receiving climate information from other sources. Steven Sotelo from CIAT brought up,

> We developed two versions of the platform, one for technicians and one for farmers. Initially, we thought the farmer version would be basic and the technician version more advanced. However, during usability testing with farmers and technicians, both groups expressed the need for help and training to use the system. The two levels that we originally considered as options for users did not turn out to be usable according to the real perceptions and needs of farmers and technicians, so we had to design a hybrid interface between the two profiles.

The project tapped into local networks that existed among stakeholders in the Colombian agriculture sector, such as local technical agro-climatic committees. These committees, composed of farmers, national and local governments, and the community at large, gather together to analyze different meteorological sources and create recommendations through bulletins, social media, radio, and local newspapers (Loboguerrero et al., 2018). The inclusion of Aclímate as a source of information helped ensure that not only farmer associations benefitted from the platform´s information, but also the individual farmers could participate in the analysis of the data and its diffusion and implementation in the field. Steven Sotelo from CIAT commented,

> It would have been difficult for a group of scientists to generate tools without the support of the associations and other local actors, who are the ones who have direct contact with the farmers, who are the ones who open the way for us to communicate with them, to be able to have those spaces. It is not easy to reach them directly, if not through these associations and partners. So, this is a co-creation process, a joint work.

4.2. Kaffrine, Senegal: Re-use of Census Data to Establish a Climate Warning System

With regard to the Kaffrine, Senegal Use Case, there was no opportunity for those involved in the project to be interviewed. Therefore the supporting quotes and information below have been sourced from "Providing timely and relevant seasonal forecasts to farmers in a region where misinformation could mean the difference between a good year and a catastrophic one is not a job to attempt alone." [6] In this blog, two key Senegalese experts who served as guides and beacons for the successful inter-institutional collaboration: Ousmane Ndiaye, director of seasonal forecast production at ANACIM and El Hadji Moussa Seck, agricultural extension program director for Kaffrine, Senegal, were interviewed.

[6] https://ccafs.cgiar.org/news/bridging-climate-forecasts-farmer-realities-story-seck-and-ousmane

Findings:

➔ **Establishing local networks to build trust, legitimacy, and sustainability**

The main challenge during the project was the initial reluctance and opposition from rural farmers to use climate information from weather forecasting. A key aspect in overcoming this barrier was to heavily involve local leaders, using local networks in the community to help bring together scientists and farmers. In this case, Ousmane Ndiaye and El Hadji Moussa Seck worked together to exchange relevant information, developing a project that meets the needs and expectations of the community. Ousmane Ndiaye from ANACIM said, "Seeing as there is no way for us to know what climate information is going to be relevant, nor how we might best translate it so that it is understandable and useful for the farmers. They need a partner like Seck to make that happen. Seck and I are complementary," and, " I have the climatic foresight, he has the reality on the ground. Our work is to join those two aspects to make climate variability less of a mystery for the farmer."

Working with the established social network of the local community provided a basis of trust and legitimacy from the community members. Ousmane Ndiaye from ANACIM said,

> Taking advantage of existing agricultural networks at the local level has been crucial for our success. These networks bring confidence and legitimacy with them, like for example with the governmental extension services and farmer associations that have been in existence for many years. Or how ANACIM has enabled equity in terms of access to climatic information, from themselves down to local extensionists and farmers.

➔ **Using a discussion participatory approach to develop a multidisciplinary solution-based strategy for farmers**

In this use case, aligning with the classification system outlined in "Taxonomy of Participatory Approach for Citizen Engagement", workshops engaging the scientists and farmers took the form of structured discussions – an "intermediate" level of engagement. This format provided a platform for the farmers to actively contribute to the decision-making process. The inclusive nature of these workshops allowed them to share valuable insights, allowing scientists to develop and refine their project according to the specific needs of the community. Most importantly, this collaborative discussion provided the ability for the scientists to introduce collected climate data and models to farmers, furthering the exchange of information. These open workshops were key for researchers in creating a multidisciplinary approach to address the variety of issues raised by the farmers in addition to using re-used data. Ousmane Ndiaye and El Hadji Moussa Seck commented, "What we've been able to do is employ an integrated and multi-disciplinary focus on the themes of climate, agriculture and food security. From the point of view of the farmers, we have been able to break down their climate information needs, their way of making decisions, and their manner of understanding the climate in a participatory way."

4.3. Pokhara, Nepal: Data Re-use to Develop an Open Street Map for Disaster Management

Findings:

➔ **Using discussion participatory approach for stakeholder training**

In this use case, the main challenge was changing citizen perception on geography and cartography data, which were often viewed as complex topics that only experts can understand and utilize. Given that the project's objective was to raise awareness among locals, government officials, and NGOs on the significance of open-source data in disaster management, active engagement in the collection, utilization, and management of the data was essential. To effectively achieve this objective, the project adopted a proactive approach by involving citizens in locally-led workshops and training sessions (*Discussion Participatory Approach*) in different aspects of the data collection, building skills in being

effective first responders. Shashish Maharjan from Kathmandu Living Labs said, "We organized a lot of trainings like open street maps sensitization events, local mapping parties, and local workshops like using UAVs for collecting data, 3D mapping and storing those data in different platforms like Geonode. Different workshops were conducted and more than 600 people were trained on different kinds of events."

➔ Building stakeholder capacity for data collection, storage, and management

This project focused on involving many different stakeholders, notably local university students, NGO members, and local government officials – each in a different role regarding the collection, utilization, and management of the data. For example, students were trained in data collection as they might be more familiar with GPS and mobile phones. During the process, these students needed to interact with other community members to collect the data. The collaboration with local government officials validated data that was collected by volunteers and students, such as the locations of the government offices, hospitals, and police stations. This data was later verified in collaboration with local government officials. Therefore, there were different levels of work, and there was collaboration with different levels of communities. Shashish Maharjan from Kathmandu Living Labs commented,

> We first reached out to a lot of local government officials. After agreeing with them, we first trained them on the use of OpenStreetMap. Then, we went to different universities and we organized a different training there as well. Also, we didn't just focus on OpenStreetMap. We were training them using drones and using other mapping tools. And we were using different kinds of field data collection tools.

Employing this approach to involve local community members has not only enhanced the quality of collected data but also has the potential to safeguard the project's sustainability beyond its initial phase. While the project's stakeholders, such as institutions and organizations, may undergo changes over time, the ongoing utilization and periodic updating of the platform by trained locals remain consistent (Observatory for Public Innovation, 2019).

➔ Establishing user friendly interface for project implementation

One of the main themes throughout this project was to make data collection, usage, and management systems available and understandable to the public members who would be involved in disaster management. It was found that accessing the open data was difficult for the majority of users; therefore, it was key to create a user-friendly interface that could be more easily accessed. Shashish Maharjan from Kathmandu Living Labs said, "Even non-tech people could go to the portal and just explore their locality and all sorts of things. Later on, in collaboration with the local government, we also published different mapbooks. So even if you don't have access to the internet, you can use those map books not just to visualize, but to plan and to do certain analysis later on."

➔ Ensuring accountability for building trust

At the project's outset, there was a notable lack of public trust regarding the sharing of data. Moreover, concerns arose regarding potential unrestricted access to the platform, leading to fears of potential errors. To address these apprehensions, the project implemented robust quality checks aimed at mitigating such issues, thereby enhancing security measures. Consequently, these measures fostered a more secure and trustworthy environment, thereby promoting accountability and cultivating trust. Shashish Maharjan from Kathmandu Living Labs said, "We convinced the people to open their data, assuring that there would always be people who are checking the quality of the data. Even if someone makes a mistake and uploads incorrect data, there will always be someone else who can check this and correct the mistake. People were eventually reassured that there would be many people checking the data to rectify mistakes."

5. Conclusions

Datafication – while bringing technological advancements and opportunities – has highlighted significant risks of asymmetries in data access and agency, especially in developing countries. The emergence of these asymmetries limits the potential re-use of data for the common good and increases disparities in access and knowledge of re-using data for informed decision-making. The introduction of a responsible framework for data re-use, such as a Social License, can help mitigate some of the identified agency asymmetries, emphasizing a need for a shift from individual consent to collective engagement.

The implementation of citizen engagement methods in data re-use projects is the first step to moving towards a responsible and legitimate data collaborative scenario. Citizen engagement in data re-use projects aims to provide a platform for negotiations around conditions and expectations on the use and re-use of data. As a result, this method ensures a level of trust and accountability in the projects. Case studies from Colombia, Senegal, and Nepal serve as compelling examples of successful data re-use project implementation, focusing on incorporating innovative public engagement. These cases highlight the vital importance of understanding the local social, cultural, political, and economic factors needed to adapt and tailor engagement strategies – i.e. recognizing the data needs of farmers to improve crop conditions, gauging the level of data literacy of involved stakeholders, and employing effectives methods of information dissemination.

The findings highlight the significance of transparent communication among the stakeholders regarding the advantages of data sharing and effective collaboration in creating synergies between the private and public sectors. In the Colombian case, this is exemplified in the exchange between farmer associations, CIAT, and the Ministry of Agriculture in sharing their database information. Furthermore, the use of advanced engagement methods such as Discussion and Co-decision, coupled with leveraging local networks, increased legitimacy and trust. These approaches in a data re-use project not only assist in its adoption, but also can provide a platform for long-term implementation. While there is some progress in developing countries in terms of data re-use, it is necessary to implement and establish a Social License to ensure that important stakeholders and actors are involved in the conversations surrounding data re-use implementation.

As for the future vision of the Social License, it will be vital to further investigate how to structure the agreement process between stakeholders on the expectations and conditions of data re-use. Currently, the agreement process can either be an official document or an informal agreement that details this information. However, as the implementation of the Social License becomes more utilized and exercised, it will be necessary to develop a template that will standardize the process of developing the agreement between the stakeholders and actors involved. This template should build on the participatory mechanisms previously outlined.

Recognizing the potential of the Social License in reshaping the methods on how data re-use projects and initiatives in developing countries are implemented, development banks and other actors working on advancing data re-use for societal benefit in developing countries, should prioritize integrating and prototyping this approach into their data investments. By prioritizing public engagement and key stakeholders, in the context of developing countries, the Social License aims to empower communities, allowing them to influence how data is used and protect themselves from exploitation. Furthermore, a Social License ensures data governance frameworks reflect local values and needs, as well as allow for a shift in power dynamics, through the promotion of transparency, accountability, and collective decision-making. To realize these objectives effectively, development banks are urged to form partnerships with local communities, establish capacity-building approaches, and equip local stakeholders with the skills and knowledge necessary for meaningful participation in shaping how data will be re-used.

6. Recommendations

Responsible data re-use has become an increasingly popular practice that looks into leveraging existing data for another purpose. It allows for improved decision making, greater innovation, and provides cross-sectoral or cross-organizational insight at a reduced cost. In Europe, the Social License framework has emerged as a suggested method in advancing policies for data re-use, i.e., repurposing private sector data for public interest purposes, focusing on public engagement. Recognizing the significance of the Social License for data re-use, particularly in the European context, it is vital for stakeholders to consider and apply best practices for responsible engagement. These recommendations guide policymakers, development agencies, and other stakeholders in effectively implementing citizen engagement when re-using data for development purposes in developing countries. This approach aims to address the challenges of data asymmetries and foster a more equitable data ecosystem.

- **Encompass Context-Specific Approaches:** Identify the distinct social, cultural, political, and economic contexts of each project and gain an understanding of the current local infrastructure and social networks. Tailor citizen engagement strategies by leveraging the insights derived from the various existing contexts, structures, and networks.
  - Invest in further research to identify the key contextual variables that impact meaningful engagement,
  - Develop and fund workshop to identify best practices on context-sensitive approaches to development and how lessons learned can be applied to data re-use engagements,
  - Train data stewards to leverage data to inform data re-use engagements and develop data driven contextual approaches,
  - Ensure that representation is at the forefront of the citizen engagement strategy to ensure that decisions made represent shared community values.

- **Implement Discussion and/or Co-decision Engagement Participation:** Encourage the active involvement of relevant, local stakeholders in either the discussions, design and/or implementation of data collection or re-use process. This participation should include a method identified in the Taxonomy of Participatory Approaches for Public engagement such as: Deliberative Discussions, Panel Discussions, Citizen Assemblies, etc.
  - Evaluate the stakeholders' data literacy and access to technological tools to know what engagement to implement, i.e. a more complex -Citizen Assemblies- or simpler method -survey-,
  - Contact the stakeholders that should be involved and present them the incentives and benefits of engaging in the discussion,
  - Provide transparent information about who will benefit and who will make money from the project,
  - Organize the meetings and decide the method: -online or in-person, and the place in which they would take place: -cities or rural areas.

- **Incorporate Alternative Communication Channels:** To effectively share the outcomes of the project, it is essential to employ a diverse range of communication channels, recognizing that different stakeholders operate within distinct contexts.
  - Identify areas with limited technological infrastructures, access to internet or digital literacy,
  - Implement alternative methods such as radio, local meetings, and printed material to disseminate information, gather input, and receive feedback.

- **Build and Maintain Trust:** Building trust is important for the sustainability of the project. Therefore it is necessary to implement different strategies that involve ideally all key stakeholders at every stage of the project.
  - Build trust through incorporating valued local leaders (religious leaders, mayors, et.c), listening to the concerns of the community and by clearly communicating the benefits of data sharing and re-use,
  - Maintain trust by continuing transparent communication and actions, as well as providing mechanisms for feedback.

- **Strengthen Local Capacity and Infrastructure:** To ensure the sustainability of the project beyond the initial investment, it is important to transfer the knowledge to the community so they can continue to implement the project and benefit from its outcomes.
  - Invest in improving the capabilities and capacities of local community members, involving relevant members (i.e. university students, first responders, etc.) in the data collection, re-use, and implementation of the project,
  - Improve local data literacy, technical skills, and/or data organizational skills.

- **Create Collaboration and Partnership:** Foster cross-sector partnerships to leverage local expertise, share best practices, and ensure sustainable project implementation. Collaboration and partnerships should emphasize the inclusion of local leaders as well as utilize existing identified local structures and networks.
  - Conduct an assessment of existing local structures and networks relevant to the project goals,
  - Design collaboration initiatives that focus on including all perspectives and backgrounds of the stakeholders and community,
  - Reinforce continuous open communication channels between collaborators and partners to ensure transparency and trust.

- **Focus on Developing Deliverables that Emerge from Public Engagement:** Too often, public engagement stops without a concrete deliverable (such as framework or licensing conditions) that specify the preferences and expectations of each of the stakeholders, revolving around how data will be collected, used, and/or re-used.
  - Collaborate with stakeholders in the development of concrete deliverables to ensure their perspectives and expectations are included,
  - Ensure that the conditions detailed in the deliverables are practical, actionable and transparent, allowing for a clear vision on how to proceed with the re-use of data.

- **Consider Artificial Intelligence (AI):** With the emergence of AI tools, consider how public engagement can be used to establish a Social License for AI - including how training data is used for LLMs.
  - Design workshops and informational sessions to inform the public about AI, its applications, potential impacts, and relationship to data re-use,
  - Develop educational materials to enhance understanding and combat misconceptions about AI,
  - Encourage open dialogue that allow for stakeholders to express their concerns, expectations, and recommendations regarding AI.

Appendices

Appendix A : Country-specific Contextual Indicators

Table 1: Contextual Indicators for Colombia

COLOMBIA CONTEXTUAL INDICATORS					
Demographic Indicators		**Social-Economic Indicators**		**Socio-Political Indicators**	
Age Distribution (%) 0-14 years 15-65 years >65 years	(2018) 22.6 68.2 9.1	**Inequality Level** Gini coefficient (Scale 0-total equality-100-total inequality)	(2021) 51.5	**Corruption Levels** Perception of the level of corruption in the public sector (Scale 0-highly corrupted-100-no corruption)	(2022) 39
Gender Distribution (%) Women Men	(2018) 51.2 48.8	**Data Literacy (%)** Access to internet connection Use of computers every day	69.79 (2020) 44.9 (2018)	**Political Participation** Political participation (0-10) Abstentionism in presidential election (%)	6.8 (2022) 47.0 (2018)
Rural-Urban Distribution (%) Cities (municipal seats) Rural and populated centers	(2018) 77.1 22.9			**Trust in Institutions** (Scale 0-100) Trust in the President Trust in the government Trust in the Congress	(2020) 39 25 14
Education Level (%) Highschool University degree	(2012) 42.0 20.0				

Table 2: Contextual Indicators for Senegal

SENEGAL CONTEXTUAL INDICATORS					
Demographic Indicators		**Social-Economic Indicators**		**Socio-Political Indicators**	
Age Distribution (%) 0-14 years 15-65 years >65 years	(2021) 41,76 55.08 3.16	**Inequality Level** Gini coefficient (Scale 0-total equality-100-total inequality)	(2019) 52.2	**Corruption Levels** Perception of the level of corruption in the public sector (Scale 0-highly corrupted-100-no corruption)	(2022) 43
Gender Distribution (%) Women Men	(2022) 50.8 49.0	**Data Literacy (%)** Access to internet connection Use of computers every day	46.2 (2020) N/A	**Political Participation** Political participation (0-10) Abstentionism in presidential election (%)	7.3 (2022) 66.27 (2019)
Rural-Urban Distribution (%) Cities (municipal seats) Rural and populated centers	(2022) 49.0 51.0			**Trust in Institutions** (Scale 0-100) Trust in the President Trust in the government Trust in the Congress	(2014) 66 59 50
Education Level (%) Highschool University degree	(2021) 49.0 16.0				

Table 3: Contextual Indicators for Nepal

NEPAL CONTEXTUAL INDICATORS					
Demographic Indicators		**Social-Economic Indicators**		**Socio-Political Indicators**	
Age Distribution (%) 0-14 years 15-65 years >65 years	(2022) 29.0 65.0 6.0	**Inequality Level** Gini coefficient (Scale 0-total equality-100-total inequality)	(2019) 41.5	**Corruption Levels** Perception of the level of corruption in the public sector (Scale 0-highly corrupted-100-no corruption)	(2022) 34
Gender Distribution (%) Women Men	(2022) 52.1 28	**Data Literacy (%)** Access to internet connection Use of computers every day	91 (2021) N/A	**Political Participation** Political participation (0-10) Abstentionism in presidential election (%)	(2022) 7 N/A
Rural-Urban Distribution (%) Cities (municipal seats) Rural and populated centers	(2022) 21.0 79.0			**Trust in Institutions** (Scale 0-100) Trust in the President Trust in the government Trust in the Congress	2020 N/A 67 67
Education Level (%) Highschool University degree	(2022) 86.0 17.0				

Appendix B: Methodology

The stakeholder consultation was undertaken by conducting semi-structured interviews with stakeholders from two key groups: 1) data experts, and 2) use case stakeholders. A total of 12 participants were interviewed:

Name	Domain	Organization	Date of Interview
Igbal Safarov	Data Expert in open data and data governance.	Coty	29/09/23
Sille Sepp	Data Expert in human-centered, personal data-based services and processes.	MyData Global	3/10/23
Tiago Peixoto	Data Expert in public sector performance, digital government, participatory democracy, and civic tech.	The World Bank	3/10/23
Shashish Maharjan	Key Stakeholder from the Nepal Use Case.	Kathmandu Living Labs	4/10/23
Ronda Železný-Green	Data Expert in digital capacity building, with a focus on equity and helping marginalized groups.	Panoply Digital	5/10/23
Isabel Cardenas	Key Stakeholder from the Colombia Use Case.	Ministry of Agriculture and Rural Development Colombia	11/10/23
Luis Armando Muñoz	Key Stakeholder from the Colombia Use Case.	CIAT	30/10/23
Steven Sotelo	Key Stakeholder from the Colombia Use Case.	CIAT	30/10/23
Daniel Jimenez	Key Stakeholder from the Colombia Use Case.	CIAT	30/10/23
Christopher Wilson	Data Expert in civic participation, multi-stakeholder governance and digital transformation in government institutions.	MyData Global	2/11/23
Natalia Carfi	Data Expert in new technologies, open data and governance.	Open Data Charter	28/11/23
Anonymous	Data Expert in open data for governments.	-	29/11/23

Data experts were interviewed in order to analyze the broader context and current state of open data and data re-use projects, particularly as related to low and middle income countries, and to gain their insights on the benefits and risks of such projects globally. The use case stakeholder interviews aimed to gather a more detailed understanding of the project implementation and outcomes, and to gain further insights into the citizen engagement processes that were used.

The selected data experts included executive directors and senior researchers of think tanks and organizations in the data sphere, as well as professionals with a data background working in recognised public and international organizations. Stakeholders from all three use cases were contacted for non-compensatory interviews, and key stakeholders from projects in Nepal and Colombia were successfully reached for interviews. The interviews with stakeholders from Colombia were conducted in Spanish, and then translated into English. Stakeholders from the use case in Senegal were unable to be interviewed.

All interviews were conducted online and audio recorded. TDT produced a transcription for each interview based on the recording. To identify the key themes and findings of the stakeholder consultation, a thematic analysis was conducted for each interview.

Appendix C: Interview Question Guide

Introduction:

- Explanation of The Data Tank and of the joint project with AFD,
- Confirmation of verbal consent to be audio recorded during the interview,
- Confirmation of whether participants are happy to be quoted in the final report, after being sent a copy of the quote for review,
- Opportunity for participants to ask any further questions before the interview.

Interview questions for data experts:

1. What do you believe is the value proposition of using and re-using data? Can you define any advantages or benefits?
2. What challenges and potential risks do you see associated with the use and re-use of open data?
   a. Follow up question: What kinds of risks and concerns do you believe the public has regarding open data and data re-use? Ask for an example.
3. Why does open data and data re-use matter for developing countries? Examples?
4. Do you believe that current consent models address privacy and security issues in the context of establishing open data and data re-use?
   a. Do you think these models accommodate both individual and community needs?
5. In your research, have you come across any examples of engaging citizens in open data initiatives/policies? What do you think is the state of public engagement in the data sphere? Do you think it is growing-scaling? Examples?

Interview questions for use case stakeholders:

<u>Questions for the use case from Nepal</u>

1. Could you tell us about the Pokhara project - what were the aims and objectives of this project?
2. We see that a major component of the project was engaging local people, community, and various government and non-government organizations in mapping activities, helping them understand various mapping tools and disaster management concepts. Why was this approach essential for the project?
3. What was the process of engaging the stakeholders? At what stage of the project did you engage the public?
4. What methods did you use to reach out to and to engage the different stakeholders/the public? Especially citizens and stakeholders that were not technicians and academics?
5. What were some challenges that you encountered (i.e.,in developing the method of engagement or in getting citizens involved)?
6. How did engaging different stakeholders benefit your project?
7. What was the impact of the project? Do you still see the benefits today? What is happening with the project today?

Questions for the use case from Colombia

1. Could you tell us about the Aclímate project? What were the goals and objectives of this project? What was CIAT's role, compared to that of the Ministry of Agriculture?
2. We see that one of the main components of the project was to involve different actors for data capture. How was the process of integrating these actors, from associations such as Fedearroz, the farmers and the more technical people from CIAT, IDEAM and the Ministry? In what phases of the project were each actor and the farmers involved?
3. How was the process of sharing the data, of building trust to share this data?
4. What methods were used to socialize the results of the model or the project as such? Especially citizens and stakeholders who were not technical or academic?
5. What were the main challenges faced by the project?
6. How has stakeholder participation benefited the project?
7. What was the impact of the project? Are you still receiving benefits today? What is happening with the project today?

Appendix D: Social License Public Engagement Toolkit (L-KIT)

The below toolkit is designed to guide organizations that want to re-use data in effectively establishing a Social License through engaging with communities and stakeholders, building upon lessons learned from cases like Nepal and Senegal. It emphasizes the need for innovative tools of engagement, building on existing local infrastructure and networks, and the importance of sharing information.

This toolkit is designed to be flexible and adaptable to various contexts and projects. Its ultimate goal is to foster a collaborative environment where data projects can thrive and benefit all stakeholders, particularly in developing countries. By following these steps and learning from successful case studies, organizations can establish a strong Social License through effective public engagement.

1. Defining the Purpose

- **Objective Identification:** Clearly define the goals and expected outcomes, including deliverables such as sharing framework or agreement, of the public engagement project,
- **Value Proposition:** Establish why the data being re-used in the project is beneficial for the community and the stakeholders involved.

2. Identifying the Stakeholders

- **Stakeholder Mapping:** Identify key stakeholders, including local leaders, community members, and relevant organizations.
  - **Establish who are the data receivers:** the stakeholders who need the data,
  - **Establish who are the data holders:** the stakeholders who own/store the data,
  - **Establish existing local infrastructure and networks:** the local intermediaries who can facilitate the collaboration between stakeholders and project implementers (i.e. local leaders, religious leaders, mayors, etc.).
- **Needs Assessment:** Understand the needs, concerns, and expectations of each stakeholder group.

3. Selecting the Type of Public Engagement

- **Engagement Identification:** Determine the level of engagement needed, ideally from Discussion and/or Co-decision[7],
- **Method Selection:** Choose appropriate engagement methods, such as round tables, deliberative discussions, committees, advisory groups, etc.

Criteria for Selecting Engagement Methods

Stakeholder Characteristics

- Demographics: Consider age, education level, and socio-economic background of the stakeholders,
- Digital Literacy: Gauge the community's comfort and familiarity with digital tools,
- Language and Cultural Nuances: Choose methods that align with local languages and cultural practices.

Local Infrastructure

- Technology Access: Evaluate the availability and reliability of technology (i.e., internet access, mobile network coverage).

Local Networks and Community Preferences

- Community Engagements: Identify important networks and leaders in the community that would facilitate discussion and project design between the stakeholders,
- Preferred Communication Channels: Identify how the community prefers to receive and share information (i.e., community meetings, radio, social media).

Project Scope and Complexity

- Technicality of the Project: For complex projects, more interactive and educational methods like workshops may be needed,
- Scale and Reach: For larger projects, scalable methods like digital platforms or mass communication tools might be more effective.

Resource Availability

- Personnel and Expertise: Assess the availability of skilled personnel to facilitate engagement methods.

Risk and Cultural Assessment

- Data Privacy Concerns: Choose methods that ensure data security and privacy, tha meet the stakeholders' demands/preferences,
- Potential for Conflict: Be aware of any social or political dynamics that could impact the engagement process.

Flexibility and Adaptability

- Changing Circumstances: Be prepared to modify engagement methods in response to unforeseen changes or challenges,
- Feedback Loop Integration: Ensure that there is room for adjusting methods based on stakeholder feedback.

[7] It is possible to implement the other methods identified in The Taxonomy for Public Engagement, however it is recommended to implement Discussion and/or Co-decision as active participation approaches.

4. Innovative Tools of Engagement

   - **Localized Tools:** Adapt engagement tools to local contexts, like using radio in areas with limited internet access,
   - **Multi-channel Approach**: Use a mix of digital and non-digital tools to reach a wider audience.

Lessons from our Case Studies

- Nepal's Approach: Highlight the use of paper maps and workshops in areas with limited digital access, emphasizing the importance of adapting to local technological landscapes,
- Senegal's Strategy: Demonstrate how leveraging existing agricultural networks and local leaders can enhance project legitimacy and trust,
- Colombia's Information Sharing: Use the Colombian case to illustrate effective data sharing and collaboration between stakeholders.

5. Building on Local Infrastructure and Networks

   - **Leveraging Existing Networks:** Collaborate with existing local networks, such as farmer associations or community groups,
   - **Community Integration:** Ensure the project is integrated into the local community infrastructure for sustainability.

6. Ensuring Information Sharing

   - **Transparency:** Maintain open and transparent communication throughout the project lifecycle,
   - **Feedback Mechanisms:** Implement systems for continuous feedback from stakeholders.

7. Engagement Strategy Development and Planning

   - **Plan Creation:** Develop a comprehensive engagement plan integrating previous steps. This includes outlining specific activities, setting timelines, and assigning responsibilities. Utilize the criteria for selecting engagement methods to ensure the chosen methods align with stakeholder characteristics, project scope, resource availability, local infrastructure, and community preferences.

8. Implementation

   - **Implementation:** Carry-out the engagement activities as outlined in the plan,
   - **Deliverables:** Create tangible deliverables from the engagement activities such as data sharing agreements or data responsibility frameworks
   - **Adaptability:** Remain flexible and responsive to feedback and changing circumstances. Be prepared to modify engagement methods if they are not meeting the intended goals or if stakeholder feedback indicates a need for change.

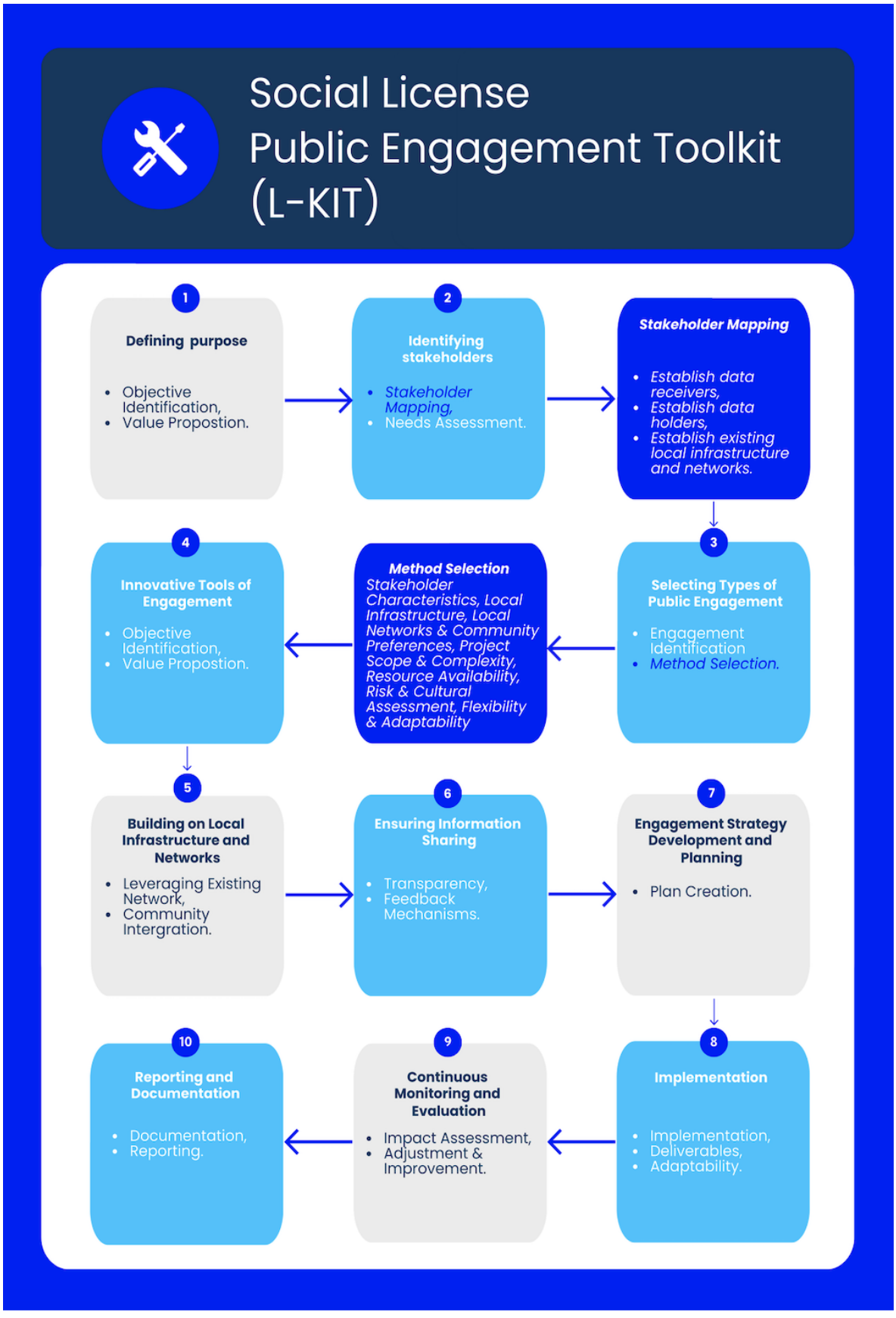
9. Continuous Monitoring and Evaluation

   - **Impact Assessment:** Regularly monitor and evaluate the effectiveness of the engagement activities. This should involve collecting feedback, measuring participation levels, and assessing stakeholder satisfaction,

- **Adjustment and Improvement:** Use the insights gained from ongoing monitoring and evaluation to make necessary adjustments to the engagement strategy. This step is crucial for ensuring the engagement process remains relevant, effective, and aligned with the stakeholders' evolving needs and contexts.

10. Reporting and Documentation

    - **Documentation:** Keep detailed records of all engagement activities and stakeholder feedback,
    - **Reporting:** Regularly report progress to stakeholders and the community.


Social License
Public Engagement Toolkit
(L-KIT)
1
Defining purpose
Objective Identification,
Value Propostion.
2
Identifying stakeholders
Stakeholder Mapping,
Needs Assessment.
Stakeholder Mapping
Establish data receivers,
Establish data holders,
Establish existing local infrastructure and networks.
3
Selecting Types of Public Engagement
Engagement Identification
Method Selection.
Method Selection
Stakeholder Characteristics, Local Infrastructure, Local Networks & Community Preferences, Project Scope & Complexity, Resource Availability, Risk & Cultural Assessment, Flexibility & Adaptability
4
Innovative Tools of Engagement
Objective Identification,
Value Propostion.
5
Building on Local Infrastructure and Networks
Leveraging Existing Network,
Community Intergration.
6
Ensuring Information Sharing
Transparency,
Feedback Mechanisms.
7
Engagement Strategy Development and Planning
Plan Creation.
8
Implementation
Implementation,
Deliverables,
Adaptability.
9
Continuous Monitoring and Evaluation
Impact Assessment,
Adjustment & Improvement.
10
Reporting and Documentation
Documentation,
Reporting.

References

Aitken, M., Toreini, E., Carmichael, P., Coopamootoo, K., Elliott, K., & 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*, 7(1). https://doi.org/10.1177/2053951720908892

Allen, K., & Wilson, J. (2008), The legitimacy of corporate social responsibility. Corporate Governance, 8(2), 36-49.

ANDI (2022), *Habilidades Digitales en Colombia ¿El futuro digital es de todos?* https://www.andi.com.co/Uploads/GAN_HabilidadesDigitales_COL_V8.pdf

Ansell, C., & Gash, A. (2008), Collaborative Governance in Theory and Practice. *Journal of Public Administration Research and Theory*, 8, 543-571. http://dx.doi.org/10.1093/jopart/mum032

Belenguer, L. (2021), *Citizens' Assemblies for a more democratically engaged and informed society on AI & Data Privacy*. Medium. https://medium.com/escapadasuk/citizens-assemblies-for-a-more-democratically-engaged-and-informed-society-on-ai-data-privacy-4d22e1c5b585

Boersma, Martijn. (2020), *Paradox of the Social License to operate.* University of Technology Sydney. https://www.uts.edu.au/about/uts-business-school/management-department/news/paradox-social-license-operate

Boscarino, N., Cartwright, R. A., Fox, K., & Tsosie, K. S. (2022), Federated learning and Indigenous genomic data sovereignty. Nature Machine Intelligence, 4(11), 909–911. https://doi.org/10.1038/s42256-022-00551-y

Coleman, M., et al. (2022), Building a social licence for data-driven innovation in health: A framework for co-creation and trust. International Journal of Medical Informatics, 155, 102400.

CCAFS, & CGIAR (2015, August 6), *Big Data for climate-smart agriculture*. Ccafs.cgiar.org. https://ccafs.cgiar.org/research/projects/big-data-climate-smart-agriculture#:~:text=The%20CSMS%20allows%20farmers%20to,both%20crowdsourcing%20and%20secondary%20databases

CCAFS, & CGIAR (2016, March 3), *The impact of climate information services in Senegal*. Ccafs.cgiar.org. https://ccafs.cgiar.org/outcomes/impact-climate-information-services-senegal

Choo, M., & Findlay, M. (2021), Data re-use and its impacts on digital labour platforms. *SSRN Electronic Journal.* https://doi.org/10.2139/ssrn.3957004

Choo, J. (2022), *The use and abuse of personal data by the PAP government*. New Naratif. https://newnaratif.com/the-use-and-abuse-of-personal-data-by-the-pap-government/

Cinnamon, J. (2019), Data inequalities and why they matter for development. *Information Technology for Development*, 26(2), 1–20. https://doi.org/10.1080/02681102.2019.1650244

Data & Trust Alliance (2020) The Data Trust Playbook: Building Trustworthy Data Ecosystems.

DANE. (2018), *DANE - ¿Cuántos Somos?*. Departamento Administrativo Nacional de Estadística DANE. https://www.dane.gov.co/index.php/estadisticas-por-tema/demografia-y-poblacion/censo-nacional-de-poblacion-y-vivenda-2018/cuantos-somos

Degeling, D., et al. (2020), Ethical and legal challenges in the use of digital health data in low- and middle-income

Dodds, L. (2017, March 24), What is data asymmetry? *Lost Boy*. https://blog.ldodds.com/2017/03/24/what-is-data-asymmetry/#:~:text=The%20term%20data%20asymmetry%20refers

Duarte, F. (2023, March 16), *Amount of data created daily (2023)*. Exploding Topics. https://explodingtopics.com/blog/data-generated-per-day

European Commission (n.d.), *What is open data.* https://data.europa.eu/en/dataeuropa-academy/what-open-data

European Comission (2024), *Data Act | Shaping Europe's digital future.* Digital-Strategy.ec.europa.eu. https://digital-strategy.ec.europa.eu/en/policies/data-act#:~:text=The%20Data%20Act%20is%20a

European Commission (2018, November 22), *Commission appoints Expert Group on Business-to-Government Data Sharing | Shaping Europe's digital future.* Digital-Strategy.ec.europa.eu. https://digital-strategy.ec.europa.eu/en/news/commission-appoints-expert-group-business-government-data-sharing

European Commission (2023), *Data governance act | Shaping Europe's digital future*. Digital-Strategy.ec.europa.eu. https://digital-strategy.ec.europa.eu/en/policies/data-governance-act

European Commission (2020), *Towards a European strategy on business-to-government data sharing for the public interest Final report prepared by the High-Level Expert Group on Business-to-Government Data Sharing.*

Eurostat. (2022), *Empowering society by reusing privately held data for official statistics — A European approach.*

Eurostat (2023), *Revision of the legal framework for European statistics - Eurostat*. Ec.europa.eu. https://ec.europa.eu/eurostat/web/main/about/opportunities/consultations/revision-of-the-legal-framework-for-european-statistics

Favaretto, M., De Clercq, E., & Elger, B. S. (2019), Big Data and discrimination: Perils, promises and solutions. *A systematic review. Journal of Big Data, 6*(1), 12. https://doi.org/10.1186/s40537-019-0177-4

Ford, E., Rees-Roberts, M., Stanley, K., Goddard, K., Giles, S., Armes, J., Ikhile, D., Madzvamuse, A., Spencer-Hughes, V., George, A., Farmer, C. and Cassell, J. (2023), Surrey and Sussex: results of deliberative focus groups. *International Journal of Population Data Science*, 5(3). https://ijpds.org/article/view/2114

Guay, J. (2017), *Phone records help Namibia clamp down on malaria.* Apolitical. https://apolitical.co/solution-articles/en/phone-records-help-namibia-clamp-malaria

Han, H. J. (2022), "How Dare They Peep into My Private Life?" *Human Rights Watch.* https://www.hrw.org/report/2022/05/25/how-dare-they-peep-my-private-life/childrens-rights-violations-governments

Lindegren, Daniel, et al. "An Evaluation of Three Designs to Engage Users When Providing Their Consent on Smartphones." *Behaviour & Information Technology*, vol. 40, no. 4, 17 Dec. 2019, pp. 398–414, https://doi.org/10.1080/0144929x.2019.1697898. Accessed 5 Jan. 2024.

McConaughey, E., Rigaudeau, B., & Duggal, K. (2023), *Social License to Operate*. Jus Mundi. https://jusmundi.com/en/document/publication/en-social-license-to-operate#:~:text=The%20term%20Social%20License%20to,by%20the%20stakeholders%20with%20the

Kenton, W. (2022, December 7), *Social License to operate (SLO): Definition and standards*. Investopedia. https://www.investopedia.com/terms/s/social-license-slo.asp

Latinobarómetro (2020), *Latinobarómetro*. https://www.latinobarometro.org/lat.jsp?Idioma=724

Loboguerrero, A. M., Boshell, F., León, G., Martinez-Baron, D., Giraldo, D., Recaman Mejía, L., Díaz, E., & Cock, J. (2018), Bridging the gap between climate science and farmers in Colombia. *Climate Risk Management*, 22, 67–81. https://doi.org/10.1016/j.crm.2018.08.001

Muller, S. H., Kalkman, S., van Thiel, G. J., Mostert, M., & van Delden, J. J. (2021), The social licence for Data-intensive health research: Towards co-creation, public value and trust. *BMC Medical Ethics*, 22(1). https://doi.org/10.1186/s12910-021-00677-5

Muyoga, C., Jimenez Cisneros, A., & Železný-Green, R. (2022), *6 steps to get started on decolonizing data for Development*. data.org. https://data.org/news/decolonizing-data-for-development/

Observatory of Public Sector Innovation (2019), *Secondary Cities (2C) Pokhara Project*. Observatory of Public Sector Innovation. https://oecd-opsi.org/innovations/secondary-cities-2c-pokhara-project/

OECD. (2022, September), OECD Guidelines for Citizen Participation Processes. https://www.oecd.org/gov/open-government/oecd-guidelines-for-citizen-participation-processes-f765caf6-en.htm

Paccoud, I., Baumann, M., Le Bihan, E., Pétré, B., Breinbauer, M., Böhme, P., Chauvel, L., & Leist, A. K. (2021), Socioeconomic and behavioural factors associated with access to and use of Personal Health Records. *BMC medical informatics and decision making*, 21(1), 18. https://doi.org/10.1186/s12911-020-01383-9

Paprica, P. A., de Melo, M. N., & Schull, M. J. (2019), Social licence and the general public's attitudes toward research based on linked administrative health data: a qualitative study. *CMAJ open*, 7(1), E40–E46. https://doi.org/10.9778/cmajo.20180099

Purwanto, A., Zuiderwijk, A., & Janssen, M. (2020), "Citizen engagement with open government data: Lessons learned from Indonesia's presidential election", *Transforming Government: People, Process and Policy*, Vol. 14 No. 1, pp. 1-30. https://doi.org/10.1108/TG-06-2019-0051

Sotelo, S., Guevara, E., Llanos-Herrera, L., Agudelo, D., Esquivel, A., Rodriguez, J., Ordoñez, L., Mesa, J., Muñoz Borja, L., Howland, F., Amariles, S., Rojas, A., Valencia, J., Segura, C., Grajales, F., Hernandez, F., Cote, F., Saavedra, E., Ruiz, F., ... Ramirez-Villegas, J. (2020), Pronosticos Aclimatecolombia: A system for the provision of information for climate risk reduction in Colombia. *Computers and Electronics in Agriculture, 174,* 105486. https://doi.org/10.1016/j.compag.2020.105486

UN. Department of Economic and Social Affairs. Division for Public Administration and Development Management (2013), *Guidelines on open government data for citizen engagement.* United Nations. https://digitallibrary.un.org/record/3907402?ln=en

Van de Sandt, S., Dallmeier-Tiessen, S., Lavasa, A., & Petras, V. (2019), The Definition of Reuse. *Data Science Journal*, 18(1), 22.DOI: https://doi.org/10.5334/dsj-2019-022

Van de Waerdt, P. J. (2020), Information asymmetries: recognizing the limits of the GDPR on the data-driven market. *Computer Law & Security Review*, 38(105436). https://doi.org/10.1016/j.clsr.2020.105436

Vandenbroele, J. (2017, March 1). *How can open data feed citizen engagement?* Citizen Lab. https://www.citizenlab.co/blog/civic-engagement/how-can-open-data-feed-citizen-engagement/

Verhulst, S., & Young, A. (2022), Identifying and addressing data asymmetries so as to enable (better) science. *Frontiers in big data,* 5, 888384. https://doi.org/10.3389/fdata.2022.888384

Verhulst, S. (2023), Operationalizing digital self-determination. *Data & Policy*, 5, e14. https://doi.org/10.1017/dap.2023.11

Verhulst, S. et al (2023), "The Urgent Need to Reimagine Data Consent"

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

Yang, Z. (2022, November 22), *China just announced a new social credit law. Here's what it means.* MIT Technology Review. https://www.technologyreview.com/2022/11/22/1063605/china-announced-a-new-social-credit-law-what-does-it-mean/

Young, A., & Verhulst, S. (2017), *Aclímate Colombia.* Open Data's Impact - The GovLab. https://odimpact.org/case-aclimate-colombia.html

Young, A., Verhulst, S., Safonova, N., Zahuranec, A. (2020), The Data Assembly: Responsible Data Re-use Framework. The GovLab. https://thedataassembly.org/files/nyc-data-assembly-report.pdf

List of Abbreviations and Acronyms

ANACIM	The National Civil Aviation and Meteorology Agency of Senegal
B2B	Business-to-Business
B2C	Business-to-Consumer
B2G	Business-to-Government
B2S	Business-to-Science
CCAFS	Climate Change, Agriculture and Food Security
CIAT	International Center for Tropical Agriculture
G2C	Government-to-Citizen
IDEAM	National Institute of Hydrology, Meteorology, and Environmental Studies
LLMS	Large Language Models
SLO	Social License to Operate
URAS	Union des Radios Associatives et Communautaires

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 Coquelicot

Legal deposit 1st quarter 2024
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