RESPONSIBLY USING AND REUSING NON-TRADITIONAL DATA SOURCES FOR SERVICE PROVISION TO CHILDREN:

A PATH TOWARDS DUE DILIGENCE

Sara Marcucci, Andrew Zahuranec, and Stefaan Verhulst

THEGOVLAB

RESPONSIBLE DATA FOR CHILDREN

unicef

TABLE OF CONTENT

Photo by charlesdeluvio on Unsplash

1. INTRODUCTION

Today's children are the first generation growing up amid the rapid datafication of various aspects of life, and humanitarian and development organizations are increasingly contributing to this increased collection and use of data. Understanding the implications of non-traditional data collection on children requires carefully considering the potential long-term impact on their trust, agency, and self-development and their unique vulnerabilities. Addressing these facts includes acknowledging:

1. **Children Are at the Forefront of Datafication:** Children are exposed to "being datafied" from–or even before–birth. Each generation sees data systems accumulate more data about them than the previous.
2. **Children's Limited Agency:** Children lack full agency to make decisions about their data participation. Indeed, they cannot legally provide consent, only assent. Even when provided with the option to opt out, they may lack the understanding to assess risks and benefits. Indeed, privacy terms and conditions are often not easily understood by adults, let alone children. There is a need for more intelligible terms.
3. **Challenges with Aggregated Data:** Aggregated, anonymized data, often considered a privacy solution, presents unique challenges when applied to children. They, as a

vulnerable group, may face disproportionate risks when data is aggregated at the group or demographic level. Aggregated data may misrepresent certain groups of children, as the aggregation may have the “dominant” (more frequent) group obscure outliers. For instance, consider a program where data is used to investigate a certain city’s rate of childhood asthma. Aggregated data might suggest that, overall, children in the city are relatively healthy with respect to asthma. However, this broad overview could mask significant disparities within the city between individual neighborhoods, especially in relation to environmental factors. Indeed, children living in areas with poor air quality, such as near industrial sites or in neighborhoods with high traffic pollution, may have a much higher incidence of asthma compared to those in cleaner, more affluent parts of the city. As a result, public health interventions may not be adequately targeted to address these disparities, as the aggregated data does not highlight the need for interventions focusing on environmental improvements in specific areas.

4. **Impact of Data Violations on Trust:** Mishandling data or data violations can result in a loss of trust, which can have an especially large impact on children who have had fewer trust-building encounters with technology. The loss of trust may have immediate as well as lifelong consequences, affecting their behavior towards technology and potentially leading to a refusal of services.
5. **Overlooking Children's Interests:** As new technologies and data volumes increase, existing obligations and protections for children may be overlooked. Children's interests are not always prioritized or adequately considered in data collection efforts, potentially leading to unintended consequences.
6. **AI and Algorithmic Bias:** The use of AI and algorithms introduces potential biases and risks, especially when applied to children. Children may have less understanding of these processes, making them more susceptible to biased decision-making. Responsible data use frameworks for children must address the role of AI and algorithms.
7. **Privacy and Children's Self-Development:** While some level of oversight is a core part of child protection work and can mitigate harms, privacy and data responsibility are crucial for children's psychosocial growth. Having the autonomy to experiment with identities without invasive data surveillance is vital for self-development. Measures such as deanonymization may not prove sufficient. Individual and group privacy empower children to engage comfortably and confidently with peers, fostering civic and political engagement.

Cover Photo by Trinity Nguyen on Unsplash

2. NON-TRADITIONAL DATA SOURCES

Non-traditional data sources encompass a wide range of information beyond the conventional data collected by government agencies and nonprofit organizations. These sources may include:

1. **Social Media:** Data from platforms like TikTok, Whatsapp, Viber, Telegram, Twitter, and Instagram can offer insights into the daily lives, preferences, and needs of children and/or their legal guardians;

2. **Mobile Data:** Mobile phone usage patterns and location data can provide information about the movements and habits of children and/or their legal guardians;

3. **Economic and Financial Data:** Transaction records and financial activities can help in assessing the economic well-being and financial inclusion of children and/or their legal guardians. While a young child is unlikely to have a bank card, understanding the financial activity of their adult caregiver linked to that child could reveal valuable insights;

4. **Geospatial Data:** Geographic information systems (GIS) data can aid in understanding the locations of children and/or their legal guardians, and identifying areas in need of services; and

5. **Crowdsourced Data**: Crowdsourced data harnesses the collective intelligence of a community to gather information that might be challenging for a single entity to collect comprehensively. For instance, Waze is a navigation app that uses crowdsourced data to provide real-time traffic updates. Users report traffic jams, accidents, and road conditions, which the app then shares with others to suggest the best routes. This collective intelligence approach allows for highly accurate and up-to-date traffic information.

When using these sources, the challenge lies in determining the feasibility of accessing this type of information while fully respecting individuals' and groups' rights and privacy. It involves exploring whether such data can be gathered and utilized ethically, and if so, identifying the most appropriate methods to do so. To this aim, this document hopes to map out a series of considerations to take into account when deciding whether or not to use and reuse non-traditional data for service provision.

3. VALUE PROPOSITION OF NON-TRADITIONAL DATA REUSE IN THE CONTEXT OF SERVICE PROVISION TO CHILDREN

1. **Validation of Traditional Data-Based Insights:** Non-traditional data sources can serve as a valuable means to validate insights derived from traditional data sets, enhancing the reliability and accuracy of information.

    a. *Example:* The Joint Data Center is in the planning phases of a project in collaboration with the UNDP in Moldova, which plans to rely on survey data on the socioeconomic conditions of host communities and refugees in Moldova. This survey would investigate economic questions, such as access to the labor market, mobility of refugees, and similar issues. Geolocation data from the telecommunications company Orange, accessible through a data-sharing agreement, may be used to validate findings from surveys on patterns of movement, alongside border crossing datasets, program beneficiary datasets, and other forms of monitoring. However, project leaders stressed that they are still assessing whether the data from Orange would provide meaningful value and had not yet come to a conclusion on its overall value relative to other, more traditional assets.

2. **Improved Identification of Needs:** Non-traditional data sources may enable more precise identification of the needs of target populations, making it easier to provide services where they are most required.

    a. *Example:* For instance, a project initiated in 2020 by Facebook in collaboration with Columbia University's Center for International Earth Science Information Network (CIESIN), was aimed at enhancing the global response to COVID-19 through the creation of detailed population density maps. Utilizing a sophisticated blend of machine learning, high-resolution satellite imagery, and comprehensive census data, the initiative mapped the distribution of populations across nearly every country and territory worldwide accurately. Ultimately, the maps aimed to identify the most effective ways to distribute COVID-19 vaccines, especially in areas where accurate population data is scarce or outdated.

3. **Timely Assistance:** Real-time data can facilitate swift response and assistance, especially in emergency situations or crises.

    a. *Example:* InForm is an internally hosted data collection and management platform for UNICEF, based on Open Data Kit (ODK). The platform aims to establish a common approach to primary data collection, processing, and

handling through mapping and visualization of previously fragmented data. In March 2019, Cyclone Idai struck Mozambique and Zimbabwe, causing widespread devastation and leaving 2.2 million people in need of urgent assistance in Mozambique alone. In response, UNICEF collaborated with the government and other organizations to establish a Cyclone Idai Rapid Response Platform using the InForm data-driven tools. The platform, headquartered in Beira City, compiled existing data on infrastructure, population, and road conditions to aid disaster response efforts.

4. **Improved Resource Allocation:** Data analytics can help optimize resource allocation, ensuring that limited resources are utilized efficiently.

    a. *Example:* The Migration Yorkshire website provides a user guide with a map illustrating a series of information about children on local authority, regional, and world maps. Efforts like these may help identify geographical areas with high concentrations of children, helping organizations to allocate resources more effectively.

4. CRITERIA TO DETERMINE CONTEXTUAL RELEVANCE AND APPROPRIATENESS OF NON-TRADITIONAL DATA REUSE

Determining when non-traditional data is appropriate and can be leveraged responsibly involves carefully considering various factors. Criteria to help assess the appropriateness of non-traditional data can include:

1. **Relevance to the Problem:** Ensure that the non-traditional data is directly relevant to the issue or challenge you are addressing. It should provide meaningful insights or solutions.

2. **Data Availability:** Assess the availability of non-traditional data sources and map relevant organizations that might provide this data.

3. **Data Quality and Reliability:** Assess the quality and reliability of the data source. Is it accurate, up-to-date, and from a credible and trustworthy origin? Poor-quality data can lead to erroneous conclusions.

4. **Individual and Group Consent:** Ensure that data is collected and used with the informed consent of individuals (in the case of children, their legal guardians would need to do this, and the children would provide their assent), especially when dealing with sensitive or personal information. Respect privacy and self/group-determination rights, especially for vulnerable populations like children.

5. **Data Security:** Implement robust security measures to safeguard the data against unauthorized access, breaches, or misuse.

6. **Ethical Considerations:** Consider the ethical implications of data collection and usage, especially when it involves children or sensitive personal information. Adhere to ethical standards and relevant frameworks.

7. **Bias Mitigation:** Evaluate the potential for bias in the data and employ strategies to identify and mitigate bias to ensure fairness and equity.

8. **Legal and Regulatory Compliance:** Ensure compliance with applicable laws and regulations, including data protection laws and regulations governing the rights of children.

4.1. RELEVANCE TO THE PROBLEM

Before exploring non-traditional data sources, it is crucial to articulate a specific question or problem that requires attention. This question serves as a **guiding principle, enabling the evaluation of whether non-traditional data can effectively address the issue at hand**. By defining the problem, the context for data collection is set, and it becomes easier to determine the minimum viable data needed to find solutions. This approach minimizes the risk of data misuse and data overload and streamlines the selection of pertinent sources. Actions that can be taken to pursue this criteria can include:

- Clearly define the specific question or problem that needs attention, ensuring it's specific and actionable.
- Assess if the identified question aligns directly with the issue at hand, avoiding ambiguity.
- Outline the minimum data requirements necessary to address the problem, preventing unnecessary data collection.
- Establish clear criteria for selecting non-traditional data sources based on their direct relevance to the defined problem.

4.2. DATA AVAILABILITY

After establishing the question and relevance of non-traditional data to the problem, the next step is to assess the **availability of non-traditional data sources** capable of addressing it. This involves identifying potential organizations or entities that may possess the required data. These sources can span governmental agencies, private companies, research institutions, or social media platforms. The goal is to pinpoint sources with relevant information to inform the subsequent stages of analysis. Actions that can be taken to pursue this criteria can include:

- Identify specific organizations, platforms, or entities that may possess the required non-traditional data.
- Investigate the accessibility and availability of data from identified sources.
- Explore partnerships with governmental agencies, private companies, research institutions, and social media platforms.

4.3. DATA QUALITY AND RELIABILITY

Once relevant non-traditional data have been identified as available, it is important to assess their quality and reliability. Ensuring that the data is **accurate, up-to-date, and sourced from trustworthy origins** is essential. Poor-quality data can lead to misguided conclusions or decisions. To verify credibility, an assessment of the provenance and reputation of the data source is necessary, promoting confidence in the subsequent analytical processes. Actions that can be taken to pursue this criteria can include:

- Establish criteria for data quality and reliability (see, for example, the UNICEF Data Quality Framework), with an emphasis on accuracy and trustworthiness.
- Conduct assessments of the provenance and reliability of the data source (see, for example, the RD4C Data Ecosystem Mapping Tool) to build confidence in subsequent analytical processes.

4.4. INDIVIDUAL ASSENT AND GROUP CONSENT

Particularly when dealing with sensitive or personal information, adherence to informed consent principles is essential throughout the data lifecycle. This becomes even more critical when handling vulnerable populations, such as children. It emphasizes the importance of obtaining explicit and meaningful consent/assent from individuals or groups before collecting and utilizing their data, respecting their autonomy and agency. While children may not be able to legally consent, it is still valuable for organizations to secure their assent to programs to guarantee that they continue to be seen as legitimate and in line with their expectations. Actions that can be taken to pursue this criteria can include: Establish criteria for data quality and reliability (see, for example, the UNICEF Data Quality Framework), with an emphasis on accuracy and trustworthiness.

- Develop explicit and comprehensive consent processes for data subjects.
- Ensure that individuals or groups fully understand the purpose and implications of data usage.
- Document meaningful consent procedures meticulously, particularly when working with vulnerable populations such as children.

4.5. DATA SECURITY

To safeguard against potential threats, robust security measures must be implemented throughout the data lifecycle. This includes **encryption, stringent access controls, and compliance with relevant data protection regulations**. The objective is to fortify the data against unauthorized access, breaches, or any form of misuse, thereby maintaining the confidentiality and integrity of the collected information. Actions that can be taken to pursue this criteria can include:

- Implement encryption and stringent access controls to secure non-traditional data.
- Ensure compliance with data protection regulations and industry best practices.
- Develop a detailed data security strategy aligned with the specific needs of the project.
- Regularly audit and update security measures to respond to evolving threats and vulnerabilities.

The process of setting up data security systems can be complex and while resources may be available to guide this process (e.g. ICRC Handbook on Data Protection in Humanitarian Action, UNICEF Policy on Personal Data Protection), a dedicated conversation with security professionals within UNICEF and UNHCR may be useful in helping the office set up proper policies and procedures and address the specific data and context in which the office is residing. The RD4C initiative cannot offer specific recommendations without knowing details of UNICEF Romania's data systems or the kinds of non-traditional data it might be interested in pursuing.

4.6. ETHICAL CONSIDERATIONS

It is vital to approach the use of non-traditional data sources for children services with a strong ethical foundation. **Meaningful individual and group consent/assent, data minimization, and transparency** in data handling are some of the principles that can be prioritized. This involves a conscientious approach throughout the data lifecycle, ensuring that practices align with **broader ethical principles** and do not compromise the rights or well-being of groups and individuals. Actions that can be taken to pursue this criteria can include:

- Establish clear ethical guidelines for handling non-traditional data, particularly for children service delivery. By explicitly outlining the interests and rights of children for the team will allow the team to be aware of what they need to uphold, protect, or guard against;
- Develop a process to handle requests to share or receive data. The team might, for example:
    - Assess the purpose of sharing and its legitimacy;

- Evaluate what is necessary to share (e.g. raw data, clean data, personal data) and the risk of each;
- Define what data is sensitive in an operational context;
- Identify what data sharing arrangements are necessary to ensure that partners uphold data responsibility principles and ethical requirements.

- Within the multi-functional team, assess what is needed to secure the data to guarantee the privacy of children and prevent unwanted exposure. Practitioners might implement:
  - Physical security (e.g. making data only accessible from certain computers or office locations);
  - Technological security (e.g. making it so that certain user profiles have restricted access; and
  - Procedural (e.g. having standard operating procedures with an approval chain to create users, exchange data).

4.7. BIAS MITIGATION

Acknowledging the potential for bias in data is a critical aspect of responsible data analysis. It is indeed important to **assess whether the collection and use of non-traditional data inadvertently perpetuate biases or contribute to unfair treatment of data subjects**, especially towards vulnerable populations like children. Evaluating and addressing existing and potential biases, whether inherent or introduced during the data and decision cycles, is imperative for establishing whether or not the use of non-traditional data is beneficial for children. Actions that can be taken to pursue this criteria can include:

- Identify potential biases throughout the non-traditional data lifecycle;
- Ensure that data has been cleaned before it is analyzed, as data that is not well organized before analysis will not generate reliable findings;
- Evaluate the type of bias and develop protocols to address biases specific to vulnerable populations like children.
- Ask critical questions to minimize biases and ensuring fairness and equity, such as:
  - What categories are relevant to the decisions that we hope to make with the data?
  - What disaggregation points are relevant to the purpose (e.g. location, country/ region of origin);
  - What are the limitations of the dataset and what can it not represent?

4.8. LEGAL AND REGULATORY COMPLIANCE

Legal and regulatory compliance is a fundamental aspect of any data initiative, including non-traditional data activities. This encompasses **adherence to established laws, regulations, and industry standards** governing the collection, processing, storage, sharing and use of data. It serves as a safeguard against legal ramifications and can reinforce ethical practices in handling non-traditional data. Actions that can be taken to pursue this criteria can include:

- Ensure adherence to specific data protection laws and regulations applicable to the project.
- Regularly review compliance measures to stay updated with evolving legal requirements.
- Establish protocols for handling data in accordance with local and international laws.
- Train personnel to understand and comply with the legal landscape governing non-traditional data use.

5. DUE DILIGENCE FOR ENSURING RESPONSIBLE DATA PRACTICES IN PRIVATE SECTOR PARTNERSHIPS

When collaborating with private sector organizations to use the non-traditional data they collect and handle, it is essential to ensure responsible data practices on their part. Due diligence is crucial to ensuring the responsibility and trustworthiness of these organizations.

Due diligence is a continuous process, and is to be revisited on a regular basis. Indeed, because of the vast diversity in contexts, organizations, and technologies, no two due diligence processes will be the same or stay the same. Hence, the process illustrated here is only meant to be a starting point and, in order to be effective, needs to be tailored to the specific legal, cultural and political contexts and needs of the organization.

In particular, we present six steps: (1) Determination of Due Diligence Scope, (2) Internal Data Collection, (3) Risk Ranking and “Red-Flag” Identification, (4) Additional Due Diligence Tool for High-Risk Cases, (5) Approval-Based on Risk Level, and (6) Post-Engagement Due Diligence.

Step 1: Determination of Due Diligence Scope

WHAT: The first step in the due diligence process is to define the scope of the assessment. Organizations identify the data subjects and sources, partners, and contexts they will be engaging with to understand the potential risks involved. This step is essential as it lays the foundation for the rest of the process, allowing organizations to focus their efforts on the areas that matter most.

WHO: This phase requires collaboration among stakeholders from legal, compliance, and strategic planning teams. Key decision-makers and project managers play a crucial role in outlining the scope. An example of a team conducting this assessment is the following:

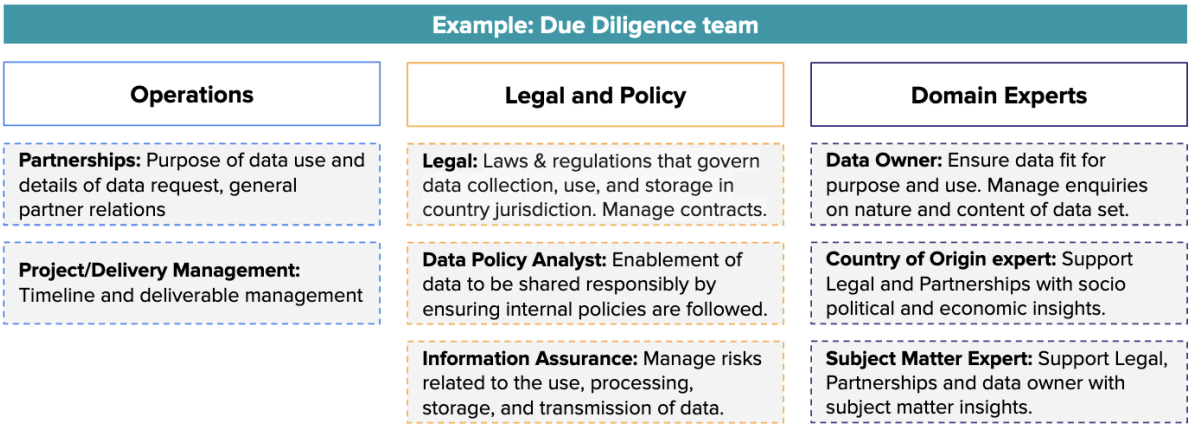

Figure 1: Example team of specialists, and their roles, who could make-up a due diligence team. We have avoided using specific job titles as these may change on a case-by-case basis. We have opted to rather name teams or departments that could be involved.

HOW:

- **Stakeholder Consultation:** Engage with relevant stakeholders to identify the specific data requirements and objectives of the partnership.
- **Legal Expertise:** Consult legal experts to ensure alignment with data protection laws and ethical standards.
- **Documentation:** Document the defined scope, outlining the boundaries within which the due diligence assessment will be conducted.

This step sets the groundwork for a focused and effective due diligence process, ensuring that the assessment aligns with the organization's objectives and ethical standards.

Step 2: Internal Data Collection

WHAT: In Step 2, the primary focus is on internal data collection, a critical phase for understanding the contextual and geographic risk factors associated with using non-traditional data from a private company. This step involves gathering vital information about the regulatory, legal, and ethical landscape in which the project will operate, as well as about the private company's transparency in community engagement, data governance practices, past performance, and ethical alignment. Key aspects include assessing data protection regulations, rule of law, due process, corruption tolerance, and human rights records.

WHO: Personnel involved in this phase include individuals with expertise in legal compliance, data governance, and contextual understanding relevant to the initiative. Collaboration with

external entities, such as data governance or legal consultants, becomes essential for a detailed exploration of the selected private company's data practices.

HOW:

- **Compliance Assessment:** Conduct a thorough analysis of the technology company's adherence to both the country's evolving data protection regulations and international conventions, such as the Convention on the Rights of the Child (CRC). Engage legal experts to ensure comprehensive compliance.
- **Transparency Evaluation:** Examine the private company's efforts to maximize transparency, especially in regions with high illiteracy rates. Assess their use of child-friendly visual materials and interactive sessions to inform local communities about data collection, usage, and the rights of families and children involved.
- **Data Governance Documentation:** Request and review explicit documentation from the private company outlining decision-making responsibilities throughout the data lifecycle. Ensure a clear decision trail is established and maintained.
- **Past Performance Examination:** Investigate the private company's past projects, emphasizing their successful implementation of non-traditional data initiatives in similar challenging environments. Assess their expertise in handling sensitive data within comparable contexts.
- **Ethical Alignment Check:** Verify that the partnership's goals align with commonly shared and established ethical principles in the context of refugee- and child-focused initiatives.

Step 3: Risk Ranking and "Red-Flag" Identification

WHAT: In this step, the focus shifts to evaluating potential risks associated with the private sector partner's data practices. The goal is to create a systematic approach for prioritizing and addressing risks, including the identification of red flags signaling non-compliance or ethical concerns.

WHO: Key personnel involved in risk management, compliance, and project management are instrumental in this phase. Collaboration with experts in data governance and security may also be crucial.

HOW:

- **Risk Matrix Development:** Create a risk matrix that considers factors such as data sensitivity, regulatory compliance, and ethical considerations. To do so, it may be useful to develop a resource along the lines of the Risks Across the Data Life Cycle Framework illustrated below.
- **Red Flag Identification:** Define specific "red flags" indicating potential issues with partners, such as non-compliance history or security breaches. Determine the degree

to which the partner has demonstrated these “red flags” and the degree to which they can and cannot be addressed.

- **Quantitative Assessment:** Assign numerical values or scores to different risk factors to quantify and prioritize risks. Tally the score of the proposed collaboration(s) and discuss what the overall score suggests about the overall risk of engagement.

Data Lifecycle Stages	Risks
Planning	Unclear purpose and goals of the project
	Financial resources and stakeholder partnerships hard to find and maintain
Collecting	Data subjects and sources and their consent not being protected
	Poor quality (duplication or inconsistencies) or dirty data, data bias/non-representation,
Processing	Poor anonymization and privacy of data subjects and sources
	Insufficient security provisions, inaccessibility, aggregation and correlations challenges
Sharing	Lack of trust and communication among partners, conflicting jurisdictions, different levels of security
	Poor balance between interoperability and context-rich, relevant data
Analyzing	Poor focus and rigor in the data analysis, inaccurate data modeling, poor problem and definition design
	Black boxes, blindspots, unexplainable automated decision making, inequity production and reproduction (eg. biased algorithms)
Using	Faulty reporting (eg. misinterpretation), malicious actors using findings for unforeseen and/or harmful purposes
	Poor communication strategy and external use and repurpose of findings

Table 1: The Risks Across the Data Life Cycle Framework.

Step 4: Additional Due Diligence Tool for High-Risk Cases

WHAT: For partnerships identified as high-risk in the previous step, additional due diligence tools and methodologies are applied to gain a deeper understanding of the partner's data management practices, so as to potentially mitigate risks.

WHO: Engage specialized professionals, such as third-party auditors, data analysts, and legal experts, to conduct a more in-depth assessment.

HOW:

- ***Additional Diagnosis Tool:*** In cases where high-risk factors are identified during the initial due diligence process, additional external due diligence becomes necessary. To this end, we have developed a Due Diligence Diagnostic Tool (Appendix A) that illustrates the critical aspects and questions that a project team may consider when faced with such circumstances. The tool is organized according to the stages of the data lifecycle (example from the Planning Stage in Table 2), and aims to determine whether the engagement is still viable through mid-way solutions or if it should be reevaluated altogether.

Stage	Risks Identified (Using the Risks Across the Data Life Cycle Framework)	Additional Due Diligence and Diagnosis: What could the project team consider for each risk?
Planning	The partnership has unclear purposes and goals	All actors involved in the partnership may consider completing a project case if one has not been provided. At a minimum, the need and purpose of the project have to be defined and agreed on. POTENTIAL QUESTIONS TO ASK: • Why does the partnership have an unclear purpose and undefined goals? • Who is responsible for defining these? • Are there established metrics and indicators of success? If not, can we establish these?

	Financial resources and stakeholder partnerships are difficult to find and maintain	It's recommended that all actors involved in the project come together and analyze the reasons for a lack of resources and partnerships. The most desired outcome is for clear processes and accountabilities to be established.
		
		POTENTIAL QUESTIONS TO ASK:
		• What financial resources are missing?
		  ○ Why are they missing?
		  ○ Who and what is needed to rectify the situation?
		• Which stakeholder partnerships are needed?
		  ○ Why are these relationships difficult to establish and/or maintain?
		  ○ What can be done to improve the current situation? More collaboration? Greater transparency?

Table 2: The Planning Stage from The Additional Diagnosis Tool.

Step 5: Approval Based on Risk Level

WHAT: Once the risk assessment is complete, a decision-making process is established to approve or reject the partnership based on the identified risk levels.

WHO: Cross-functional teams involving legal, compliance, and project managers play a crucial role in decision-making.

HOW:

- **Informed Decision-Making:** Ensure that decision-makers are well-informed about specific risks and mitigations identified during due diligence.
- **Documentation:** Document the decision-making process, including any conditions or requirements for approval.
- **Threshold Definition:** Clearly define criteria and thresholds for acceptable risk levels based on the risk assessment. To facilitate the approval process, the due diligence team develops a checklist using insights from the Risks Across the Data Life Cycle framework illustrated above (Table 1).

Step 6: Post-Engagement Due Diligence

WHAT: Step 6 involves the process of post-engagement due diligence monitoring. After the data initiative has been initiated, this phase focuses on continuous monitoring and assessment of the private sector partner's data practices. Due diligence is not a one-time process; it requires ongoing assessments to adapt to changing contexts, regulations, and organizational needs. Regular reviews help organizations stay accountable and responsive to evolving data risks. It is thus important to establish channels for ongoing communication and involve relevant stakeholders in periodic reviews.

WHO: The responsibility for post-engagement due diligence monitoring may lie with the head of programs, supported by the project team and internal lawyers. Additionally, involved data subjects, whose data is being used, are considered key stakeholders and may provide valuable insights during the monitoring process.

HOW:

- **Regular Audits:** Scheduled audits are conducted to review the data initiative's activities, ensuring that the initial risk assessment and compliance requirements continue to be met. These audits help identify any new risks or ethical concerns that may have emerged during the engagement.
- **Data Subjects Engagement:** Ongoing engagement with data subjects is maintained to gather feedback and assess their experiences with the data initiative. This direct interaction helps gauge the impact on data subjects and identify any potential privacy or ethical issues.
- **Responsive Actions:** If any new risks or ethical challenges are identified during post-engagement monitoring, the project team, in collaboration with internal lawyers, takes prompt and responsive action to address and mitigate these issues. Proactive measures are essential to maintain data responsibility throughout the data initiative.
- **Periodic Reporting:** Regular reporting on post-engagement due diligence findings is shared with relevant stakeholders, including organizational leaders, data subjects, and partnering organizations or governments. Transparent reporting ensures accountability and builds trust among stakeholders.

6. CONCLUSION

In navigating the opportunities and challenges of using non-traditional data sources for service provision to children, it is evident that the private sector's innovative use of data and technology holds great potential. However, this potential must be harnessed responsibly, especially considering the unique vulnerabilities of the children involved. The framework presented in this document aims to address critical aspects from ethical considerations to legal compliance, offering a comprehensive guide for organizations venturing into this space.

The recognition that children face particular vulnerabilities when it comes to the use of non-traditional data underscores the importance of a cautious and principled approach. The challenges associated with surveillance, agency limitations, and potential long-term impacts on trust necessitate a careful balance between leveraging non-traditional data sources and protecting the rights and well-being of children.

The exploration of non-traditional data sources, including social media, mobile data, economic indicators, geospatial data, and crowdsourced information, unveils a wealth of possibilities. Improved targeting of services, timely assistance, and efficient resource allocation emerge as key benefits, promising a transformative impact on service delivery to children.

To harness these opportunities effectively, due diligence is paramount. The six-step due diligence process outlined here, from determining the scope to post-engagement monitoring, hopes to provide a framework for ensuring responsible data practices in private sector partnerships.

In conclusion, the responsible use of non-traditional data sources for children services demands a holistic and proactive approach. Balancing innovation with ethical considerations, legal compliance, and ongoing diligence ensures that the private sector can be a valuable ally in addressing the challenges faced by children, ultimately contributing to their well-being and development.

APPENDIX: ADDITIONAL DUE DILIGENCE AND DIAGNOSIS - WHAT TO CONSIDER IN CASE HIGH RISK IS IDENTIFIED FOR A PARTICULAR PROJECT OR DATASET

The essential goal of due diligence is to assess the potential risks of using private sector non-traditional data. Additional due diligence may be necessary when potential risks are identified, to ultimately find potential solutions or–alternatively–cancel the engagement.

This tool illustrates the additional due diligence and diagnosis to be conducted **in case high risk is identified** during the due diligence process, particularly unfolding Step 4: Additional External Due Diligence. The aim is to show the issues and questions the project team could consider in case of high risk, when trying to determine whether the engagement is still worth establishing through **mid-way solutions** or not.

1. **Questions to be asked for each risk across the data lifecycle**

FRAMEWORK: The mitigation strategies are very broad and only serve as a starting point for the project team to consider different elements and options in case of identified high risk. These will need to be tailored to each case and further refinement of such strategies will have to take place once the specific context is identified.

Stage	**Risks identified**	**Additional Due Diligence and Diagnosis: What could the project team consider for each risk?**
Planning	The project has unclear purposes and goals	All actors involved in the project may consider completing a project case if one has not been provided. At a minimum, the need and purpose of the project have to be defined and agreed on. POTENTIAL QUESTIONS TO ASK: • Why does the project have an unclear purpose and undefined goals? • Who is responsible for defining these? • Are there established metrics and indicators of success? If not, can we establish these?

	Financial resources and stakeholder partnerships are difficult to find and maintain	It's recommended that all actors involved in the project come together and analyze the reasons for a lack of resources and partnerships. The most desired outcome is for clear processes and accountabilities to be established.
		
		POTENTIAL QUESTIONS TO ASK:
		• What financial resources are missing?
		  ○ Why are they missing?
		  ○ Who and what is needed to rectify the situation?
		• Which stakeholder partnerships are needed?
		  ○ Why are these relationships difficult to establish and/or maintain?
		  ○ What can be done to improve the current situation? More collaboration? Greater transparency?

Collecting	The consent of data subjects and sources not being protected	Identify the reasons for this and then determine the next appropriate steps to take. As a guide: • If the cause has to do with national or local policy and legal drivers, the engagement may have to be canceled. • If the cause has to do with poor management, but there is indeed willingness to uphold consent, responsible actors are to be identified and made accountable. • If the cause has to do with negligence, it should be clear whether the actors involved are indeed interested in protecting data subjects' consent, and if and how it may be feasible to do so.

	Poor quality (duplication or inconsistencies) or dirty data, data bias/non-representation	The actors involved in the collection of the data are to be involved in the mitigation strategy, which is to be co-designed with them.
		• If data is duplicated or dirty, delete double entries and clean the data.
		• If data is biased, further investigations need to be carried out to understand why.
		
		POTENTIAL QUESTIONS TO ASK:
		• Who is responsible for cleaning the data?
		  ○ Why is there poor quality data? How can we ensure we minimize the chances of this happening again in the future?
		• In what way is the data biased? E.g. Due to poor representation, discrimination against certain groups and perpetuation of stereotypes.
		  ○ Can this be addressed?
		  ○ Who should be involved in the process of improving the quality of the dataset so as to minimize the risks of unintended consequences ?
		  ○ How can we improve data quality assurance processes over time?

Processing	Poor anonymization and privacy of data subjects and sources	Anonymizing data ensures the risk of re-identification is minimized. If it is discovered that data has not been, or has been poorly anonymized, then you may wish to consider applying the following treatments to the dataset in order to safeguard privacy: • Suppression > Remove data from a dataset. Best applies to any direct identifiers. • Randomization > Add noise or shuffle values while maintaining patterns in the dataset. • Pseudonymization > While not a method of anonymization, this method can minimize the chances of data subjects and sources being reidentified. Techniques include encryption, tokenization and hashing. POTENTIAL QUESTIONS TO ASK: • Considering the nature of the dataset, which is the best anonymization technique to apply? Once anonymization techniques have been applied: • Is it possible to single out and identify an individual? • Can public records be linked together and used to identify an individual in the dataset?
	Insufficient security provisions, inaccessibility, aggregation and correlations challenges	All stakeholders could meet to discuss the severity and variety of harms that are possible should the data be released. Consider delaying the launch of the data/project until the risk is sufficiently reduced. QUESTIONS TO ASK: • Do we have the required resources to tackle the challenge and minimize risk to a satisfactory level? • What actions need to be taken to minimize the severity and variety of harms that could result from this data?
Sharing	Lack of trust and communication among partners, conflicting jurisdictions, different levels of security	It is important to understand the legal and ethical basis for sharing personal data. This will be geography and context-dependent. POTENTIAL QUESTIONS TO ASK: • What is fueling the distrust between stakeholders? What can be said and done to reduce fear and improve cooperation? • Do we understand the nature of the data and the reasons for sharing it? • Are we exercising the principle of data proportionality? • Are adequate security measures in place to keep data safe for as long as it is needed?

	Poor balance between interoperability and context-rich, relevant data	Successful sharing of data requires datasets to be interoperable across many systems. It also requires data to be meaningful and useful. POTENTIAL QUESTIONS TO ASK: • What needs to be done to improve the interoperability features of this dataset? • What context can be added to the data to improve its relevance? • What processes do we need to consider implementing during the collection phase to ensure that data is more meaningful in the future?
Analyzing	Poor focus and rigor in the data analysis, inaccurate data modeling, poor problem and definition design	If it is discovered that data has been poorly or incorrectly analyzed, then focus stakeholder attention on what can be improved before any data is shared. POTENTIAL QUESTIONS TO ASK: • What needs to be done to improve the quality of analysis of this dataset? • Do we have the required resources to tackle the challenge and minimize risk to a satisfactory level?
	Black boxes, blindspots, unexplainable automated decision making, inequity production and reproduction (eg. discriminatory algorithms)	Unexplainable systems have the potential to cause a number of harms. Consider delaying the launch of the data/project until the risk is sufficiently reduced. POTENTIAL QUESTIONS TO ASK: • Can we improve the explainability of this data system? • If not, is it possible to implement a system that can be explained?
Using	Faulty reporting (eg. misinterpretation), malicious actors using findings for unforeseen and/or harmful purposes	You may wish to employ the skills of an experienced PR or communications professional during a situation of this nature. POTENTIAL QUESTIONS TO ASK: • Has this potentially been a misunderstanding? Could we contact the actors and attempt to set the record straight? • If not, can we develop a fact-correcting communications strategy in order to rectify any disinformation and misinformation? • Who could we partner with to amplify these messages and ensure we get the facts in front of more people?

	Poor communication strategy and external use and repurpose of findings	Working with an experienced PR or communications professional during a situation of this nature is recommended. • POTENTIAL QUESTIONS TO ASK: • What immediate and short-term tactics can we use to improve the public narrative around this work? • Can we lean on any partners or influential individuals to support us during this endeavor? • What is our long-term strategy to ensure we course correct over time and are still able to achieve our desired outcomes?