DATA INNOVATION RISK ASSESSMENT TOOL

CHECKLIST

Rationale for the checklist: Large-scale social or behavioural data may not always contain directly identifiable personal data and/or may be derived from public sources. Nevertheless, its use could potentially cause harm to individuals.

Data use should be always assessed in light of its impact (negative or positive) on individual rights**.** This risk assessment tool (or checklist) outlines a set of minimum checkpoints, intended to help you to understand and minimize the risks of harms and maximize the positive impacts of a data innovation project (and is intended primarily for projects implemented within international development and humanitarian organizations).

How to use the checklist: The checklist should be considered before a new project is launched, when new sources of data or technology are being incorporated into an existing project, or when an existing project is substantially changed. In particular, this assessment should consider every stage of the project's data life cycle: data collection, data transmission, data analysis, data storage, and publication of results. If possible, the questions raised by the checklist should be considered by a diverse team comprised of the project leader as well as other subject matter experts, including – where reasonably practical – a representative of the individuals or groups of individuals who could be potentially affected. Consider consulting with data experts, data privacy experts, and legal experts so that they can assist with answering these questions and help to further mitigate potential risks, where necessary.

Note that the checklist was developed by Global Pulse as part of a more comprehensive Risk, Harms and Benefits Assessment, consisting of Two Steps: (I) Initial Assessment and (II) Comprehensive Risks, Harms and Benefits Assessment. This checklist is an Initial Assessment that should help to determine whether a more comprehensive Risk, Harms and Benefits Assessment should be conducted.

Nature of the checklist: This checklist is not a legal document, and is not based on any specific national law. It draws inspiration from international and regional frameworks concerning data privacy and data protection. The document provides only a minimum set of questions and guiding comments. The checklist and guiding comments are designed primarily as a general example for internal self-regulation. As this checklist offers only minimum guidance, you are encouraged to expand the list depending on the project's needs, risks, or specific context, or in response to the evolving data landscape.

Depending on the implementing organization (its legal status/nature) and applicable laws, the guiding principles, standards and basis for answering these questions may need to be changed.

The latest version of this checklist and the full version of the comprehensive assessment will be made available at a later stage (independently of this publication) and will be available at www.unglobalpulse.org/privacy. For more information or to provide input on the checklist, please contact dataprivacy@unglobalpulse.org. This checklist is a living document and will change over time in response to the evolving data landscape.

Instructions for completion

Please be sure to answer all of the questions by choosing at least one of the following answers: "Yes," "No," "Don't Know," or "Not Applicable." Please use the comments column to explain your decision where necessary.

For every "Not Applicable" answer, please provide an explanation in the comments. Every "Don't Know" answer should be automatically considered a risk factor that requires further consultation with a domain expert before a project is undertaken. Once you have properly consulted with an expert regarding the issue, please be sure to go back to the checklist and change your answer in the form to finalize your checklist.

A final decision based on the checklist should not be made if there is any answer marked "Don't Know."

DATA INNOVATION RISK ASSESSMENT TOOL

Part 1: Type of Data

Personal Data: For the purposes of this document, personal data means any data relating to an identified or identifiable individual, who can be identified, directly or indirectly, by means reasonably likely to be used related to that data, including where an individual can be identified from linking the data to other data or information reasonably available in any form or medium. If you are using publicly available data, note that this data can also be personal, and therefore may involve some of the same considerations as non-public personal data.

1.1 Will you use (e.g. collect, store, transmit, analyse etc.) data that directly identifies individuals?

Personal data directly relating to an identified or identifiable individual may include, for example, name, date of birth, gender, age, location, user name, phone number, email address, ID/social security number, IP address, device identifiers, account numbers etc.

- ☐ Yes
- ☐ No
- ☐ Don't Know
- ☐ Not Applicable

Comments:

1.2 Will you use data that does not directly identify an individual, but that could be used to single out a unique individual by applying existing and readily accessible means and technologies?

Keep in mind that de-identified data (e.g., where all personal identifiers - such as name, date of birth, exact location, etc. - are removed), while not directly linked to an individual(s) or group(s) of individuals, can still single out an individual(s) or group(s) of individuals with the use of adequate technology, skills, and intent, and thus may require the same level of protection as explicit personal data. To determine whether an individual(s) or group(s) of individuals is identifiable, consider all of the means reasonably likely to be used to single out an individual or group(s) of individuals. Factors that influence a likelihood of re-identification include availability of expertise, costs, amount of time required for re-identification and reasonably and commercially available technology.

- ☐ Yes
- ☐ No
- ☐ Don't Know
- ☐ Not Applicable

Comments:

1.3 Will you use sensitive data?

Any data related to (i) racial or ethnic origin, (ii) political opinions, (iii) trade union association, (iv) religious beliefs or other beliefs of a similar nature, (v) physical or mental health or condition (or any genetic data), (vi) sexual orientation; (vii) the commission or alleged commission of any offence, (viii) any information regarding judicial proceedings, (ix) any financial data, or any information concerning (x) children; (xi) individual(s) or group(s) of individuals, who face any risks of harm (physical, emotional, economical etc.) should be considered as sensitive data. Consider that the risk of harm is much higher for sensitive data and stricter measures for protection should apply if such data is explicit personal data or is reasonably likely to identify an individual(s) or a group of individuals.

- ☐ Yes
- ☐ No
- ☐ Don't Know
- ☐ Not Applicable

Comments:

NEXT STEP:
As you go through the remaining sets of questions, please keep the data type you identified in the section above in mind. If you answered "YES" to at least one of the question above, the risk of harms is increased.

Part 2: Data Access

2.1 Means for data access

This question aims to help you understand the way in which you have obtained your data, to ensure that there is a legitimate and lawful basis for you to have access to the data in the first place. It is important to understand that whether directly or through a third party contract, data should be obtained, collected, analyzed or otherwise used in conformity with the purposes and principles of the Universal Declaration of Human Rights and the International Covenant on Civil and Political Rights and other applicable laws, including privacy laws.

DATA INNOVATION RISK ASSESSMENT TOOL

How was the data obtained? (Choose from one of the answers below)

- ☐ A: Directly from individual(s) (e.g., survey)
- ☐ B: Through a data provider *(e.g. website, social media platform, telecom operator)*
- ☐ C: Don't know

Comments:

NEXT STEP:
If you answered "A," please proceed to section 2.2, "Legitimacy, lawfulness and fairness of data access and use." If you selected "B," you can skip 2.2 and proceed to point 2.3, "Due Diligence on third party data providers." If you selected "C", consult with your legal expert before proceeding further.

2.2 Legitimacy, lawfulness and fairness of data access and use

Lawfulness, legitimacy, and fairness. Any personal data must be collected and otherwise used through lawful, legitimate, and fair means. Personal data use may be based, for example, on one or more of the following legitimate bases, subject to applicable law: i) consent of the individual whose data is used; ii) authority of law; iii) the furtherance of international (intergovernmental) organizational mandates (e.g. in case where an international intergovernmental organization is the holder of the mandate and is the implementer of a data project); iv) other legitimate needs to protect the vital interest of an individual(s) or group(s) of individuals. Keep in mind that the legitimacy and lawfulness of your right to use the data must be carefully assessed, taking into account applicable law, the context, legal status of your organization; and the above bases (i- iv) are only included as examples for the purposes of this document.

Data should always be accessed, analyzed, or otherwise used taking into account the legitimate interests of those individuals whose data is being used. Specifically, to ensure that data use is fair, data should not be used in a way that violates human rights, or in any other ways that are likely to cause unjustified or adverse effects on any individual(s) or group(s) of individuals. It is recommended that the legitimacy and fairness of data use always be assessed taking into account the risks, harms, and benefits of data use.

Informed consent should be obtained prior to data collection or when the purpose of data re-use falls outside of the purpose for which consent was originally obtained. Keep in mind that in many instances consent may not be adequately informed. Thus, it is important to consider assessing the proportionality of risks, harms and benefits of data use even if consent has been obtained.

While there may be an opportunity to obtain consent at the time of data collection, re-use of data often presents difficulties for obtaining consent (e.g., in emergencies where you may no longer be in contact with the individuals concerned). In situations where it is not possible or reasonably practical to obtain informed consent, as a last resort, data experts may still consider using such data for the best or vital interest of an individual(s) or group(s) of individuals (e.g., to save their life, reunite families etc.). In such instances, any decision to proceed without consent must be based on an additional detailed assessment of risks, harms and benefits to justify such action and must be found fair, lawful, legitimate and in accordance with the principle of proportionality (e.g., any potential risks and harms should not be excessive in relation to the expected benefits of data use).

Do you have a legitimate basis for your data access and use?

- ☐ Yes
- ☐ No
- ☐ Don't Know
- ☐ Not Applicable

Comments:

2.3 Due diligence on third party data providers

This question usually applies when you are not a data collector, but rather obtained data from a third party (e.g. telecom operator, social media platform, web site). It is important that you verify, to the extent reasonably practical, whether your data provider has a legitimate basis to collect and share the data with you for the purposes of your project. For example, have you checked whether your data provider has obtained adequate consent (e.g. directly or indirectly through the online terms of use) or has another legitimate basis for sharing the data with you for the purposes compatible with your project? (See notes on "Lawfulness, legitimacy, and fairness" above)

Does your data provider have a legitimate basis to provide access to the data for the purpose of the project?

- ☐ Yes
- ☐ No
- ☐ Don't Know
- ☐ Not Applicable

Comments:

DATA INNOVATION RISK ASSESSMENT TOOL

Part 3: Data Use

3.1 Purpose specification

The purpose of data use should be legitimate and as narrowly defined as practically possible. Furthermore, requests or proposals for data access (or collection where applicable) should also be narrowly tailored to a specific purpose. The purpose of data access (or collection where applicable) should be articulated no later than the time of data access (or collection where applicable). In answering this question, concentrate on the reason why you need the data. Also, think about articulating your answer prior to or at the time of request for data.

Have you defined the purpose for which you will be using the data as narrowly, reasonably and practically as possible?

- ☐ Yes
- ☐ No
- ☐ Don't Know
- ☐ Not Applicable

Comments:

3.2 Purpose compatibility

Any data use must be compatible to the purposes for which it was obtained. Mere difference in purpose does not make your purpose incompatible. In determining compatibility consider, for example, how deviation from your original purpose may affect an individual(s) or group(s) of individuals; the type of data you are working with (e.g. public, sensitive or non-sensitive); measures taken to safeguard the identity of individuals whose data is used (e.g. anonymization, encryption). There must be a legitimate and fair basis for an incompatible deviation from the purpose for which the data was obtained. (See notes on "Lawfulness, legitimacy, and fairness" above)

Is the purpose for which you will be using the data compatible with the purpose for which you obtained the data?

- ☐ Yes
- ☐ No
- ☐ Don't Know
- ☐ Not Applicable

Comments:

3.3 Data minimization

Data access, analysis, or other use should be kept to the minimum amount necessary (to fulfill its legitimate purpose of use as noted in points 3.1 and 3.2). Data access, collection, analysis or other use should be necessary, adequate, and relevant in relation to the purposes for which the data has been obtained. Data should only be stored for as long as necessary, and any retention of data should be lawful, legitimate, and fair. The data should be deleted and destroyed at the conclusion of the necessary period. In answering this question, consider if at any point in time in your project cycle you have the minimum data necessary to fulfill the purpose of intended use.

Are all the data that you will be using (including its storage) necessary and not excessive?

- ☐ Yes
- ☐ No
- ☐ Don't Know
- ☐ Not Applicable

Comments:

3.4 Regulation and legal compliance

Make sure that you have obtained all regulatory and other required authorizations to proceed with the Project. (For example, the use of telecom data may be restricted under telecommunication laws, and additional authorizations may be needed from a telecommunication regulator; or the transfer of data from one country to another may need to comply with rules concerning trans-border data flows). Furthermore, to ensure that you have complied with the terms under which you have obtained the data, you should check existing agreements, licenses, terms of use on social media platforms or terms of consent. If you are uncertain about this question, you should consult with your privacy and legal expert.

DATA INNOVATION RISK ASSESSMENT TOOL

Is your use of the data compliant with (a) applicable laws and (b) the terms under which you obtained the data?

- ☐ Yes
- ☐ No
- ☐ Don't Know
- ☐ Not Applicable

Comments:

3.5 Data quality

Data experts as well as domain experts should be consulted, if necessary, to determine the relevance and quality of data sets. Data accuracy must be checked for biases to avoid any adverse effects, including giving rise to unlawful and arbitrary discrimination.

Is your data adequate, accurate, up to date, reliable and relevant to the purpose of the project?

- ☐ Yes
- ☐ No
- ☐ Don't Know
- ☐ Not Applicable

Comments:

3.6 Data Security

Taking into account the available technology, cost of implementation and data type, robust technical, organizational safeguards and procedures, including efficient monitoring of data access and data breach notification procedures, should be implemented to prevent any unauthorized use, disclosure or breach of data. Embedding principles of privacy by design and employing privacy enhancing technologies during every stage of the data life cycle is recommended as a measure to ensure robust data protection. Note that proper security is necessary in every stage of your data use.

In considering security, special attention should be paid when data analysis is outsourced to subcontractors. Data access should be limited to authorized personnel, based on the need-to-know principle. Personnel should undergo regular and systematic data privacy and data security trainings. Prior to data use, the vulnerabilities of the security system (including data storage, way of transfer etc.) should be assessed.

When considering the vulnerability of your security, consider the factors that can help you identify "weaknesses" - such as intentional or unintentional unauthorized data leakage: (a) by a member of the project team; (b) by known third parties who have requested or may have access, or may be motivated to get access to misuse the data and information; or (c) by unknown third parties (e.g., due to the data or information release or publication strategy).

It is generally encouraged that personal data should be de-identified, where practically possible, including using such methods as aggregation, pseudonymization or masking, to help minimize any potential risks to privacy. To minimize the possibility of re-identification, de-identified data should not be analyzed or otherwise used by the same individuals who originally de-identified the data.
It is important to ensure that the measures taken to protect the data do not compromise the data quality, including its accuracy and overall value for the intended use.

Have you employed appropriate and reasonable technical and administrative safeguards (e.g. strong security procedures, vulnerability assessments, encryption, de-identification of data, retention policies, confidentiality/non-disclosure, data handling agreements) to protect your data from intentional or unintentional disclosure, leakage or misuse?

- ☐ Yes
- ☐ No
- ☐ Don't Know
- ☐ Not Applicable

Comments:

DATA INNOVATION RISK ASSESSMENT TOOL

Part 4: Communication about your project

4.1 Transparency

Transparency is a key factor in helping to ensure accountability, and is generally encouraged. Transparency can be achieved via communication about your project (including providing adequate notice about the data use, as well as the principles and policies governing the data use). Making the outcomes of your data innovation project public can also be important for innovation.

Note, that making data (produced as an output of your project) open is an element of transparency. If you decide to make a data set open, you must conduct a separate assessment of risks, harms and benefits. In this case, you may also want to provide transparent notices on the process and applicable procedures for making the data set open.

Did or will you communicate about the data use (publicly or to other appropriate stakeholders)?

- ☐ Yes
- ☐ No
- ☐ Don't Know
- ☐ Not Applicable

Comments:

4.2 Level of transparency

Being transparent about data use (e.g., publishing data sets, publishing an organization's data use practices, publishing the results of a data project, etc.) is generally encouraged when the benefits of being transparent are higher than the risks and possible harms. Also note, that level of detail (e.g., the level of aggregation) in a data set that is being made open should be determined after a proper assessment of risks and harms.

Particular attention should be paid to whether, for example, publishing non-sensitive details about a project or making non-identifiable datasets open can cause a mosaic effect with another open datasets. Accidental data linking or mosaic effect can make an individual(s) or group(s) of individuals identifiable or visible, thus exposing the individual(s) or group(s) of individuals to potential risks of harms.

Are there any risks and harms associated with the publication of the collected data or resulting reports and are they proportionately high compared to the benefits?

- ☐ Yes
- ☐ No
- ☐ Don't Know
- ☐ Not Applicable

Comments:

Part 5: Third Parties

5.1 Due diligence in selecting partner third parties (e.g., research partners and service providers, including cloud computing providers, etc.).

Frequently, data related initiatives require collaboration with third parties-data providers (to obtain data); data analytics companies (to assist with data analysis); and cloud or hosting companies (for computing and storage). It is therefore important that such potential collaborators are carefully chosen, through a proper due diligence vetting process that also includes minimum check points for data protection compliance, the presence of privacy policies, and fair and transparent data-related activities.

It is also important to ensure that third party collaborators are bound by necessary legal terms relating to data protection. These may include: non-disclosure agreements and other agreements containing appropriate terms on data handling; data incident history; adequate insurance, data transfer and data security conditions among other matters.

Cloud hosting. Many projects may use cloud or other hosting services, meaning that your organization does not maintain security of the hardware. It is important to ensure that your chosen cloud or hosting provider, and the data center in which they operate, have appropriate standards of security. Security certifications could be good evidence of your cloud provider's security compliance. When considering cloud storage and computing, take into account where the data will be actually located to understand potential vulnerabilities, compliance with laws, the special status of an implementing organization, including their privileges and immunities, where applicable, or rules concerning trans-border data flows.

Are your partners, if any, compliant with at least as strict standards and basic principles regarding data privacy and data protection as outlined in this checklist?

- ☐ Yes
- ☐ No
- ☐ Don't Know
- ☐ Not Applicable

Comments:

Part 6: Risks and Harms

Any risks and harms assessment should take into consideration the context of data use, including social, geographic, political, and religious factors. For example, analysis of the movement of vulnerable groups during humanitarian emergencies in conflict-affected zones could also be used by non-intended users of data to target them with discrimination or persecution.

Any Risk, Harms and Benefits Assessment should consider the impact that data use may have on an individual(s) and/or group(s) of individuals, whether legally visible or not, and whether known or unknown at the time of data use.

When assessing your data use, consider how it affects individual rights. Rather than taking rights in opposition to each other, assessing the effect of data on individual rights in conjunction is recommended wherever possible. Use of data should be based on the principle of proportionality. In particular, any potential risks and harms should not be excessive in relation to the positive impacts (expected benefits) of data use. In answering questions 6.1 and 6.2 below also consider any potential risks and harms associated with (or that could result from) every "No" answer or "Don't Know" answer that you selected in the Sections above.

.

6.1 Risks: Does your use of data pose any risks of harms to individuals or groups of individuals, whether or not they can be directly identified, visible or known?

Risks should be assessed separately from harms. Note that not all risks may lead to harms. In answering this question, it is important to concentrate on the likely risks. Types of risks may vary depending on the context. For example, some of the risks that should be considered include data leakage, breach, unauthorized disclosure (intentional or unintentional), intentional data misuse beyond the purposes for which the data was obtained/or intended to be used by your organization, risk of re-identification or singling out, data not being complete or of good quality, etc.

Note that typically data analytics result in the production of a new data set. Such an outcome should be considered as a risk as well, and must be separately assessed for risks, harms and benefits before any further use/disclosure. Also, consider bias as a risk that can be produced as a result of data use. (In many cases, bias can negatively affect an individual(s) or group(s) of individuals and lead to harms).

If you have identified potential risks, please ensure to employ the necessary mitigation measures to reduce such risks to a minimum. Ensuring proper data security is one of many strong mitigation measures (see Section 3.6). If you do not know what kind of risks exist or whether the risks are likely, it is recommended that you perform a more comprehensive Risk, Harms and Benefits Assessment (as a Step 2).

- ☐ Yes
- ☐ No
- ☐ Don't Know
- ☐ Not Applicable

Comments:

6.2 Harms: Is your project unlikely to cause harm to individuals or groups of individuals, whether or not the individuals can be identified or known?

No one should be exposed to harm or undignified or discriminatory treatment as a consequence of data use. An assessment of harms should consider such key factors as i) the likelihood of occurrence of harms; ii) the potential magnitude of harms; iii) the potential severity of harms.
The assessments should account for potential physical, emotional, or economic harms, as well as any harms that could result from infringement of individuals' rights.

Note that the risks of harms may be higher for sensitive data. Decisions concerning use of sensitive data may involve consultation with the individual(s) or a group(s) of individuals concerned (or their representative), where reasonably practical, to mitigate any risks. If you do not know what kind of harms exist or you have identified significant harms, try to perform a more comprehensive Risk, Harms and Benefits Assessment (as a Step 2 mentioned in the introduction section).

- ☐ Yes
- ☐ No
- ☐ Don't Know
- ☐ Not Applicable

Comments:

DATA INNOVATION RISK ASSESSMENT TOOL

Part 7: Decision and rationale for decision

Final Assessment

Based on your answers in Sections 1-7, explain if the risks and resulting harms are disproportionately high compared to the expected positive impacts of this project.

Questions 1.1 – 1.3; 4.2; 6.1-6.2 answered as "Yes" mean that the risk is present.
Questions 2.1 – 2.3; 3.4-3.6; 4.1; 5.1 answered as "No" mean that the risk is present.

If you answered "Don't Know" to any of the questions, consider it as a "risk factor". You should not complete this assessment unless all questions are answered "Yes", "No" or "Not Applicable".

If you have answered "Not applicable", you should make sure that you explained why it is not applicable in the Comments column.

If you found any risks, you should assess the likelihood of the risks and likelihood, magnitude and severity of the resulting harms and make sure to mitigate them before the project is undertaken.

If you identify that some of the risks or harms are unclear, or high, then you should perform a more comprehensive Risk, Harms, Benefits Assessment as a Step 2 (as mentioned in the Introduction) and engage data security, privacy and legal experts.

If you have found that the likelihood of risks and harms is very low (or non-existent) in comparison to the probability of the positive impact, you should now proceed with your project. Always bear in mind you should implement as many mitigation measures for the identified risks (even if low).

Review team

Person who performed the assessment

This should be filled out and signed by the lead person responsible for conducting the assessment

Name:
Title:
Sign:

Comments:

People who participated in or reviewed the assessment (e.g. Project Lead, Data Security, Privacy, Legal Expert)

This should be filled out by those who assisted the lead person in making the decision or who have been consulted on specific questions raised above, if any (add additional reviewers, if neccessary). You can indicate the specific questions that this person answered. If this person also helped to determine the final outcome of the overall assessment, please indicate in the comments section.

Name:
Title:
Sign:

Comments: