IMPACT Initiatives

Research Cycle Data Management at IMPACT: Personally Identifiable Information

Standard Operational Procedure

1. Overview

This document outlines the first phase in standardising protection of research data across IMPACT. Specifically, it describes the processes governing personally identifiable information within the research cycle, by specifying four immediate overall action points for implementation. The first chapter (Introduction) presents the motivation for this in the context of growing attention to data protection and the general development of standards across IMPACT. It then defines "personally identifiable information", and outlines related data protection goals. In order to achieve those goals, the second chapter introduces four overall action points. Each action point is then laid out in detail, through *(1) a requirement* with a brief explanation of its *(2) context* and *(3) details*, the step by step *(4) process* for implementation, and finally the associated *(5) accountability / responsibility structure*.

The action points below are immediate measures taken to protect personally identifiable information during the research cycle. They do not constitute an exhaustive overview of data protection within IMPACT and should be viewed as complementary to other global or in country SOPs. They are concerned specifically with research data, not with general data held by the organisation. They do not cover best practices for data protection in general, for which further guidance is forthcoming, *(the ICRC handbook on data protection is a good resource for this in the meantime)*.

2. Introduction

Since the inception of IMPACT's research department in 2017, minimum standards have been strengthened, alongside with mechanisms for centralised quality control of research design, data, analysis and reporting. Given the sensitive nature of the information we collect and the continued growth of IMPACT, it has become imperative to formalise and apply organisation-wide standards on data protection and put in place a mechanism for quality control.

As part of the research-data protection strategy, specific attention needs to be paid to personally identifiable information –that is all "information that can be used on its own or with other information to identify, contact, or locate a single person, or to identify an individual in context"[1][2][3]. This document is concerned primarily with this type of information, and aims at three main goals:

- **Minimisation of personally identifiable information:** Ensure we collect and store as little personally identifiable information as possible.
- **Limited, controlled storage and internal sharing of personally identifiable information:** Minimise the number of devices and servers holding personally identifiable information, by limiting the number of access points it passes through.
- **Personal ownership and accountability:** Assign formalised and limited access rights for all dataset that contain personally identifiable information, to specific individuals. They then hold formal accountability to protect the personally identifiable information; including ensuring that dataset storage devices are held in a protected space as far as possible.

[1] *"Management of Data Breaches Involving Sensitive Personal Information (SPI)". Va.gov. Washington, DC: Department OF Veterans Affairs. 6 January 2012. Archived from the original on 26 May 2015. Retrieved 25 May 2015.*

[2] *Stevens, Gina (April 10, 2012). "Data Security Breach Notification Laws" (PDF). fas.org. Retrieved Jun 8, 2017.*

[3] *Greene, Sari Stern (2014). Security Program and Policies: Principles and Practices. Indianapolis, IN, US: Pearson IT Certification. p. 349. ISBN 9780789751676. OCLC 897789345. Retrieved June 8, 2017.*

A summary of the overall country team action points are outlined in the box below, with a timeline on the following page. Further detailed explanations of each action point are provided in the subsequent sections.

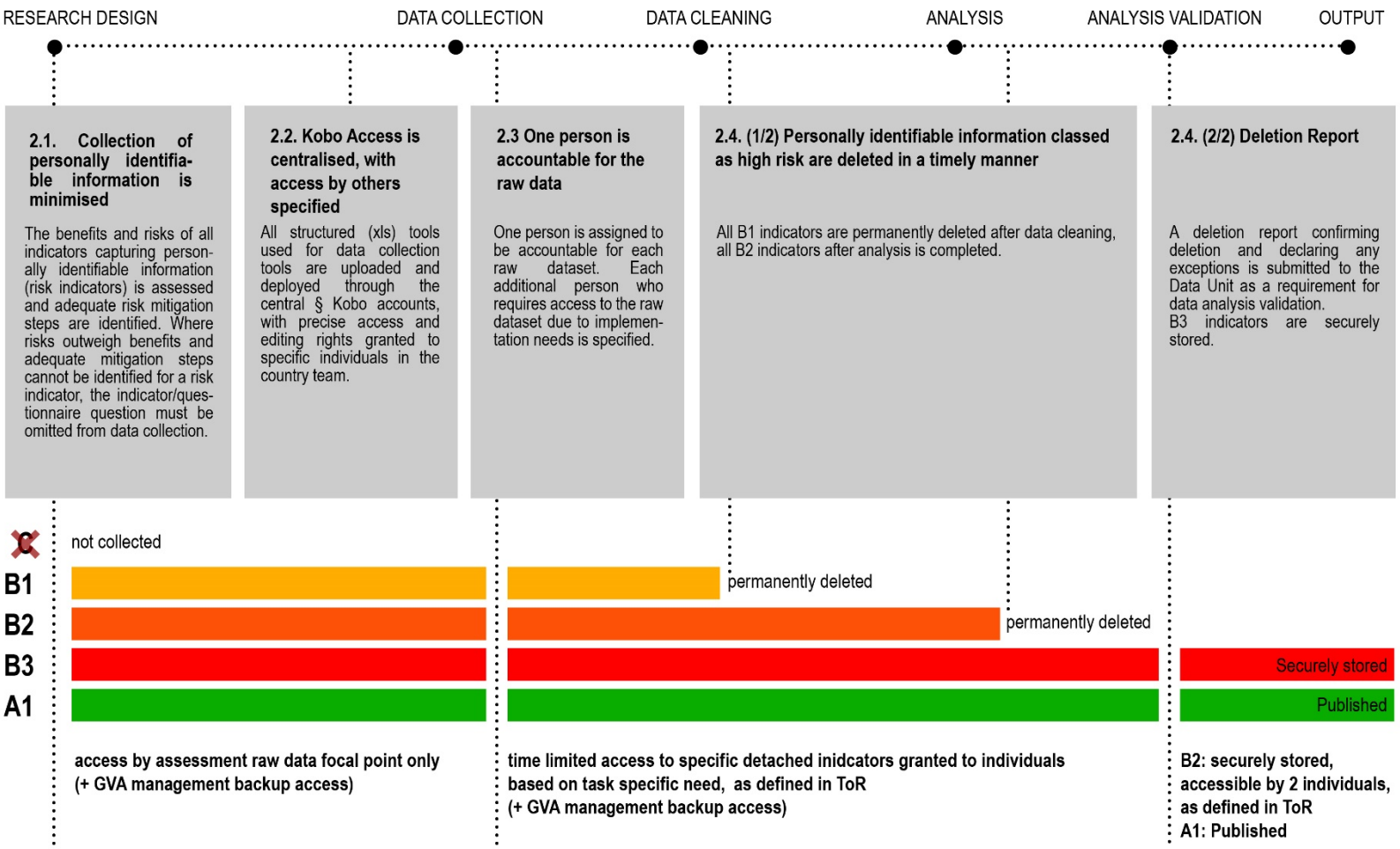
Summary of Country Team Action Points

To achieve the data protection goals outlined above, this section defines four overall action points for country teams:

1. **Ensure that the amount of personally identifiable information collected is minimised** as far as possible, as defined in the ToRs that are validated by the Research Department.
2. **Ensure that all structured (xls) tools used for data collection are deployed through the central IMPACT Kobo accounts** managed by the Research Department, and that precise access rights to any deployed Kobo tool and associated data are implemented, as defined in the ToRs that are validated by the Research Department.
3. **Ensure that one person is assigned to be accountable for each raw dataset** and that each additional person who requires access to the raw dataset due to implementation needs, is specified in the TORs that are validated by the Research Department.
4. **Ensure the timely deletion of all personally identifiable data,** as confirmed in the IMPACT Data Deletion report.

NB: Each of the above action points are outlined in detail in the sections below.

IMPACT Initiatives



2.1 Collection of personally identifiable information (risk indicators) is minimised

Requirement	**The benefits and risks of all indicators capturing personally identifiable information (risk indicators) is assessed and adequate risk mitigation steps are identified. Where risks outweigh benefits and adequate mitigation steps cannot be identified for a risk indicator, the indicator/questionnaire question must be omitted from data collection.**
Accountable	**CFP (Field); Research Design Unit (HQ)**

To minimise the amount of personally identifiable information we collect, the risk of all indicators collecting data that can be used to identify/locate persons, will be evaluated during the Research Design phase and explicitly validated by the Research Design unit, during the validation of TORs. Our data collection decisions often involve a trade-off between collecting the largest amount of useful information, and the risk involved in collecting it. For example, while GPS locations of the interviews are useful to ensure the quality of enumerators' work, interviewees' GPS location is often a high-risk indicator. At the research design stage, the benefits and risk need to be assessed by considering questions such as: Do I really need this information? If so why? Can I obtain this information in another way? If not, how big is the risk for the interviewees if the location of the interview becomes known, given the specific context? Can the risk be mitigated? If so, how? Does my need of the information justify the risk of collecting it?

Some indicators bear the same type of identification risk across assessments (e.g. phone numbers that can be used to identify and contact individuals directly), while others carry an identification risk in a specific context (e.g. HH gender and age composition at a sparsely populated location, which can be used to identify the household indirectly). Implications of disclosure will also vary according to the context (loss of privacy will always be an implication but may be accompanied by a risk of direct action being taken against a respondent in specific contexts). Typical risk indicators types of identification risks are listed in Table 1 below.

Table 1: Example Risk indicators list

1. Risk indicator	2. Type of identification risk
KI_phone number	Direct contact/identification of KI
GPS_latitude of HH GPS_longitude of HH	Direct identification of HH
Exact Age of all HH members	Indirect identification of HH using other information like locality, gender, income sources etc.

The indicator risk assessment is recorded in the TORs, where all risk indicators considered for the assessment, are classified according to levels of risk and their collection is explicitly approved by the HQ Research Department, upon validation of the TORs. The classification is done by the Research Design Unit, and is based on risk analysis → benefit analysis → mitigation analysis. Indicators that are classed as **C** i.e. where no sufficient benefit has been identified, must be removed and no corresponding data collected. A clear justification of benefits and application of the best possible mitigation strategy will usually mean an indicator can be collected. Indicators that are classed as **B1, B2** or **B3** i.e. for which the (mitigated) risk is deemed acceptable given the benefit of collecting the data, are approved for collection only if the associated mitigation measures are taken. An example of the Risk Assessment table where this process is captured within the TORs, is outlined below in Table 2 below and the definition for each class is detailed in the Process section that follows.

Table 2: Example Indicator Risk Assessment table, see Data Management section in the Terms of Reference or Methodology template

1. Risk indicator	2. Type of identification risk	3. Disclosure implications	4. Benefit	5. Class	6. Required mitigation
KI_phone number	Direct contact/Identification of KI	loss of privacy/potential target of armed actors	Follow up for data cleaning, where need to verify / request additional information	**B1**	Deleted directly after verification/cleaning
GPS_latitude GPS_longitude of KI interview	Direct identification of KI	loss of privacy/potential target of armed actors	To know where the KI is for this assessment (locality)	**C**	Benefit not identified: remove

In the table above, columns 1-4 are populated by the person responsible for completing the TORs for the assessment (as designated by the CFP). Once the TORs are submitted to the Research Department, the Research Design Unit identifies that Class and corresponding required mitigation (columns 5 and 6. In this example, a clear benefit has been identified AND risk can be minimised through a mitigation measure for the first indicator but not for the second. The second indicator is therefore classed as "C" and should be removed unless a clear benefit that justifies the risk is identified OR a sufficient mitigation can be applied that renders it a "B" level indicator.

NB: Level of risk associated with a specific indicator will vary depending on context. A clear outline of what the disclosure implications consists of in a specific context; along with clearly identified benefits of collecting the data, helps the Research Design Unit classify indicators more precisely.

Process:

1. Complete the Risk Assessment table in the Data Management section in the Terms of Reference / Methodology Note.

- If absolutely no information that potentially (directly or indirectly) allows identification of individuals is to be collected, fill out and sign the corresponding line in the ToRs.

- If information that potentially allows identification of individuals is to be collected, complete columns 1-4 in the Indicator Risk Assessment table in the TORs.
  - Ensure that the "2. Type of identification risk" column takes into account potential interaction between indicators (e.g. profession AND gender indicators may, when combined, enable identification of individual). Examples of types of identification risks are outlined in Table 1 above but this is not exhaustive and tailored to context; country teams should therefore consider all indicators critically based on the specific data collection context and declare any potential identification risks that can be identified.

2. The Research Design Unit reviews the entries provided and assigns a "5. Class" to each indicator, in addition to indicating what "6. Required mitigation" is to be undertaken. Classes include:

- **A** – no identification risk; no mitigation required
- **B** – benefit identified and risk is mitigated: implement proposed mitigation →
  - **B1**: deleted directly after verification/cleaning *(e.g. key informant phone number)*
  - **B2**: deleted after analysis/visualisation *(e.g. GPS coordinates)*
  - **B3**: stored *(e.g. GPS coordinates of dataset that will be used as a baseline for future data collection)*
- **C** – benefit does not justify the risk: remove/do not collect

Accountability structure

- **Responsible** for completing the Risk Assessment in the TORs → assigned by CFP
- **Responsible** for action points in risk assessment table → as defined under Required Mitigation.
- **Accountable** for completion of TORs and implementation of mitigations → Country Focal Point
- **Accountable/ Responsible** for validation of TORs → Research Design Unit

NB: Any changes to the terms outlined in the validated TORs regarding the above must be flagged to the Accountable HQ unit for validation.

2.2 Kobo access is centralised

Requirement	**All structured (xls) tools used for data collection are uploaded and deployed through the central IMPACT Kobo accounts, with precise access and editing rights granted to specific individuals in the country team. Each CFP creates an emergency assessment account on Kobo to use in exceptional cases (e.g. tool deployment over the weekend)**
Deadline	**For all data collection starting after 01.03.19**
Accountable	**CFP (Field); Data Unit (HQ)**

To formalise storage and access on Kobo, account structures are centralised by deploying all structured (xls) tools through the central IMPACT Kobo accounts, to which all collected data is submitted as a result. This way, we aim to ensure that no data Kobo remains on Kobo servers after an assessment has ended; to prevent Kobo servers with unknown or not individualised access structures; and to avoid situations of locked Kobo accounts (e.g. where the person with access leaves the organisation or forgets their login).

All Kobo tools will be deployed by the Data Unit through the central IMPACT Kobo accounts. From there, access is granted to field accounts following the access rights table defined in the ToRs. No collection can occur on any other Kobo accounts. Uploading tools is a formalised requirement for the Data Unit and specific increased capacity is provided to reflect that. The responsible focal point for Kobo tool uploads will be communicated when the requirement comes into effect.

Process:

1. Complete the Kobo Access Rights table in the Data Management section in the TORs/ Methodology Note

Table 3: Example Kobo Access Rights table to a deployed form, see Data Management section in the Terms of Reference or Methodology template

Kobo Access	Person	Account name
View Form		
View and Edit Form	Karl Phantom	Kphan
View Form and Submit Data	Enumerator 1	afg1803enumerators1
Download data	Jane Doe	Jdoe

NB: "Download data" access to a deployed form can only be awarded to one single individual. This person must be the same person that is accountable for the raw data (see section 2.3 below).

2. As soon as there is a very first draft for a Kobo tool ready – send the tool draft to the HQ Data Unit along with a link to / attached ToRs.

3. The Data Unit deploys the form and share access rights in line with the Access Rights table in the ToRs. This takes at most two working days; to avoid delays please ensure that you share a very rough early draft of the tool as early as possible to enable early deployment.

4. Piloting and data collection can start via the deployed form. The person with form editing rights can change /deploy updated versions of the form without any action from the Data Unit.

Kobo access rights when working with decentralised data cleaning
As noted above, "Download data" access to a deployed form can only be awarded to one single individual. This person must be the same person that is accountable for the resulting raw data (see section 2.3 below). However, in specific contexts, the research cycle process require multiple people to have access to the incoming raw data. This occurs, for instance, when multiple field coordinators use the same form, collect data in different parts of the country, and clean the data on site before sending the cleaned data to the country base.

Where it is impossible to avoid having multiple people access the raw data, each "section" of the assessment is treated as a separate assessment from the data protection perspective. In this example, the form is deployed **separately for each** field coordinator, essentially creating individual accounts for each field coordinator/section of the assessment. The field coordinator is assigned "Download data" access to the form deployed specifically for her, hence enabling cleaning to be conducted for the data submitted for the respective section, without exposing the field coordinator to the raw data coming in for other sections of the assessment. Each field coordinator then follows the risk indicator classifications and mitigation measures outlined in the risk assessment table in the TORs (see above) before submitting the cleaned dataset for her specific section of the assessment to the person who consolidates all sections into an overall dataset. Where this process is followed, the **Kobo Access Table (Table 3 above) should be duplicated in the TORs, with one table completed for each section of the assessment.**

Kobo access rights when working with partners
As a general rule, data is provided to partners the same way as it is made publicly available e.g. via HDX; meaning only the validated / clean datasets are shared. Where partners are collecting data for an assessment facilitated by REACH, they are simply given "View Form and Submit data" access (see table above) but do not generally require, and are thus not provided with, access to the incoming raw data.

However, for some assessments, raw data may need to be accessed by a partner during and/or after data collection. In this case, two access options are available, with Option A being preferred by IMPACT:

Option A: A MoU is signed with the partner, where IMPACT provide the partner with the raw dataset, (including uuids uniquely created for the partner), for which the partner is solely accountable to protect.
Option B: A MoU is signed with the partner, where the partner is given "Download data" access rights to the relevant Kobo account. The partner is accountable for the protection of any data the partner downloads from the Kobo account. If IMPACT is still required to clean / analyse data, the usual process outlined above is followed, with "Download data" access rights assigned to one designated IMPACT focal point.

Urgent tool deployment when the Data Unit is unavailable: Pre-uploaded blank forms
Given the volatility of the contexts we work in, under rare circumstances, tools need to be deployed unexpectedly/at extremely short notice.

1. Preparation:
    a. Each CFP may hold one emergency assessment kobo account (no tools are to be deployed directly from that account).
    b. All essential rights in the table above ("view and edit form" and "download data") for a blank tool from the GVA IMPACT account are shared only with the CFP. Further data submission rights are given to a with a predefined enumerator account.
2. Using the emergency account:
    1. Inform the Research Design and Data units, detailing the exceptional circumstances that call for the use of the emergency tool.
    2. Ensure that the preferred Kobo Access Rights table in the Data Management section in the TORs/ Methodology Note are filled out and sent to the Research Design and Data units, so that access rights can be adjusted as soon as the Data Unit is available again.

3. Replace the blank tool with Kobo's "replace project" and "redeploy" functionality with the actual tool for the assessment and conduct the assessment.
4. As soon as the Data Unit is available again, the access rights for the Kobo project are changed to those specified in the ToR.

Accountability structure

- **Responsible** for completing the Kobo Access Rights table in the TORs and submitting an early tool draft to Data Unit → assigned by CFP
- **Accountable** for completion of TORs and submission of tool draft → CFP
- **Accountable / Responsible** for deployment of tool and sharing of access rights according to TORs within two working days → Data Unit

NB: Any changes to the terms outlined in the validated TORs regarding the above must be flagged to the Accountable HQ unit for validation.

2.3 One person is accountable for the raw data, with access by others specified

Requirement	**One person is assigned to be accountable for each raw dataset. Each additional person who requires access to the raw dataset due to implementation needs is specified.**
Accountable	**CFP (Field); Research Design Unit (HQ)**

To minimise the risk of accidental leakage of personally identifiable information without disturbing current workflows, access to raw data is limited and specified. Risks of data leakages increase with the amount of people who keep personally identifiable data in their inbox, on their laptops or on any shared folder. This can be circumvented because personally identifiable information is often not necessary to everyone working on the assessment. For instance, names of Key Informants may be needed for follow up during data cleaning, but not for the analysis. Formalised access rights per dataset will limit access points to personally identifiable data to the minimum necessary. To achieve this, one person accountable for the raw data is defined in the ToRs. This person is responsible for detaching any personally identifiable indicators from the dataset before the data is shared internally or externally. They can only share the raw data with people specified in the ToRs as having a justified need to use the personally identified indicators. If anyone who is not specified in the ToRs gains access to any personally identifiable data, that is considered a data breach and the person accountable for the raw data is directly held accountable for the breach.

As outlined above, in specific contexts, multiple people need access to the raw data, for instance where multiple field coordinators deploy the same form, collect data in different parts of the country, and clean the data on site before sending the cleaned data to the country base. In this case, the Raw Data Access Rights table (Table 4) below must be duplicated in the TORs and completed for each "section" of the assessment. Following completion of the assessment, a deletion report (Annex 1) should be completed for each section of the assessment.

Process:

1. Complete the Raw Data Access Rights table in the Data Management section in the ToR / Methodology Note.
   a) Ensure the person accountable for the raw data set is identified.
   b) Ensure any additional individuals that need access to the raw data set (including personally identifiable indicators i.e. risk indicators) are identified. This may for example include GIS officers need access to data that contains GPS points in order to produce maps.
   c) Ensure the person trusted with the final backup copy of any raw data that cannot be deleted at the end of the assessment (indicator Class B.3 above).

Table 4 Example Raw Data Access Rights table, see Data Management section in the Terms of Reference or Methodology template

Raw Data Access	Reason	Person
Accountable	Accountable	Jane Doe
Access	GIS- choropleth maps using GPS points	Karl Phantom
Access	Final back-up of B.3 indicators	Alice Bob

NB: The person who is Accountable for the raw data must be the same person as the one who has "download data" rights in the Kobo access rights table (Table 3).

1. During data cleaning, the person Accountable for the raw dataset ensures that all personally identifiable information (risk indicators) is detached from the dataset and stored in one location only (where storage is permitted, as per TORs) plus one backup for B3 indicators – or deleted.

Risk indicators may be detached using a **pseudonymised approach**[4] as follows:

i) Add a variable populated by anonymous IDs, unique to each record, in the master dataset.
ii) Copy all risk indicators from the master dataset into a new file, along with the variable holding the new IDs → this new risk indicator dataset is stored by the person Accountable for the raw dataset only, no other copies exist.
iii) Delete all risk indicators from the master dataset.
iv) The anonymous ID variable that now exists in both the master dataset and the risk indicator dataset provides a linkage, should this be needed during the assessment, between the pseudonymised master dataset and the risk indicators.

2. If it is not essential to be able to associate the risk indicator by record with the original data, an **anonymised approach**[5] should be used instead. Here the data is detached as per the steps outlined above but no IDs are added and the order of the records are changed, thereby preventing the linkage between risk indicators and individual records.

3. The person Accountable for the raw dataset then shares the anonymised and/or pseudonymised data internally and with HQ Data Unit as per the data rights access table.

Accountability structure

- **Responsible** for completing the Raw Data Access table in the TORs → assigned by CFP
- **Responsible** for detaching risk indicators from dataset before sharing → as defined in TORs.
- **Accountable** for completion of TORs and assigning person accountable for raw dataset → CFP
- **Accountable/ Responsible** for validation of TORs → Research Design Unit

NB: Any changes to the terms outlined in the validated TORs regarding the above must be flagged to the Accountable HQ unit for validation.

[4] Enabling reattachment of detached data for each record.
[5] Preventing reattachment of detached data for each record.

2.4 Personally identifiable information classed as high risk are deleted in a timely manner

Requirement	**All data is deleted from Kobo after the assessment ends. All personally identifiable information (risk indicators) is deleted from other datasets. A deletion report is submitted to the Data Unit as a requirement for data analysis validation.**
Accountable	**Person accountable for raw dataset (Field); Data Unit (HQ)**

To further minimise the risk of accidental leakage of personally identifiable information, the conservation of personally identifiable data must be considerably reduced. Personally identifiable data should be kept beyond the date of the assessment only if it is absolutely necessary, as defined by the TORs validated by the HQ Research Design Unit. Risks of data leakages multiply with the amount of people who keep personally identifiable data in their inbox, on their laptops or on any shared folder, which is why we restrict the number of people with access to these risk indicators (see previous section 2.3). This can further be circumvented by removing personally identifiable information from the few locations where it remains, after an assessment has been completed. To formalise this step, a deletion report confirming deletion and declaring any exceptions, is submitted to the Data Unit as a requirement for data validation. The goal is for all communication to eventually occur over a secured server which will make it possible to track data sharing. New official communication channels will be shared as soon as they become available.

Process:

1. After the end of the assessment, the person responsible for the raw data confirms that all personally identifiable data has been deleted from all devices as specified in the TORs.
   a) This is done by filling out the deletion log (see Annex 1) and sharing it with the HQ Data Unit when data is submitted for validation.
   b) Any request for exceptional conservation of data is recorded in the deletion log.

2. Any person besides the accountable who had access to the raw data fills out a “mini deletion report” which confirms they deleted the copies of the raw data from their personal phones and computers and did not share it with anyone else, as specified in their responsibilities.

Accountability structure:

- **Responsible** for deleting all personally identifiable data → all staff with access to personally identifiable information as defined in the ToRs.
- **Accountable** for deletion of all personally identifiable data and submission of deletion report → person defined as accountable for the raw dataset in the ToRs.
- **Accountable** for validating deletion report → Data Unit.

NB: Any changes to the terms outlined in the validated TORs regarding the above must be flagged to the Accountable HQ unit for validation.

ANNEX 1: IMPACT Data Deletion Report

Basic Dataset information:

Research Cycle ID	
Assessment name	
Kobo server project name	
Dataset name	
Individual accountable for raw data (as per ToR):	
Individual holding final back-up copy of B3 indicators (as per ToR):	
Other individuals with access (as per ToR):	

Confirm deletion of information as defined in ToR:

Requirement	Signature
All data deleted from Kobo Server	
All B1 and B2 classified indicator data deleted irreversibly from all copies of the data, including raw data and cleaned data	
All B3 classified indicators detached from associated data and deleted irreversibly from all devices except two securely kept copies, accessible only to 1) individual accountable for raw data and 2) individual holding final back-up copy of B3 indicators	
No C classified indicators have been collected	
No B or C classified indicators have been recorded in cleaning logs	
No copies of B1, B2 or C classified data has been available to anyone other than the people listed in the ToR at any point	

Could all requirements be confirmed (signed for) in the table above?

☐ Yes
☐ No – please explain:

_____
_____
_____
_____

Signature : _____