Insight into Data Management

This training module is provided under the terms of the Creative Commons License: Attribution-NonCommercial-ShareAlike 4.0 International
This training module was developed by CartONG. Created in 2006, CartONG is a French H2H/Support NGO specialised in Information Management. Our goal is to put data at the service of humanitarian, development and social action projects.

We are dedicated to improving the quality and accountability of field activities, in particular through better needs assessments and monitoring and evaluation.

We act as a multidisciplinary resource and expertise centre, accompanying our partners’ strategies and operations.

Our staff and volunteers also support the community as a whole by producing documentation, building capacities and raising awareness on the technical, strategic and ethical challenges of digital technologies.
This training module was developed by CartONG, as part of the project “Strengthening Information Management within francophone CSOs”. This project, led by CartONG, is co-financed by the French Development Agency (AFD) over the 2019-2022 period.

For more information, please refer to the article published on our website or contact us by email.

This training module is supported by the French Development Agency (Agence Française de Développement - AFD). Nevertheless, the ideas and opinions presented in this toolbox do not necessarily represent those of AFD.
## TABLE DES MATIÈRES

- A few basics
- Data Management cycle & Data Protection
- Data collection
  - Defining data collection needs
  - Mobile data collection
  - Sampling
- Data analysis
- Data visualisation
- Data sharing
- Data retention, archiving, and deletion
- Useful resources
A few basics
What is the difference between data, information and knowledge?

- **Data** is a standalone element that has not been interpreted or put into context. **Raw data** refers to data that has not been cleaned, interpreted or analysed...

- Aggregation of data through processing results in **information**. Therefore, data **must be interpreted** to become information.

- If this information is **useful**, then it becomes **knowledge**.
How is data characterised?

- Data is generally classified according to **several criteria** that determine how it will be collected, analysed, visualised, shared and protected.

Refer to the [Data analysis toolbox](#).

[Diagram showing classification of data into quantitative and qualitative categories, with subcategories for continuous, discrete, nominal, and ordinal data.]
# Quantitative vs. Qualitative data

<table>
<thead>
<tr>
<th>Quantitative data</th>
<th>Qualitative data</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Example:</strong> “The number of people in a household”</td>
<td><strong>Example:</strong> “The main source of drinking water”</td>
</tr>
<tr>
<td><strong>Quantity</strong>, most of the time quantifiable and expressed in a specific unit (here the unit is “people”).</td>
<td><strong>Characteristic</strong> of an element, is not quantifiable.</td>
</tr>
<tr>
<td>Generally (but not exclusively) collected via quantitative surveys such as mobile or paper data collection forms.</td>
<td>Generally (but not exclusively) collected through formal or informal interviews, discussions and observations.</td>
</tr>
<tr>
<td><em>Example question:</em> How many people are there in your household?</td>
<td><em>Example question:</em> What is your household’s principal source of drinking water?</td>
</tr>
</tbody>
</table>
Other types of data

Audiovisual data (image, photo, audio recording, video, etc.) are data that are generally classified more as qualitative.

Example of a question for which the answer will be image: “Could you take a picture of your household’s main source of drinking water?”

Spatial and geographic data are not generally considered qualitative or quantitative data.
• **Personal data**: any information relating to a natural person, from which that person can be directly or indirectly identified.

Examples: name, picture, address, voice recording, phone number.

• **Personally Identifiable Information (PII) or “Direct Identifiers”**: specific personal data that can directly identify an individual.

Examples: name, address, date of birth, bank details, or ID/passport number of a respondent.

Not all personal data is PII

Examples: the “number of persons in the household” and the “household income” are examples of personal data that are not PII, as they can be applied to multiple different households.
• **Sensitive data**: personal data, which if disclosed or accessed without proper authorisation, may cause harm, lead to discrimination or repression against the data subject.

*Examples: may include data relating to health, race or ethnicity, or affiliation to religious and political groups.*

Sensitive data **may be harmful to an individual** or have a negative impact on an organisation’s ability to carry out its activities. **The sensitivity of the data** - as well as the appropriate safeguards - depends on the context.
Database and dataset

Raw data, including quantitative data, is typically represented as a data set or base.

- **Dataset:** structured data collection based on a specific data collection or survey.
  → Often: a table where each column of the table represents a variable (number of baskets received) and each row represents a unit of analysis (household).

```
<table>
<thead>
<tr>
<th>Name of the head of household</th>
<th>Food baskets received in November</th>
<th>Food baskets received in December</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tyvi</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>Ortha</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Zebka</td>
<td>3</td>
<td>3</td>
</tr>
</tbody>
</table>
```

- **Database:** an organized collection of data stored as multiple sets or datasets.
  → Generally stored/accessible in electronic form from a computer system allowing easy data access, manipulation and update.

*Example: Excel file*
Data Management cycle & Data Protection
This is the complete sequence of specific steps required for effective Information Management.

It describes the different processes needed to produce and use data.
Data protection: processes, systems and practices used to safeguard information from being lost, corrupted or accessed by unauthorised parties.

→ The right to data protection derives from the right to privacy (fundamental rights of individuals).

New technologies have allowed for easier and faster processing of personal data, which in turn leads to concerns about invasions of privacy.
Data protection therefore has implications at all stages of the data management cycle → **Responsible Data Management**.

The guiding principle is “**do no harm**” when producing and using data.
General principles of responsible data management in the sector

Minimisation of collected and analysed data

→ Reduce risk by collecting only what will be used.

Refer to the Webinar - “L’info-sobriété en pratique : comment l’appliquer à mon OSC?” (in French only).

Default protection, when collecting, storing, analysing, viewing, sharing and at the end of the project.

• Access rights and anonymization of the collected data
• Physical protection of data via access rights, sharing procedures (passwords, encryption, etc.).

Upholding the rights of survey respondents, when collecting, storing, analysing, viewing, and sharing their data; and at the end of the project.

• Informed consent
• Respect for the individual : respect for their time, experience, feelings and rights (access, rectification, deletion, etc.)
• Sharing of results and feedback to respondents on the use of their data
Data collection
Data collection: the process of assembling data via qualitative or quantitative methods.

There are several types of data collections:

- So-called quantitative survey methods: allow to extract statistical information about a situation and must therefore be representative of the surveyed population, either because they survey the entire target population, or because the sampling method used guarantees the representativeness of the outcomes.
- So-called qualitative survey methods more generally aim at highlighting the challenges, issues, and at providing details on the context to better understand a situation. There are several types of qualitative studies, such as: observation, semi-structured interviews, discussion groups.
## Data collection - Tools

<table>
<thead>
<tr>
<th>Collection Tools\Type of data to be gathered</th>
<th>Quantitative</th>
<th>Qualitative</th>
<th>Qualitative (media)</th>
<th>Geographic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Focus-group type discussions</td>
<td>Possible</td>
<td>Well-suited</td>
<td>Possible: audio recordings</td>
<td></td>
</tr>
<tr>
<td>Interviews with key informants (informal, formal, semi-structured, with interview grid, etc.)</td>
<td></td>
<td>Well-suited</td>
<td>Possible: audio recordings</td>
<td></td>
</tr>
<tr>
<td>Observations (structured, unstructured)</td>
<td>Mostly, because well-suited</td>
<td>Possible, but less suited and more limited than interviews</td>
<td>Possible, but less well-suited than on mobile</td>
<td></td>
</tr>
<tr>
<td>Data collection on paper (survey questionnaire)</td>
<td>Mostly, because well-suited</td>
<td>Possible, but less suited and more limited than interviews</td>
<td>Well-suited: GPS position, etc.</td>
<td></td>
</tr>
<tr>
<td>Data collection on mobile (survey questionnaire)</td>
<td>Mostly, because well-suited</td>
<td>Possible, but less suited and more limited than interviews</td>
<td>Well-suited: GPS position, etc.</td>
<td></td>
</tr>
</tbody>
</table>
Quality data enables better decision-making. Quality depends on:

1. **Relevance**: the extent to which data meets **user needs**.
2. **Accuracy**: how well the data describes what it is trying to **measure**. Data is uncorrupted when accurate (complete and clean).
3. **Timeliness**: how ‘up-to-date’ the data is, thereby impacting relevance and accuracy. **Temporality**: the fact that the data must be available when we want to use and share it. They are no longer useful when they arrive too late to inform decision-making. → A factor that plays an important role in data collection planning.
4. **Accessibility**: **ready availability** of data.
5. **Comparability**: **ability to compare** and analyse data in relation to other sources.
Defining data collection needs
Defining data collection needs

• Data collection and analysis improve knowledge and facilitate evidence-based decision-making. Program data are collected throughout the project cycle, primarily to:
  • Carry out a diagnosis
  • Enable implementation (operations)
  • Track results - indicators
  • Assess a program

• First and foremost, determine what data needs to be collected to answer the questions that underpin your analysis (diagnosis, monitoring, evaluation, etc.). To do this:

  1. List the already existing and available data.
  2. Assess the resources (human resources, budget) that you have to optimally fine-tune the survey.
  3. Formulate a Research Plan.
     → Refer to the Data analysis toolbox
Two types of data can thus be distinguished according to their sources.

- **Primary data**: collected directly by the field team to answer one or more research questions.
  
  *Examples: Monitoring and Evaluation (M&E) data collected by the M&E field teams.*

- **Secondary data**: available, pre-existing and collected by sources external to the field team.
  
  *Examples: reports from other actors, open data that can be analysed in more detail in relation to affected populations.*
Defining data collection needs - Existing and available data

• Please review the secondary data beforehand to:
  • Determine the research approach, formulate hypotheses and research questions
  • Define primary data requirements
  • Minimise the amount of data to collect
  • Save time on data collection and analysis
  • Check the plausibility of the outcomes produced by the analysis of the primary data, comparing them with the secondary data (triangulation).

Warning: Make sure that the source is reliable
  • Who produced this data?
  • What methodology was used? What is the level of detail of the calculations?
  • What is the accessibility of the raw data?
  • Etc.

Examples of shared databases in the humanitarian and development sectors:
• The Humanitarian Data Exchange (HDX): https://data.humdata.org/
• https://www.humanitarianresponse.info/
• https://reliefweb.int/
Defining data collection needs - Assessing resources

In order to **calibrate and plan your survey**, assess the following resources:

**Skills within your team (or your own):** Do you or your team have past experience with this kind of work? Do you know how to use the tools that you have available for the analysis?

**Time:** How much time do you have to spend managing this data?

**Knowledge of the subject matter:** In addition to considering your relevant practical skills, how familiar are you with the theoretical foundations of this topic? Do you understand the phenomena that you are investigating and the factors that are relevant to consider?

**Tools available for the analysis:** Do you have access to tools that will help you investigate your data? Are these tools something that you’re familiar with?

If you need to reduce the length of the survey so that the team can complete the collection within the budget, it is better to realise it at the beginning of your analysis process, rather than halfway through!
Developing a research plan will help you better understand the data you already have and the data that you really need.

<table>
<thead>
<tr>
<th>Research question</th>
<th>Indicator/variable</th>
<th>Questionnaire question</th>
<th>Data collection unit</th>
<th>Desired disaggregation</th>
<th>Analysis type</th>
</tr>
</thead>
<tbody>
<tr>
<td>Specific research question you are trying to answer</td>
<td>Indicator being calculated (from your logframe!)</td>
<td>Relevant question on the survey</td>
<td>Individual, household, community, etc.</td>
<td>Village, district, region, country, etc.</td>
<td>Frequency distribution, descriptive statistics, correlation, etc.</td>
</tr>
</tbody>
</table>

Add details on how data will be collected (collection tools), viewed, shared, stored...

Specify which will be sensitive, personal, shared or not, etc.
Mobile data collection
Mobile data collection: the *use of mobile technology* (smartphones/tablets) *to collect data*. This allows you to:

- **Improve the quality of the data**, the information and the analysis, and consequently, the decision-making.
- **Save time on analysis** and processing (even if the collection takes longer to prepare than a paper collection).

Many mobile data collection solutions exist, which will allow you to find one adapted to your needs.

→ Refer to the [Mobile data collection toolbox](#).
Mobile data collection

MDC compared to Paper-based data collection
Advantages and disadvantages

MDC Advantages
- Improved data quality and analytical capacities with integrated calculation, data validation constraints, skip logic, no handwriting issues, ...
- Integrated tool to collect different types of multimedia: GPS points, pictures, signatures, audio recordings, barcodes, ...
- Time saving: easier and faster analysis
- Better day-to-day monitoring of the data collection process and possibility of remote access to data
- Cost saving in the long run: less human resources (no need for data entry clerks)
- Centralized online archiving: reduced risk of data loss if well managed
- Interviewers go lighter in the field

MDC Disadvantages
- Longer preparation phase for designing, coding and testing the survey
- Can be intimidating and create distance with the person interviewed. Not appropriate in some contexts
- Can be a safety issue for staff in some contexts (targeting, theft, ...) Requires more technical skills in the project system implementation (longer training, adapted skills)
- Significant initial investment (purchase of mobile phones)
- Fragile devices and risk of technical issues
- Dependence on electricity for use and internet connexion for data synchronization
- Not adapted for qualitative surveys

Produced by CartONG/Terre des Hommes
Mobile data collection

MDC versus Paper-based data collection workflow
Approximative time spent at each step

<table>
<thead>
<tr>
<th>Methodological preparation &amp; paper form conception</th>
<th>MDC</th>
<th>Paper-based data collection</th>
</tr>
</thead>
<tbody>
<tr>
<td>Form coding</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Form testing</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Enumerators &amp; managers training</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pilot site</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Data collection</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Preparation of data entry forms</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Data entry</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Data cleaning</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Data analysis &amp; visualization</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Approximative time spent at each step *

Additional time needed for MDC
Additional time needed for paper-based data collection

* This illustration is a simplified representation, the associated time can vary significantly from data collection to data collection

Produced by CartONG/Terre des Hommes
Mobile data collection - Choosing your tool

- Is there a **specific tool** for your sector?
- Explore **comparative studies** of the tools and **criteria** examined (organisational management, user experience, data quality, data protection, case management, etc.).

**Data protection**: user rights management (limiting access), encryption, server location, etc.

### Typology of data collection tools

<table>
<thead>
<tr>
<th>Generic MDC tools usable across sectors</th>
<th>Sector-specific tools for a full project cycle management process</th>
</tr>
</thead>
<tbody>
<tr>
<td>Any generic MDC tool used for scoring (e.g., for nutrition)</td>
<td>RedRose used for post distribution monitoring</td>
</tr>
<tr>
<td>Any generic MDC tool used for WASH infrastructure follow-up</td>
<td>e-Wash tools adapted for other case management uses</td>
</tr>
<tr>
<td>[ ... ]*</td>
<td>[ ... ]*</td>
</tr>
</tbody>
</table>

**Generic tools** (that could be applied to all thematic sectors)

- Focus on **data collection** (survey or regular data); **Form-based**
- Can be used easily by **any program team, IM or M&E**
- **Supporting** an operational methodology; Can be used at **any step of the project cycle management**

**Different tools per thematic sector**

- For the digitalization of a **full business workflow**; Workflow-based
- Requests involvement of ICT4D team, strong thematic expert involvement
- Tool is at the center of the operational methodology

*Tools mentioned are examples to support the understanding of the illustration*

Produced by CartONG/Terre des Hommes
There are different tools to create a mobile data collection survey form:

<table>
<thead>
<tr>
<th>Tool</th>
<th>XLSForm (Excel)</th>
<th>Online forms creation interface (KoBoToolbox, ODK build)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ease-of-use</td>
<td>≈</td>
<td>✓</td>
</tr>
<tr>
<td>Internet connection not required</td>
<td>✓</td>
<td>✗</td>
</tr>
<tr>
<td>Advanced function/features</td>
<td>✓</td>
<td>✗</td>
</tr>
<tr>
<td>Compatibility</td>
<td>KoBoToolbox, ONA, SurveyCTO, etc.</td>
<td>KoBoToolbox, ONA, SurveyCTO, etc.</td>
</tr>
</tbody>
</table>

This choice depends on:
- The skills **of your team**
- The **complexity of the form**.
Online interface

Example: KoBoToolbox

→ Compatible with other ODK world standard tools.
→ Accessible for beginners.
→ See the tutorials for this tool in the Mobile Data Collection toolbox.

XLSForm
The research plan → helps to draft the questions most appropriate to the information sought and the desired indicators.

Keep in mind: the order of the questions, their wording, or the length of the questionnaire can lead to biases.

There are different types of questions:

- **Open question**: the respondent can answer freely, speaking for him or herself and not by choosing from among proposals.

- **Closed question**: the respondent must choose one of the proposed answers, or is restricted in his or her answer.
Mobile data collection - Types of questions

- Mobile survey questionnaires allow the use of more or less complex types of questions. These are not all questions as such but rather data inputs. Examples include the following:

<table>
<thead>
<tr>
<th>Type of question</th>
<th>Answer to enter</th>
</tr>
</thead>
<tbody>
<tr>
<td>integer</td>
<td>Integer input</td>
</tr>
<tr>
<td>text</td>
<td>Free text response.</td>
</tr>
<tr>
<td>select_one [options]</td>
<td>Single choice question: only one answer can be selected.</td>
</tr>
<tr>
<td>select_multiple [options]</td>
<td>Multiple choice question: multiple answers can be selected.</td>
</tr>
<tr>
<td>select_one_from_file [file]</td>
<td>Single choice from a file: only one answer can be selected.</td>
</tr>
<tr>
<td>note</td>
<td>Displays a note on the screen, requires no input. Can also display media (an audio, video, or picture file), see XLSForm.org for more information).</td>
</tr>
<tr>
<td>geopoint</td>
<td>Collection of a single GPS coordinate. An accuracy threshold can be added (see XLSForm.org).</td>
</tr>
<tr>
<td>dateTime</td>
<td>Accepts the input of a date and time.</td>
</tr>
<tr>
<td>image</td>
<td>Takes a picture.</td>
</tr>
<tr>
<td>audio</td>
<td>Makes an audio recording.</td>
</tr>
<tr>
<td>barcode</td>
<td>Scans a barcode, which requires the installation of the barcode scanner application.</td>
</tr>
<tr>
<td>calculate</td>
<td>Performs a calculation</td>
</tr>
</tbody>
</table>

- To learn more about the types of questions that can be asked in mobile data collection, refer to the “Standard Questions” tab of CartONG’s XLSForm memo.
The use of mobile data collection improves the quality of raw data by harmonising it and avoiding non-responses through logic compelling responses, such as:

**Skip logic:** makes it possible to display (or not) certain questions depending on the answer to a previous question. *Example: The question “Are you pregnant” only appears if the person answered “woman” to the question “What is your gender?”.*

**Constraints or validation criteria:** allows you to limit the response range and avoid recording invalid responses. *Example: regarding an age, limit the response range to between 0 and 110.*

**Cascading list:** a form construct that allows you to filter the possible answers to a question based on the entry of previous questions. *Example: displays only French cities, if in the previous question, you indicated living in France.*
Sampling
Sampling

A population is the entire group that you want to study, while a sample is just a subset of that population. For an analysis to be correct, the sample surveyed must be representative of the population you wish to study.

Example: in a KAP survey (Knowledge, Attitudes, Practices), you interview a portion of the people living in the commune (= sample of the population), but not all the inhabitants (= population).

To ensure that the estimates derived from the sample are representative (not biased), it is important to:

• Calculate descriptive statistics on the structure of survey respondents (age, sex, commune, etc.) to ensure that there are no large differences with the structure of the population.

• Specify the sources in all statistical elements (graphs, tables) so that the reader understands the scope of the survey.
Defining the sample

In order to define your sample, you must:

1. Know:
   - The size of the population (example: number of individuals in a city)
   - The sampling unit (example: individuals, households).

2. Choose the sampling method:

<table>
<thead>
<tr>
<th>Main non-random methods</th>
<th>Main random methods*</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Convenience</td>
<td>• Simple random sampling</td>
</tr>
<tr>
<td>• Reasoned choice</td>
<td>• Systematic random sampling</td>
</tr>
<tr>
<td>• Quotas</td>
<td>• Cluster sampling</td>
</tr>
<tr>
<td>• Snowball</td>
<td></td>
</tr>
<tr>
<td>• Auto-select</td>
<td></td>
</tr>
</tbody>
</table>

Limits the representativeness of the sample, which may bias the observation but respond to research choice.

Provides statistically representative results but is not always feasible because it requires sufficient population information to calculate the sample size.

*Random methods require the calculation of the sample size, e.g. the number of units to be surveyed, so that the sample is representative of the entire target population. You can calculate the sample size using certain sites, such as this one - available in English. Pay attention to the margin of error and confidence level you can tolerate. You can also find more information about this in the Data Analysis Toolbox.
Informed consent - Focus

• It is important to adapt and readjust the way you seek consent and what underlies it for each data collection (based on what you plan to do with the data, on population characteristics, etc.). Here are a few things to keep in mind:

  • Easy to understand! Also think about the translation.
  • Mention the fact that a mobile device will be used for collection if this is the case (or allow them to refuse the fact that a mobile device will be used if it makes them uncomfortable).
  • Obtaining a child’s consent is different from obtaining an adult’s: you will need to make sure that you also have the guardian’s consent.
  • Use wording like “I understand that…”: ensures that you obtain informed consent and that the interviewee is aware of their rights and the issues related to the collection of data that concerns them.
Data analysis
Data analysis: the process of applying techniques to data in order to discover useful information and support decision-making. After data collection, each separate piece of data can be brought together (aggregated or disaggregated), and through data analysis techniques provide more useful information.

→ The information that is discovered through the analysis can then be used to facilitate decision-making.

• The most widely used tools in the industry are:

  - Spreadsheets
  - Business intelligence
  - Statistics
  - Geographical analyses
Data analysis

Example of data analysis with Excel (descriptive statistics and PivotTable):

Measures of central tendency and variance of the ages of the sample population

<table>
<thead>
<tr>
<th>Quartile</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>10</td>
</tr>
<tr>
<td>Median</td>
<td>21</td>
</tr>
<tr>
<td>3</td>
<td>38</td>
</tr>
<tr>
<td>IQR</td>
<td>28</td>
</tr>
<tr>
<td>Lower limit</td>
<td>-32</td>
</tr>
<tr>
<td>Upper limit</td>
<td>80</td>
</tr>
</tbody>
</table>

* Using the QUARTILE function of the age variable.

Mean age of head of household by region

<table>
<thead>
<tr>
<th>Row Labels</th>
<th>Average of HHHAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Region 1</td>
<td>45,3</td>
</tr>
<tr>
<td>Region 2</td>
<td>45,8</td>
</tr>
<tr>
<td>Region 3</td>
<td>42,5</td>
</tr>
<tr>
<td>Grand Total</td>
<td>44,6</td>
</tr>
</tbody>
</table>

* Using Pivot Tables with showing the value as Average in the value field settings
Data cleaning

- Data cleaning is a necessary step just prior to analysis to ensure an up-to-date database (data may be old, but there should be no errors in the data: consistency, logic, etc.) that can be analysed without risk of misleading those who will use the information from the analysis.

- Database cleaning is **predominantly a logical process** that notably entails data **consistency analyses** and **triangulation** with other available information.

- For more information about this, refer to the Data Analysis Toolbox.
- To learn how to clean your data in Excel, see the Excel training module and the Excel toolbox.
Data cleaning

1. **Prior to any edits**, export your data and **create a backup copy of the original raw data** in a separate workbook.

2. **Format your data** in a readable database: display the data in tabular form with **clear column names** with **1 row per entry and no merged cells**!

3. **Spot errors and inconsistencies** by screening the dataset (duplicates, errors, outliers, missing data, etc.).

4. Check **the consistency and logic between the answers** (per row).

5. Take action (correction, deletion or modification of the data).

6. **Code or group certain variables** to facilitate analysis.

7. Before, during and after each of these steps, you should review the data to confirm that you have made no **unexpected changes** (visually or by using filters, etc.)!

---

When you clean your data, log changes to your dataset in a “**change log**”.
Triangulation consists in comparing the obtained data and results with other data sources and other results. This makes it possible to check the plausibility of the collected data so as to better grasp what angle of approach to adopt when analysing and representing the data, and thus inform the decision-making as best as possible.

In practice, this involves:

• Crossing the data collected using different methods (qualitative and quantitative).
• Crossing with external secondary data.
A bias is a systematic error of conscious or unconscious interpretation that can distort the reliability of the analysis results.

- **Biases are multiple** and can occur at different stages (from data collection to analysis and visualisation of results).

  *Examples of biases: cultural biases, related to personal beliefs; gender bias; cognitive bias; tool format bias, etc.*

- Whilst they are sometimes unavoidable, it is important to acknowledge them in order to understand and recognise the limits of analysis.

- It is therefore essential to understand which methods of data collection and analysis are appropriate according to the context and the target population. Also pay attention to how you ask questions, the language you use, as well as to the need to create a climate of trust.
### Bias

<table>
<thead>
<tr>
<th>Step in the data cycle</th>
<th>Bias examples</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Sampling</strong></td>
<td>A non-random sample can lead to an over-representation of respondents with similar characteristics.</td>
</tr>
<tr>
<td><strong>Survey design</strong></td>
<td>A poorly formulated, poorly translated question, or a particular set of questions can mislead or influence the answers. The choice of a question type, certain categories of answers or the length of the questionnaire as well.</td>
</tr>
<tr>
<td><strong>Collection</strong></td>
<td>A question incorrectly asked or incorrectly explained by the investigator, the way the questionnaire is communicated (on paper, with a phone, remotely by phone, etc.), or the investigator’s gender may influence the answer.</td>
</tr>
<tr>
<td><strong>Collection</strong></td>
<td>Lack of confidence or misunderstanding of the purpose of the survey (which can result in assistance being provided) can lead individuals to not respond or give false answers.</td>
</tr>
<tr>
<td><strong>Analysis and visualisation</strong></td>
<td>Cognitive bias affect the reasoning that guides our analysis, which then become inherent when we want to draw conclusions from our data. A statistical representation or choice (of an indicator, specific categories) can influence the visualisation of a trend. → To learn more about analysis biases, you can consult this technical brief.</td>
</tr>
</tbody>
</table>
Data visualisation
Data visualisation affects decision-making processes by allowing more people to understand and interpret data more quickly. It provides insight into what information means through a clear visual representation. This helps the human brain to apprehend data more “naturally”, thereby identifying patterns, trends, outliers, etc.

Some data can be more complicated to represent than others; to help you through this, ask yourself:

• What am I trying to say?
• Who am I trying to say it to?

To learn more data visualisation in Excel, consult the training module on Excel.
How to represent **quantitative data**?

**Multiple statistical representations** are possible (graphs, charts, table, etc.). However, it is important to carefully select a graph that best applies to your data.

**Comparison**

- Simple multiple variables
- Simple over time: visualise a tendency
- Multiple (series and categories) and over time
Data visualisation

Composition

- Part of a whole
- Several parts of a whole

Distribution

- Single variable: Classes
  *Examples: number of children, age groups*
- Correlation: Multiple variables
Data visualisation

Example of data visualisation analysis with Excel (pie chart):

To learn how to represent your data in Excel, see the Excel training module.
Data visualisation

How to represent qualitative data?
• Qualitative data may, for example, be represented in the form of diagrams or maps.

• Some qualitative data can be made quantitative. For example, it is to turn them into quantifiable items by creating categories, during collection or analysis.

  Example: 30% of respondents responded “moderately satisfied” to the question on latrine cleanliness.

How to represent audiovisual data?
• Categorisation can also facilitate the representation of audiovisual data. However, the latter must be analysed on a case-by-case basis to determine their best representation.
Data visualisation

How to represent geographical data?

They can be represented in the form of geographical and statistical maps.

Note: the distinction made here between these 2 map categories is abusive - a map is, by definition, geographical - but it allows to magnify the trait for educational purposes to understand the possibilities of visualisation based on your data.

uMAP - Map of CartONG members
Static maps

Heat diagram according to the concentration and value of the elements

Source: IGN
Data sharing
Data Sharing - Principles to apply

• Identify **the risks associated with specific data** and classify the data as strictly **confidential**, which can be shared in a restricted way, which can be **public**.
  • **Anonymize the data.** Pay attention to **aggregated data** and make sure that it does not allow **re-identification**.
  • **Respect what has been consented to in informed consent.**

**How, what and to whom to share?**
• All or part of certain databases...
  ✓ ... in **free or limited access** on sites such as HDX or other platforms with **terms and conditions of use** governing the sharing and use of data.
  ✓ ... **bilaterally** to another organisation by drafting a **Data Sharing Agreement** (DSA) that governs the sharing and use of data.
• **Minimise sharing of sensitive or personal data** to minimise risk.
Data retention, archiving, and deletion
1. **Map your data** to find out through where the data you produced for your analysis went. Example:

<table>
<thead>
<tr>
<th>Step in the data cycle</th>
<th>Storage</th>
<th>Users/Access</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Collection</td>
<td>• Peripherals: USB flash drive, external</td>
<td>• Enumerator</td>
</tr>
<tr>
<td>• Analysis</td>
<td>hard drive</td>
<td>• Analyst</td>
</tr>
<tr>
<td>• Sharing</td>
<td>• Server</td>
<td>• Partner</td>
</tr>
<tr>
<td>• Archiving</td>
<td>• Cloud</td>
<td>• Audience</td>
</tr>
<tr>
<td></td>
<td>• Phone, tablet</td>
<td></td>
</tr>
</tbody>
</table>

The [data flow mapping](#) tool can be useful for this.
2. Determine **which data is retained and for how long**, which data is archived and which data can be deleted. Depending on the type and purpose of the data, one option may be more relevant and appropriate.

- Particularly consider personal or sensitive data
- Respect what has been consented to

<table>
<thead>
<tr>
<th>Retention</th>
<th>Archiving</th>
<th>Deletion</th>
</tr>
</thead>
<tbody>
<tr>
<td>For how long?</td>
<td>Choose an appropriate storage device.</td>
<td>Make sure you have deleted on each storage device.</td>
</tr>
<tr>
<td>Manage and/or restrict access (consider the presence of personal and/or sensitive data).</td>
<td>Manage and/or restrict access (consider the presence of personal and/or sensitive data).</td>
<td>Keep in mind that a deleted file can sometimes still be recovered (in the computer trash, for example).</td>
</tr>
<tr>
<td>Choose an appropriate storage device.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Useful resources
Resources developed by CartONG

- Excel Toolbox, developed by CartONG

- Data analysis toolbox, developed par CartONG

- Mobile Data Collection Toolbox

- Self-assessment tool in program data management
  including the data flow mapping tool:
Mobile data collection (MDC)


- XLS Form Tutorials: https://xlsform.org/en/
Bias Data Management


Open Data - Data Protection

- Shared databases in the humanitarian and development sectors:
  - The Humanitarian Data Exchange (HDX): https://data.humdata.org/
  - https://www.humanitarianresponse.info/
  - https://reliefweb.int/

- Tutorial “Becoming RAD: How to Retain, Archive and Dispose of data responsibly” by the Engine Room:

- “L’info-sobriété en pratique : comment l’appliquer à mon OSC?” webinar – in French only:
  https://www.youtube.com/watch?v=z2wFhn34j3M&t=8s

- Data Protection Toolbox:
CartONG
23 Boulevard du Musée
73000 Chambéry (France)
+33 (0)4 79 26 28 82

info@cartong.org
www.cartong.org

All photos: © CartONG
Icons used made by freepik from www.flaticon.com