Responsible Data Governance
for Monitoring and Evaluation
in the African Context

DECEMBER 2021

02.

GUIDANCE ON RESPONSIBLE DATA GOVERNANCE IN M&E

MERL Tech

UNIVERSITY OF THE WITWATERSRAND, JOHANNESBURG

clear
Centers for Learning on Evaluation and Results
ANGLOPHONE AFRICA

Responsible Data Governance for Monitoring and Evaluation in the African Context

02 GUIDANCE FOR RESPONSIBLE DATA GOVERNANCE IN M&E

Authors: Jerusha Govender, Monet Durieux, Ilse Flink, Desiree Jason, Mark Irura, Rachel Sibande and Linda Raftree

ACKNOWLEDGEMENTS

We acknowledge the working group members who contributed to the framing, writing and revisions of this document. Without their ongoing support and work, this document would not have been possible.

- Monet Durieux, Senior Associate, Genesis Analytics, South Africa.
- Ilse Flink, Researcher, VVOB, Rwanda.
- Jerusha Govender, Co-founder and Director, Data Innovators, South Africa.
- Mark Irura, Technical Advisor, GIZ Kenya.
- Desiree Jason, Director for Policy and Programme Evaluation, National Department of Social Development, South Africa.
- Jessica Musila, Founder and Lead Consultant, Shomer Consulting, Kenya.
- Brian Tshuma, Partner (Data & Capital Markets) at Deme Attorneys, Zimbabwe.
- Rachel Sibande, Programme Director, Data for Development, Digital Impact Alliance, Malawi.

We thank Talitha Hlaka, Communications Officer, CLEAR-AA, South Africa, Linda Raftree, Independent Consultant and Co-Founder of MERL Tech, United States of America, and Dugan Fraser formerly of CLEAR-AA and currently at the Global Evaluation Initiative (GEI), South Africa, for guiding and supporting this process from start to completion.

Many thanks also to Dr Candice Morkel, Director of CLEAR-AA, and Steven Masvaure, Senior Monitoring and Evaluation Technical Specialist at CLEAR-AA for funding this work and providing feedback on the various versions of the document.

Summary of the publication: This publication includes two sections. The first section gives an overview of data governance and how it relates to the field of monitoring and evaluation (M&E), with a focus on M&E in the African context. The second section provides guidance on how M&E practitioners can more responsibly manage data in their practice. The publication was developed and compiled by a group of M&E professionals and data privacy experts over the course of 2020 and 2021.

Suggested citation: Centre for Learning on Evaluation and Results – Anglophone Africa (CLEAR-AA) and MERL Tech (2021) "Responsible Data Governance for Monitoring and Evaluation in the African Context. Part 1: Overview of Data Governance and Part 2: Guidance for Responsible Data Governance in Monitoring and Evaluation." Faculty of Commerce, Law and Management | University of the Witwatersrand, Johannesburg, South Africa.

TABLE OF CONTENTS

INTRODUCTION

In this section of the Responsible Data in M&E Guide, we describe key practical aspects of responsible data governance for M&E. The guidelines are accompanied by tip sheets, and comprehensive reviews on each of the specific areas. We offer added orientation on specific issues and challenges related to responsible data governance in M&E.

Within the M&E Cycle (see Figure 1 for a simple model of the cycle), data collection and analysis take place at different stages, for example:

- During diagnosis or needs assessment and consultation;
- When designing and planning a programme;
- During the baseline or mid-line;
- During implementation when conducting monitoring and when programme staff are collecting data during programme operations (sign-in sheets for workshops, or health data, for example);
- During the evaluation stage

Stages in the Data Lifecycle

The Data Lifecycle illustrates the journey of data throughout each stage of the M&E activity, from designing the overall exercise, data collection or acquisition, analysis, sharing or using, through to the point that it is retained or destroyed. Each data-related process follows its own Data Lifecycle**.**

There are seven stages in the Data Lifecycle (see Figure 2).

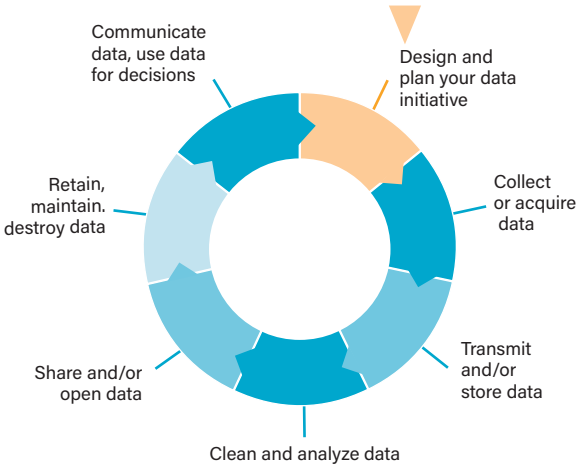
☑ Responsible data governance is key at every stage, as outlined below:

1. Design and plan your data-related process

- ☐ Develop Terms of Reference, identify your partners.
- ☐ List the various data-related processes needed for your programme.
- ☐ Outline the rules and roles for stakeholder participation in design and planning.
- ☐ Plan and budget for how you will manage data from beginning to end and how long you will retain data (National Data Privacy Legislation may stipulate for how long personal or sensitive data can or should be held).
- ☐ Determine what laws govern data collection and/or lawful basis for data collection.
- ☐ Decide how you will obtain consent.
- ☐ Obtain ethical clearance (if needed).
- ☐ Put data sharing and/or data processing agreements in place to protect your organisation and the people whose data is being collected.
- ☐ Conduct a risk-benefit assessment and a risk mitigation plan to check whether your plan will keep data and the people who provide it safe and secure.

Figure 1. Monitoring and Evaluation Cycle

Figure 2. Data Life Cycle



2. Collect or acquire data

- [ ] Be aware of the potential risks of collecting or acquiring data from or about vulnerable people (prepare a risk mitigation plan).
- [ ] Be sure you have a clear reason for collecting every data point that you plan to capture – do not collect personal or sensitive information unless you have a valid and justified purpose for doing so.
- [ ] Define your tools and procedures for data collection.
- [ ] Document your consent process or other legal processes.
- [ ] Provide a clear explanation to data subjects why you are collecting data, what you will do with it, who you will share it with, and how long you will keep it.
- [ ] Ensure that you have clear redress mechanisms for any complaints, and ways to address any requests from data subjects to correct their data, remove their data from systems, or withdraw consent for data processing.

3. Transmit and/or store your data

- [ ] Securely transfer the data to the place where it will be analysed or stored.
- [ ] Store the data safely on a secure laptop, organisation server or database, or in the cloud.
- [ ] Set access limits and strong security for your systems so that the data is kept safe from any unauthorised access, changes, downloading, deletion, or other data breaches or hacks (also see #5 on 'data sharing').
- [ ] Ensure a clear and transparent process for storing consent documentation.

4. Clean and analyse your data

- [ ] Clean the data and conduct data quality assurance.
- [ ] Tag or describe the data properly so that it can be located in the system if and when necessary.
- [ ] De-identify, anonymise, or aggregate the data to protect individuals from being identified.
- [ ] Conduct analysis on the data.
- [ ] Document the processes that have been performed on the data.

5. Share and/or open your data

- [ ] Before sharing data, review your consent documentation (see Step 2). Do not share personal or sensitive data without consent from the data subjects.
- [ ] Use safe sharing practices and limit who can access the data.
- [ ] Ensure that data sharing or data processing agreements with clear accountability protocols are in place.

6. Visualise and/or communicate data or use data for making decisions

- [ ] Decide how best to communicate insights to the specific target audiences.
- [ ] Disseminate and communicate the information and knowledge gained in ways that enable better decisions, better services, and better learning.
- [ ] Ensure that data is properly anonymised or de-identified to avoid harm.

7. Retain, maintain, and destroy your data after the project or initiative is complete

- [ ] Review how long data will be held at the design stage (see Step 1) and plan accordingly.
- [ ] If possible, aggregate or anonymise data so that it is less likely to be linked to an individual.
- [ ] Delete data as soon as possible to reduce liability.
- [ ] Ensure that data and devices are closed down when a programme is complete.

STAGE 1

Designing and planning

SNAPSHOT

This section offers guidance on planning and designing for the full Data Lifecycle, including legal aspects, data sharing plans, risk-benefit assessment, and budgeting.

At the design and planning stage you will focus on mapping out processes and resources for the entire Data Lifecycle. Design and planning are critical for ensuring that any data collected is fit for purpose, meets the internal project requirements, and fulfils any standards and legal requirements for data protection and ethical data collection. The first stage of the Data Lifecycle is related to the M&E cycle stages of diagnosis and consultation, and programme planning and design.

In some ways, the Design and Planning phase is like an executive summary. While this stage involves strategising at a higher level, it relies on the plans and details from the other Data Lifecycle stages. So, while you will start with Stage 1, you will plan all the other stages (2 to 7) in the Data Lifecycle and then return to complete the Design and Planning (Stage 1). Stage 1 is applicable to independent research, M&E, and operational data processes. The guidelines in this section apply specifically to M&E, but the concepts and principles can be extrapolated to other fields and practices.

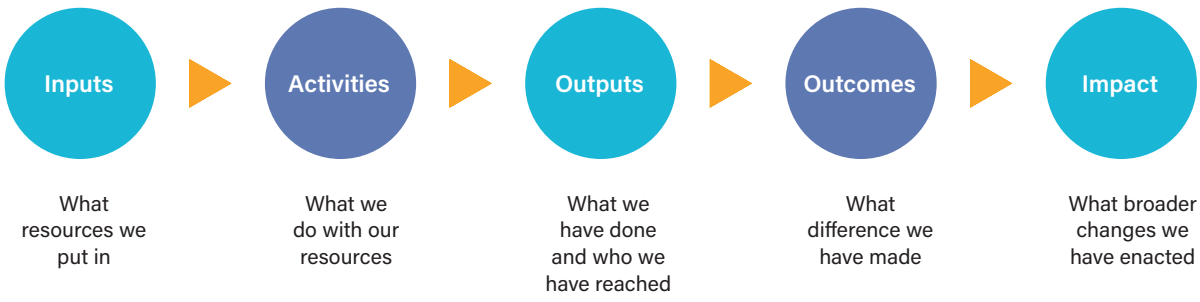
Guidelines for designing and planning in M&E

1.1 Outline the high-level data-related processes needed for your M&E effort

A results-based management approach guides the Theory of Change and M&E Framework. The Theory of Change (TOC) outlines how you plan to achieve the desired impact in a particular context, programme, or intervention. The impact pathways refer to how the intended programme activities will lead to the intended impacts.

Data collection activities should derive from the Theory of Change (TOC), the associated list of indicators for tracking the achievement of outputs, outcomes and impacts, and the data-related activities required at the distinct stages of the M&E cycle. These activities are research and consultation, needs assessments, baseline and midline data collection (e.g. for a midline evaluation), and ongoing M&E (See Figure 3). At the Design and Planning stage of the Data Lifecycle you should sketch out your high-level M&E framework.

Figure 3



See Stage 2 of the Data Lifecycle: Data Collection and Acquisition for more detail about what data you will collect from whom, how, and when.

1.2 Outline the rules and roles for stakeholder participation in your M&E effort

Clarity is a key aspect of governance and applies to participation, decision making, and accountability. In terms of data governance, it is important to decide who will be involved in various roles, whose voices will be included, and who has the power for decision making. This is especially important when designing and planning M&E data collection as the various stakeholders will have different interests and priorities in terms of their needs. It is particularly important to ensure that the voices of the programme beneficiaries are acknowledged throughout the data collection process.

When assessing project success, it is necessary to involve the various stakeholders to decide what their own unique definition of success would be for a particular project. For example, how a country defines success will be quite different from how a citizen defines it. Stakeholder mapping is a useful tool for identifying stakeholders and determining how they interact with one another.

After identifying potential stakeholders, consider possible avenues for their participation, how their needs will be met, and how involved they will be in interpreting the data and in decision making. Participation can be in the form of focus group discussions, key informant Interviews and surveys, amongst others.

Stakeholders are the individuals, groups, or organisations which are affected by an effort (in other words, those that have a 'stake' in it) or that have the power to influence decisions.

See Stage 6 for guidance on communication which may support stakeholder engagement and inform requirements for the steps in Stage 1.

1.3 Identify the laws that govern your data collection and use and/or decide on the lawful basis for data collection that fits your situation

Many African (and other countries) have data privacy laws or are developing or updating them to adapt to new realities of digital societies. These laws are often based on the European Union's General Data Protection Regulation (GDPR). Many bilateral donors are bound by the GDPR and this may extend to those they fund. It is important to be aware of national privacy laws as well as any regional frameworks or laws applicable to donors or partners if you are working in partnership.

Privacy laws use different terminology to describe types of data, and different national regulations may cover different aspects of data. It is important to understand certain aspects of these laws – outlined below. Even in the absence of a specific national privacy law, there may be other laws that need to be adhered to. For example, there may be electronic marketing laws, laws pertaining to research and the establishment of internal review boards, laws governing non-governmental organisations, laws that protect vulnerable individuals and groups, and laws related to protection of financial data or consumer data.

See Section 1 on data governance in Africa for further information on global data privacy, legislation and African-centred data governance.

Did you know?

According to UNCTAD, 28 (52%) African countries have some type of data privacy legislation, and 9 (17%) have introduced draft legislation. Some countries are developing Personal Data Protection Bills or Acts and/or are moving towards comprehensive data privacy regulation (see South Africa's updated Protection of Personal Information Act (POPIA) of 2020; Nigeria's Data Protection Bill of 2020; Kenya's Data Protection Act 2019; and Uganda's Data Protection and Privacy Act of 2019). Some have older data protection regulations, for example, Ghana's Data Protection Act of 2012, and others have laws that touch on aspects of data privacy and protection, such as telecommunications laws, ICT laws, cybercrime laws, and electronic communications laws, and/or consumer protection laws, but have not enacted a single overarching data privacy regulation.

Globally, many of the emerging data privacy laws, including those in the African context, are modelled on the European Union's General Data Protection Regulation of 2018. India's Personal Data Protection Bill of 2019 is another model that some countries find applicable. The United States does not currently have a comprehensive data privacy law, though it does have sectoral data privacy regulations (covering areas such as children's data, health data, educational data, and financial data). State-level laws are becoming more common (California, Virginia and others have introduced or passed bills by 2021). China has recently passed its data privacy/security law aimed at protecting data-related rights of individuals and organisations, ensuring the lawful, free and orderly flow of data and improving the digital economy.

At this step in the Data Lifecycle, review data privacy legislation to understand what types of data the law covers and how you need to address aspects such as:

- Personal, sensitive, and other types of data as defined in the law.
- Consent or other lawful bases or restrictions for data collection and processing.
- Requirements related to collection and use of data from children or other sensitive categories of data. In some cases, it may be necessary to establish research ethics boards.
- Data sharing and access to data, including any mandates for government access to data.
- Responsibilities for those controlling and processing data.
- Transfer of data across borders.
- Fines and other punitive measures that can be imposed for the mishandling of data.
- Time limits for data storage and maintenance.
- The rights of those about whom data is being collected or processed (data subject rights).
- Mechanisms required for handling complaints about data collection and use.
- Processes and time frames for reporting data breaches.
- Requirements to hire or assign a data protection officer.

It is necessary to implement the requisite technical and organisational measures to meet legal requirements for accountability and demonstrate compliance and align organisational processes with laws and legislation related to data governance. Embedding data governance processes within daily programme activities and documenting them helps ensure that all staff are aware of the requirements. Thereafter, repeatable and standardised processes can be deployed across an organisation or institution. Staff tasked with implementing data governance should monitor the application and adherence of data governance processes across the organisation.

1.4 Put data sharing and/or data processing agreements in place to protect your organisation and the people whose data you are collecting

In addition to following national legislation, ensure that legal agreements are in place with partners and/or contractors who will be involved in data collection, processing, storage, or other data management aspects. This includes agreements with anyone with whom you will share data. Data sharing agreements and data processing agreements outline the roles and responsibilities of everyone who has access to data. They generally include clauses related to confidentiality and privacy, data use restrictions, onward sharing restrictions, data security requirements, and data retention and destruction instructions. It is important to have agreements in place even if no funding is changing hands, or in cases where data will be shared with partners who are providing technical support, even if that support is in kind.

Stage 3 and Stage 7 of this guide include orientation on agreements and policies for data transmission, storage, and retention.

1.5 Conduct a data privacy impact assessment or a risk-benefit assessment

Once you have set up your initial plans for the full Data Lifecycle, it is necessary to assess whether your plan ensures the safety and security of the data and the people who provide the data. Conducing a privacy risk assessment or a risk-benefit assessment helps to flag any potential areas of risk, develop mitigation efforts, and to then finalise a plan that reduces potential of data-related risk or harm.

More information on conducting a privacy risk assessment can be found at the ICO website. Data Privacy Impact Assessment templates are available here and here. A risk-benefits assessment framework is also available here.

1.6 Budget for the management of data from the beginning to the end of the lifecycle

At the Design and Planning stage you will formulate the plan and the budget for the full monitoring and evaluation effort, including the time, systems, people, and money involved. This will require working through the details of all the steps of the Data Lifecycle. This will enable sound planning for the financial and human resources needed to responsibly manage data throughout the full lifecycle, the staff needed to implement an effective data governance

strategy, and whether it is necessary to hire or assign dedicated staff to implement the data governance policy. It is important to conduct a skills audit and, if necessary, include a budget for training.

See Stage 5 for cleaning and analysis processes which must also be considered during planning

Some useful resources include the National Strategy for the Development of Statistics, The Path to Becoming a Data-Driven Public Sector, and Data Governance: Definitions, Challenges, and Best Practices

Designing and planning a South African government evaluation

Background and context

The global COVID-19 pandemic has had massive social and economic impacts on poor and vulnerable communities in South Africa. The lockdown period resulted in severe economic hardships on already impoverished communities, and it was necessary for the Government of South Africa had to devise response mechanisms to ensure the wellbeing of vulnerable citizens, as it is mandated to do in the Constitution.

The temporary Special COVID-19 Social Relief of Distress (SRD) grant was implemented with effect from May 2020 until January 2021 under the auspices of the Department of Social Development. It provided social grants to qualified people between the ages of 18 and 59. To qualify for a grant, a person had to be unemployed and not be receiving other government social assistance. Targeted beneficiaries used various platforms to apply for a grant, including cell phones, email, SMS, USSD and WhatsApp.

Empirical study

The Government of South Africa needed to optimise the service delivery of this key social protection measure in a tough economic climate. It conducted an empirical study to identify problematic issues; to determine what was working well or was not; and to understand key lessons to improve the payment of social grants.

Data Laws/regulations

The Protection of Personal Information (POPIA) Act 2013 (Act. No 4 of 2013), signed by the President on 19 November and published in the Government Gazette Notice 37067 on 26 November 2013 is the mandate of the Information Regulator under the Department of Justice. The purpose of POPIA is to promote the protection of personal information by the public and private sector to ensure a balance between the rights to privacy, the need for free flow of and access to information and to regulate how personal information is processed.

Application process to the information regulator for the empirical study

To proceed with data collection and conduct the learning evaluation of the SRD, the responsible government department had to apply to the Data Regulator for exemption. According to Section 11 (1) (e) of the POPI Act, a public body can perform a public law duty in terms of Section 9(1)(b)(vi) of the Social Assistance Act.13 of 2004.

The sample design of the survey required the contact information (cell phone number) for successful applicants as well as for those whose participation was rejected. The survey form asked data subjects for consent and included information on the purpose of the survey. It explained further that the exercise was voluntary and that participants could withdraw at any time; the length of the survey; the potential risks of participating in the survey (that it could cause anxiety or stress); that the personal details of participants would not be disclosed to third parties; and that all data collected would be anonymised.

In accordance with the POPI Act, approval was sought from the Information Regulator to utilise the personal contact details of the participants to be surveyed. The application to the Data Protection Authority outlined that the Department of Social Development would adhere to strict privacy and confidentiality protocols for the study. There was an undertaking that personal details of those participating in the anonymous online survey would not be disclosed and the survey excluded names, surnames, ID numbers and physical addresses and assurance was given that information was not traceable. Email addresses, URL address, IP address would not be retained or given to a third party. Participation would in no way impact on the participants' relationship with government as the survey was anonymous.

STAGE 2

Collecting or acquiring data

SNAPSHOT

This section covers aspects of data collection such as risk-benefit, data minimisation, informed consent, and ethical review/institutional review boards. Tip sheets include data collection tools, data subject rights and lawful bases for data collection. There is also a comprehensive review on the use of data from mobile networks.

At the data collection or acquisition stage, the necessary data is gathered to answer the specific M&E questions and/or to produce evidence related to the various indicators that have been established in the M&E framework and based on the Theory of Change. This might be data that is collected directly by an organisation by means of traditional processes such as surveys, focus group discussions, or interviews. Many organisations are beginning to use other, non-traditional data sources such as administrative or programme data. In some cases, this may include open data or big data such as satellite data, cell phone data records, or other sources of data that are available but have not been collected directly from the data subject.

Guidelines on collecting and acquiring data for M&E

2.1 Review the potential risks of collecting and acquiring data from or about vulnerable people

Before commencing data collection consider the vulnerability of the population of interest and ensure you have the resources to reach and effectively engage with respondents for data collection purposes. If you are collecting data through mobile devices or online, it is especially important to conduct an ethical review of the exercise paying particular attention the aspects of inclusion, bias and representation considering that some of the most marginalised populations might not have access to the internet or use digital devices. If acquiring datasets that were not specifically collected for your M&E efforts (i.e. secondary sources), it is also important to check for bias and representation in those datasets, and to ensure that the data is sufficiently anonymised to reduce risk. Note that you may require additional approval from a government department or ministry before commencing data collection.

It is important to also consider cultural aspects when designing the data collection process. The American Evaluation Association developed a Public Statement on Cultural Competence in Evaluation with the aim of ensuring recognition, accurate interpretation, and respect for diversity. Evaluators should brief members of the evaluation team on cultural competencies.

Regardless of how you collect or acquire data, it is essential to thoroughly assess the short- and long-term risks and harm that could result from collecting or acquiring data and thereafter it should be decided if and how risks can be sufficiently mitigated to avoid any potential harm as a result of your M&E. This will involve finding data collection methods that are appropriate for use in vulnerable populations. If you are collecting data about Gender Based Violence (GBV) or child abuse, for example, be aware that survivors of such violence might find it difficult to respond to phone calls in their homes. While large-scale in-person surveys are expensive to conduct, digital surveys may not be accessible to those who do not use digital devices and might even pose additional risks.

See Stage 1, 1.4 and 1.5 on related planning steps and risk assessment

2.2 Minimise the amount of data that you collect

The only way to avoid data-related risk is to avoid collecting data. However, as M&E is a data-heavy undertaking, it is inevitable that you will need to collect data to fulfil your M&E objectives. Rather than completely avoiding the collection of data, you should endeavour to minimise the amount of data that you collect.

Ask yourself: What data do you really need to collect and why? Do you have a specific use and purpose for each data point that you plan to collect? How can you minimise the personal or sensitive data that is collected?

All data collection and processing processes must be ethical, lawful, and transparent. This means that the data tools should only collect data that is necessary and proportionate to achieve the specific task or purpose for which it is intended.

See these guidelines on data minimisation for a more in-depth orientation.

2.3 Define your tools and procedures for data collection.

What is the suitability of existing data collection tools and methods for your programme? Is a bespoke approach needed? Can you afford it?

Data collection software refers to computer programs for the collection and storage of qualitative and quantitative data in electronic format and these are increasingly replacing paper-based/manual data collection methods. Further, computer-assisted telephone interviewing (CATI) is telephone surveying technique in which interviewers follow a script that is designed to flow according to previously answered questions. Another emerging technology is automated computer telephone interviewing (ACTI). This is an AI-based system where computers with speech recognition capabilities ask respondents a series of questions and store the answers through the scripted logic. This is also referred to as interactive voice response (IVR).

See Tip Sheet 1 for a list of some common tools for data collection

2.4 Ensure that you have a thorough and transparent informed consent process in place

Consent is the foremost ethical requirement for research or M&E involving people. Consent requires that participation is voluntary and that people are fully informed about the handling of their data throughout the Data Lifecycle. For this reason, it is important to decide how to manage the data throughout the Data Lifecycle so that this information can be included in the request for consent from participants involved in the data collection process.

Set clear redress mechanisms for any complaints and establish internal systems to address requests from data subjects to correct their data, withdraw consent for data processing, and/or remove their data from the system.

See Tip Sheet 2 on Data subjects' rights for details that must be provided during the consent process

It is essential that information about data collection and use is shared with the people from whom you are collecting data. Such information must be provided in plain, clearly comprehensible language and should include information on both potential benefits and risk of harm that could result from providing this data. It is highly unethical to coerce people into giving their consent, or to proceed if it is felt that they have not clearly understood what will be done with their data and what the potential risk and harm might be.

Some people or groups may be unable to give consent due to their age or capacity. In these cases, the collection of data should be avoided or, alternatively, consent should be obtained from a legal guardian. In all cases, power dynamics will influence consent. Some groups will feel pressure to consent due to these power imbalances. For example, people in rural areas might feel obligated to consent if M&E practitioners are from more urban areas. Education and the perceived economic disparity between M&E practitioners and those whose data is being collected can also influence this power dynamic, as can identification of a M&E practitioner with a government or development agency. In addition, the new ways that data is processed and used are often complex and may be difficult for people to understand, thus limiting their capacity to make a voluntary decision[1]. These issues are normally reviewed by

a national level Research Ethics Board before any data is collected. Ethics Review Boards will also likely assess whether there is bias in the study design.

In some contexts, despite the custom of local leaders or community authorities giving consent on behalf of their communities, individual consent is required for legal compliance and for ethical reasons. Research on the various options and local customs could assist in the compilation of a culturally adequate, ethical consent process that also meets legal requirements for local community members. If a person appears reluctant to provide their data or to participate in the M&E process, they should not be shamed, coerced, or forced to do so, even if a local leader has provided consent or permission.

The issue of data transparency in data governance processes may seem contrary to privacy and security; however, the two concepts are perfectly compatible. Power dynamics in the research interaction are inevitable and these affect the way in which knowledge is generated. Transparency calls for disclosure: in this case, keeping participants (and data subjects in general) apprised of research methods and how data will be used, shared, and stored strengthens accountability and fairness throughout the data value chain and it also sets the precedent for people to speak out if there is bad conduct (either by researcher or others) thereby improving data governance.

2.5 Review ethics and consent related to any 'passive' data collection

There is less 'active' data collection when collection processes become more digitised, in other words, there is less direct interaction between the person collecting data and those providing it. Active data collection and consent processes align more closely with traditional data collection methods where informed consent is standard practice.

'Passive' data collection methods are those where data is collected from secondary sources and does not require direct interaction with individual people. Passive data collection has proven to be more ethically challenging to privacy. Researchers are usually not required to obtain consent from users and this may be ethically questionable because this data already exists. For example, do users of social media platforms consider their data to be private or public?[2] Increasingly, even when data is aggregated or anonymised, certain new technologies can de-anonymise data.

Current recommendations on ethical practice for passive data collection suggest that researchers and data collectors should seek advice from an external ethics committee (more info in Section 2.6) which will critically examine any potential harm, vulnerabilities, benefits of data processing and use, even if such research does not engage directly with human subjects and does not require legal or formal ethical reviews.[3] This is an important process which can assist to reduce bias and any potentially negative effects of research or M&E.

See Deep Dive 2 on mobile network operator data

2.6 Ensure that you have research approval from relevant authorities and have complied with other legal requirements

Before proceeding with data collection, it is essential to obtain approval from any institutional review boards or ethics committees. In some countries (for example, Tanzania[4] or Kenya[5]), it is standard practice to obtain clearance from the national science and research institutions or other ethical boards before collecting human subject data. In South Africa, for example, approval to use personal information of beneficiaries must be obtained from the Information Regulator (see the POPI Act). It may be necessary to inform sub-national authorities at the county/district levels to ensure that data collection is a smooth process. Briefings with these institutions include formally introducing your organisation and any partners to these authorities.

Review any national data laws to ensure that you are compliant and that there is a lawful basis for data collection.

See Tip Sheet 3: Lawful bases for data collection

See Tip Sheet 4: Consent for orientation on how to manage an ethical informed consent process

Below we summarise some of the digital tools commonly used for M&E.

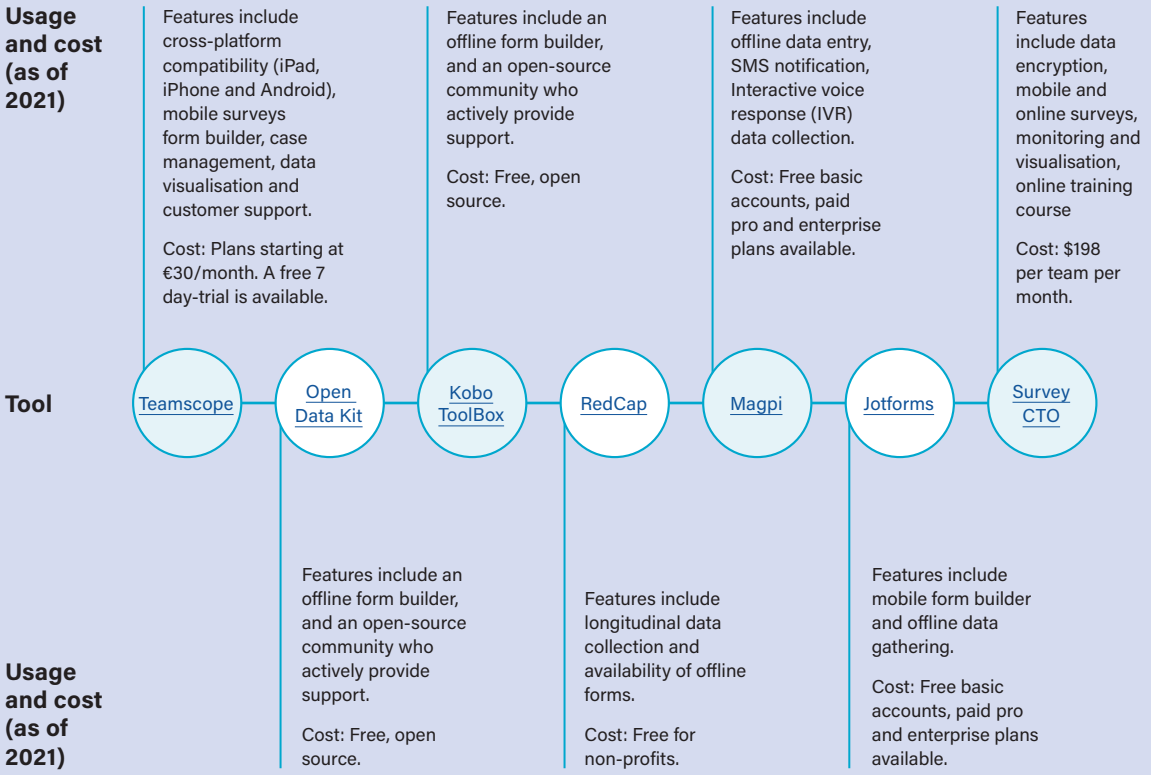
TIP SHEET 1

Common digital data collection tools and software

Digital data collection software allows for qualitative and quantitative data to be collected and stored in electronic format. Digital tools are increasingly replacing paper-based, manual data collection processes.

Most digital data collection apps can store data in the device offline until an internet connection is detected, making their use feasible even in rural areas with limited mobile network coverage. However, in many contexts digital data collection devices are still too expensive for widespread use. There are also some concerns about ensuring that data is protected from unauthorised access. Many of the abovementioned tools are constantly improving their data security and protection features yet, in some cases, improved data protection is only available after purchasing a license and is not with free versions. For this reason, M&E practitioners should be wary of using free online tools and should always review privacy policies and terms of use to ensure that data collected with a digital tool or software is protected from further use or sharing.

Below we summarise some of the digital tools commonly used for M&E.



Mobile Network Operator data

The concept of using non-traditional data for M&E purposes in the development and humanitarian sectors has recently taken centre stage. One example of non-traditional data that has generated insights in development and humanitarian contexts is Mobile Network Operator (MNO) data. This includes call detail records (CDRs), profile data (age, gender, etc.); location data (number of people in specific locations); usage data (number and duration of calls and text messages); and spend data (monthly charges, currency, on-time payments). MNO data can be combined with other types of data, such as satellite data and traditional forms of data, to produce valuable insights for development and humanitarian use. The most common type of MNO data used by M&E practitioners in development and humanitarian aid is CDR data. MNOs collect these data points for billing and other commercial purposes. This data contains personally identifiable information (PII) including phone numbers, the times at which calls are made, the duration of the call, and the source and destination phone numbers.

When combined with other datasets, MNO data can help governments and development and humanitarian stakeholders understand short-term, long-term, and seasonal mobility patterns. These are key insights for understanding how to best provide communities with essential services thereby contributing to the SDGs. MNO data can also help generate insights that help understand human behaviours and income profiles, amongst others. Such insights are useful for tracing the impact of development and humanitarian initiatives.

Why use non-traditional data, in particular, MNO data?

MNO data can provide a lower-cost alternative to traditional data gathering methods. It is more granular and real- time than other traditional datasets. Traditional methods of data collection, such as censuses or household surveys, provide vital baseline information for development initiatives. They are, however, very costly as they are both time and resource intensive. A dataset, like a national census, might only be updated every decade by government institutions, meaning that up-to-date national data or statistics are not always unavailable.

In some countries, insights derived from MNO data offer an alternative for constructing basic population statistics when gaps exist in census data and/or household surveys are unavailable. In certain circumstances, MNO data can be acquired at little (or no) cost.

What to consider when accessing MNO Data?

- Understand the MNO business model and the incentives that can drive MNOs to share data or analytics for a development or humanitarian use case.
- MNOs are for-profit entities. Their core revenue streams include voice, USSD, SMS and internet. MNO data or analytics is not widely considered to be a core revenue stream. It will therefore be necessary to consider how to incentivise MNOs to share data or analytics in a way that benefits them.
- Data obtained via MNOs requires dedicated personnel and technical expertise due to the volume of data, its veracity, the necessity for frequent data extraction, processing, transfer and storage of data, among other attributes necessary for development purposes. It is thus important to consider the various options available to alleviate the burden on an MNO of data extraction and/or data analysis. This could be done by providing technical assistance, infrastructure, in-kind support, or co-shared value propositions or commercial incentives to ease the burden of partnership from the MNO's perspective.
- When the offering of commercial incentives to MNOs seeking profits is not a viable option, consider co-shared value propositions. For example, a project can offer technical aid to generate analytics from MNO data for a development cause while also generating analytics for the MNO's business intelligence unit.
- Consider joint fundraising with MNOs. Although, this may be slower, strong joint proposals may be more attractive to MNOs than the option of self-funding. MNOs can factor in their capital costs and time during a stated period of project implementation (e.g. 2–3 years) while partners consider future sustainable business models and exit strategies early on. Additionally, a national model with multiple MNOs, government, development, and humanitarian agencies may receive support or funding at bilateral funding level.
- Understand the governance structure and leadership of the MNO to decide on the right level of engagement with the MNO. Some MNOs are best engaged at group level – possibly through their foundation arms – whilst others are best engaged at country level.
- Understand which departments within MNOs would be most suited to champion Mobile Data for Development (MD4D). This could be the executive, marketing, value-add services, enterprise or other units. It is an advantage to have an internal MD4D champion who ultimately understands the value and business case. Where possible, it is advisable to involve legal and compliance units within MNOs at the outset.
- Prior to engaging with MNOs, consult legal counsel for an opinion of the modalities of data sharing in that particular jurisdiction. Doing so will reassure the MNO's legal and compliance team as they may be embarking on this type of partnership for the first time and be concerned about issues of privacy, ethics, and consumer perception.
- Provide examples of similar data sharing agreements between MNOs and development initiatives that have been successful. This will enhance the credibility of a proposition when engaging MNOs as they are likely to have little knowledge of the use of MNO data for development or M&E purposes.
- To align with data privacy and ethics, it is preferable that data is anonymised in a safe and responsible way before it leaves the MNO environment. It is worthwhile to consider extracting analytics instead of data.

TIP SHEET 2

Data subject rights

Data subject rights are the rights that individuals have over any personal or sensitive data that is collected or managed about them. National legislations may have slightly different terms and explanations of data subject rights but the following rights are generally included in African data legislation, including established laws and regulations or those currently in the process of approval:

Right to be informed about the collection and processing of personal data. This information must be provided in a concise, transparent, intelligible, and easily accessible form, using clear and plain language. Data controllers must supply information to data subjects within one month of receiving a request for information about how their data is collected and processed.

Right of subject to access. A person has the right to obtain a copy of their personal data and an explanation of the categories of data being processed, the purposes of processing, the categories of third parties to whom the data may be disclosed, the period for which the data will be stored (or criteria for determining that period), and information about their other rights as data subjects.

Right to rectification. A data subject can request that any errors in their data be corrected.

Right to erasure (aka the 'right to be forgotten'). Data subjects can request deletion of their personal data if it is no longer needed for the original purpose or where processing is based on consent and the person withdraws their consent (and no other lawful basis for processing exists).

The right to restrict processing. A person can withdraw their consent for data processing and ask that their data is no longer processed or that its processing is limited, even if it is still stored by a data controller/processor.

Right to data portability. Data subjects can request that the data controller provides them with a copy of their personal data in a commonly used machine-readable format. They can also request that their data be transferred from one data controller to another or ask to have their data transmitted directly between data controllers.

The right to object to processing of their personal data on certain grounds. This includes the right to withdraw consent or object to processing carried out for the purposes of profiling or direct marketing. In this case, for a data controller to continue processing, it must prove that it either has compelling grounds for continuing the processing, or that the processing is necessary in connection with its legal rights.

The right not to be evaluated on the basis of automated processing. Aside from a few certain narrow exemptions, data subjects have the right not to be subject to decisions based solely on automated processing which could significantly affect them or cause them some type of harm.

The right to representation and compensation. Data subjects have the right to obtain assistance from a national supervisory authority in accessing their rights. They have a right to make a claim if they consider that their data has not been treated according to law or has been interfered with. They have the right to be represented and, in a case of proven harm, to be compensated.

TIP SHEET 3

Lawful bases for data collection and processing

To ensure the safeguarding of data that is collected and processed, data legislation normally identifies specific lawful bases under which data can be collected and processed. The lawful bases for data processing have different rules about how data is treated or managed.

In general, one of the following lawful bases must exist for personal data to be processed. (There may be slight variations in different national legislation, so it is therefore important to check the actual wording of the law for the country in which data collecting and/or processing the data.

Lawful bases for data collection and processing:

(a) Consent: the individual has given clear consent for their personal data to be processed for a specific purpose.

(b) Contract: the processing is necessary for a contract with an individual, or because the individual has asked that specific steps be taken before entering into a contract.

(c) Legal obligation: the processing is necessary for compliance with the law (not including contractual obligations).

(d) Vital interests: the processing is necessary to protect life (normally in an emergency situation).

(e) Public task: the processing is necessary for you to perform a task in the public interest or for your official functions, and the task or function has a clear basis in law (generally used by governments).

(f) Legitimate interests: the processing is necessary for your legitimate interests or the legitimate interests of a third party. However, if your interest in processing an individual's data could infringe on the rights and freedoms of the individual or put them at risk of harm, you cannot use this lawful basis, and/or you should not collect the data.

In some cases, you might be collecting part of your data under one lawful basis and using different bases for other data. For example, you might collect and process some data under 'legitimate interest' but obtain consent for more sensitive data. Your processes and systems will need to reflect the lawful bases that you have chosen, and these should be decided during the design phase of your data collection and processing. If you change your mind and want to switch to a different lawful basis once you have started collecting the data, you will need to re-collect the data so that you have everything in place according to the lawful basis requirements.

The UK Information Commissioner's Office site provides helpful resources and tools to help you determine which lawful bases is correct for the type of processing you will be doing.

Legitimate bases for UN agencies

When working in humanitarian situations, some agencies, such as United Nation (UN) bodies, are exempt from national legislation, and a separate set of rules apply. The UN and the international Committee of the Red Cross (ICRC), for example, generally work under the principle of Legitimate and Fair Processing and have established legitimate bases for data collection.

The UNHCR Data Protection Policy uses the following legitimate bases:

- Consent of the data subject.
- Vital or best interests of the data subject.
- To enable UNHCR to carry out its mandate.
- Beyond UNHCR's mandate, to ensure the safety and security of persons of concern or other individuals.

See the UNHCR's Policy on the Protection of Personal Data of Persons of Concern

TIP SHEET 3 (continued)

The ICRC data policy includes the following legitimate bases:

- Consent of the data subject.
- Vital interest of the data subject or of another person.
- Public interest, in particular, based on the ICRC's mandate under IHL and/or the Statutes of the Movement.
- Legitimate interests of the ICRC.
- Performance of a contract.
- Compliance with a legal obligation

See the ICRC's Rules on Personal Data Protection

Additional tips

- Before collecting data, determine your lawful basis (See additional tip sheets on consent and legitimate interest for more detail).
- Choose the right lawful basis very carefully because you cannot change to a different lawful basis halfway through your data collection without a particularly good reason.
- For most development work, undertaken by NGOs, INGOs and consulting firms or individuals working on M&E the lawful basis will be consent, contract, or legitimate interest.
- For government work, public Interest or vital interest are likely to be the lawful basis. In some cases, governments might be allowed to collect and use any data they choose from their citizens.
- Be extremely cautious about switching from consent to a different basis partway through your data collection exercise simply because you begin to encounter difficulties getting consent.
- Document how you selected the lawful basis so that you have it in writing (This interactive tool is useful for making the right decision).
- The language you use to explain why you are collecting data (e. g. your privacy notice or language that you include on a survey) should include your lawful basis for processing as well as the purpose of the processing.
- If working in a humanitarian or emergency situation, carefully consider and document how your lawful basis matches one of the UN legitimate bases. In such cases it is possible that you would use either consent or legitimate interest, but you may also be obligated to work under one of the legitimate bases established by the humanitarian agency leading the response.

Keep the following in mind:

- No single basis is 'better' or more important than the others – which basis is most appropriate to use will depend on your organisation's purpose and relationship with the individual.
- Different countries might use distinct categories and lawful bases, so it is important to become familiar with data privacy laws in your country of operation as well as the country which is the source of your funding. You may need to comply with laws of more than one place if your data is crossing borders.
- If you can reasonably achieve the same purpose or goal without processing personal data, then you will not have a lawful or legitimate basis and should not be collecting the data.
- If your purposes for processing the data change, you might be able to continue processing under the original lawful basis if your new purpose is compatible with your initial purpose. However, if your original lawful basis was consent and you want to use the data for a purpose other than for what you originally planned, it might be necessary to re-obtain consent from the data subjects.
- If you are processing special category data, you need to identify both a lawful basis for general processing and an additional condition for processing this type of data. Special category data includes highly sensitive data such as biometrics.

More helpful resources about lawful bases are available here

TIP SHEET 4

Consent

Regardless of the lawful basis for processing data, it is important to be transparent about the data you collect, what you will do with it, who you will share it with, and for how long you plan to keep it. You'll also need to inform people about their data subject rights (see Tip Sheet 2).

To comply with most data privacy laws and to ensure that your data collection is ethical, you must ensure that you have active, informed consent from those whose data you are collecting. When collecting data from or about children, you will need to obtain consent from a parent or guardian. It is a good practice to also obtain a child's 'assent' so that both the child and guardian are clear about the process and have had the opportunity to voice their wishes.

☑ Some key elements to include in your informed consent form are:

☐ **What data is being collected?**
Are you collecting personal or sensitive data? Explain in plain language exactly what kind of data is being collected.

☐ **Who is collecting it?**
Be transparent about who is collecting the data. Include yourself and/or your organisation and any partners involved. You may also want to include the donor that is funding the data collection.

☐ **How the data will be collected?**
Explain how data will be collected. On paper? Digitally? Via an app? Over the phone? It's also a good practice to let data subjects know approximately how long the data collection will take.

☐ **Why is the data being collected?**
Explain why data is being collected and what the anticipated benefits are for the individual or community.

☐ **How will the data be used and by whom?**
Who are the intended users of the data? How will the data be used by the M&E practitioner, the home organisation, any partners, host governments, and/or donors?

☐ **With whom will the data be shared?**
Include all parties with whom you plan to share the data. Are you contractually obligated to share data with a government department or a donor?

TIP SHEET 4 (continued)

☐ **What are the potential negative effects of the data collection?**
Is there any potential harm for individuals or groups if they choose to provide their data? Are there risks related to loss of privacy and confidentiality? If so, these should be explained.

☐ **How long will you retain the personal or sensitive data?**
Note how long you plan to store any personal of sensitive data. Will data be aggregated and retained anonymously or will you retain the raw data?

☐ **What are the individual's or community's rights related to their data?**
Provide a clear explanation of people's data rights as outlined by relevant privacy laws (See Tip sheet 2 for an overview of common data subject rights).

☐ **How can someone contact you/your organisation with questions, concerns, or complaints?**
Provide a phone number, address, and/or an email address where people can contact you or your partners for more information.

☐ **How can someone withdraw consent?**
What is the process for revoking consent or correcting data that is held about an individual?

Other considerations

Storing consent documentation

You must ensure that consent documents are securely stored and are easily accessible for future use in case questions regarding consent are raised or if someone wishes to withdraw their consent.

Re-using data

You must obtain 're-consent' from individuals in cases where you wish to share data subjects' personal or sensitive data with additional partners, retain raw data for longer periods, use data for purposes other than those outlined in the consent form, and in cases where changes have been made to the original consent form.

Community versus individual consent

Even if a local authority or village head has authorised the collection of data in a community, it is still necessary to obtain consent from individuals before collecting their data.

Online or mobile consent

When collecting data digitally, consent is still required. Consent can be obtained via an application or on a mobile device as part of a survey or self-reporting process. Audio consent can also be recorded and stored as proof of consent.

STAGE 3

Responsible M&E data transmission and storage

SNAPSHOT

This section covers data security during transmission and storage, a review of third-party vendors, and data breach protocols.

It is essential to safeguard data – especially personal or sensitive data – by protecting and securing it from unauthorised surveillance, capture, modification, downloading, copying, transfer, sharing, tampering, unlawful destruction, accidental loss, and/or improper disclosure during transmission and storage. This requires that certain precautions are followed when transmitting data (when data is sent from one place to another) and when storing data whether this is on a laptop or mobile device, on desktop computers, on servers that your organisation owns and manages, or in the cloud.

The core functions of the IT team include the secure transmission and storage of data and this is done by employing appropriate protection procedures, processes, and systems. If you manage data on behalf of an organisation or as part of a contract to collect and process data, it is especially important to follow guidance from IT, and to implement the necessary cybersecurity precautions to ensure that data is stored as securely and privately as possible.

Guidelines for responsible transmission and storage of data

3.1 Implement good security measures to protect data in transmission or storage

- Use secure passwords and apply a strong password policy.
- Make use of 2-factor authentication.
- Encrypt files during transmission and encrypting devices where possible.

- Use virtual private networks (VPNs) for private browsing where possible (in some countries, e.g. Uganda, governments might prohibit the use of VPNs).
- Ensure that Wi-Fi networks are private and secure; avoid the use of public Wi-Fi networks (such as free WiFi in coffee shops or hotels) when working with personal or sensitive data.
- Avoid sharing personal or sensitive data by email and use secure file transmission processes such as SharePoint, DropBox or Box.net.

- Set personal or sensitive data access levels to a 'need to know' basis and regularly update the list of people who can access data in accordance with consent processes and national data laws.
- Prohibit the storage of personal and sensitive data on flash or other portable drives, unless it is necessary to do so, and then only for as long as it is necessary.
- Make sure that devices are protected from viruses, spyware, and malware and that all staff are aware of how to spot breaches so that they do not fall prey to them. Ensure that staff know how to identify and report a laptop or device intrusion or data breach and are comfortable with reporting it so that it can be resolved as quickly as possible.
- Ensure that personal and sensitive data is anonymised, de-identified, and/or aggregated as soon as possible after collection, unless the raw, personal, or sensitive data is needed for a specific purpose.

- Follow national and international laws related to cross-border data transmission and the location of data storage if data is stored in the cloud.

3.2 Ensure that third-party vendors protect your data

If you are hiring a tech vendor or firm to manage, store, analyse, or otherwise process your data or you are considering pro-bono offers for these types of services from a company or advisory firm, there are critical issues to bear in mind regarding data privacy and security.

For example, it is necessary to enter into a legal agreement which states that the data will be protected and not used or shared with any other party. You should include any third-party vendor information in your consent form so that individuals are aware of who the data will be shared with.

☑ Here are some questions to ask when contracting or securing pro-bono services with a company or vendor with whom you might share data about your staff or data from or about people or partners that you are working with. Asking these questions does not guarantee that the vendor, contractor, or company will manage data in a perfectly private and secure way, but it can give you a sense of whether they take security and privacy seriously, and if they have the capacity to protect the data.

- ☐ What measures do they have in place to ensure and demonstrate compliance with data privacy principles and/or national data protection legislation, health data protection regulations or other similar data standards?
- ☐ Do they maintain records of their processing activities that are compliant with national regulations and/or industry standards? Can they produce these records if needed?
- ☐ Would they be working with any sub-processors or sharing your data with others? If so, who? And how do they ensure that sub-processors are held to the same standards of data protection?
- ☐ How do they deal with data subject access requests, such as requests to correct, delete, or restrict data processing? Are they able to trace consent? How do they authenticate users of their systems? Do they use two-factor-authentication?
- ☐ Who has access to the data and how is access determined? What controls are in place and how often are they reviewed or updated?
- ☐ What experience do they have in conducting data privacy impact assessments?
- ☐ Are they prepared to approach this work from the lens of privacy by design/privacy by default?
- ☐ What information security compliance measures are in place? Do they stress-test their systems?
- ☐ Have there been any past data security issues, breaches or documented/public criticisms of their tools and services? How, and how quickly, did they respond?
- ☐ How do they govern their data? Who is responsible for data security and data breaches?
- ☐ What are their data security breach management and notification policies and procedures?
- ☐ Have any of their staff been trained on data privacy laws and/or do they have legal counsel that is aware of privacy legislation? How have they prepared internally for compliance with data privacy laws?
- ☐ How likely is it that the vendor will be around in the long-term? If they shut down or are acquired, what will happen to the data they are holding?
- ☐ Do they have a privacy policy? Do they share, provide, or sell personal data or data profiles to third parties? What is their business model and does it involve acquiring and profiting from personal data?

See Tip Sheet 9 for more on how to develop a Data Retention Policy.

TIP SHEET 5

Developing a data breach protocol

As you collect, transmit, store, and share increasing amounts of data, the possibility of a data breach increases. A data breach refers to any incident involving unauthorised access to a system containing personal data, theft of a device containing electronic personal data, or loss of physical or electronic data. Data corruption is also considered a data breach, as is any other incident that affects the availability of personal data, such as a ransomware attack. When you hold personal and sensitive data, a breach or leak can expose vulnerable groups involved in your M&E efforts to harm.

Having a *data breach protocol* in place is essential to help organisations prevent data breaches and, if breaches do happen, to respond speedily and appropriately. Irrespective of how well secured data is, the possibility of a breach is always there. For this reason, it is critical to be prepared to react quickly when a breach occurs. A personal data breach is defined as "a breach of security leading to the accidental or unlawful destruction, loss, alteration, unauthorised disclosure of, or access to, personal data transmitted, stored or otherwise processed".

Plan ahead so that you are ready to respond rapidly and appropriately in the case of a breach or leak. Various teams will have a role in preventing and responding to data breaches.

The first step in protecting against a data breach is prevention. Different teams have specific roles in prevention and preparation. The following are important points to consider:

- While all teams might have a role in prevention, a smaller sub-set should be identified as those responsible for managing a data breach, e.g. designated persons from IT, HR, legal, communications/PR, finance, and a member of the team affected by the breach.
- Individuals in the data breach sub-set should receive prior training and meet periodically to discuss their roles and responsibilities so that any breach can be dealt with swiftly and efficiently.
- A simulation exercise should be organised at least once a year to maintain the vigilance of the team in preparation for a possible breach.

ORGANISATIONS AND CONTRACTORS	▪ Ensure that updated data privacy and security protocols are in place in order to reduce the likelihood or severity of potential breaches.
IT	▪ Ensure cybersecurity is top-notch to reduce the possibility of a breach. ▪ Provide guidance on how to manage devices and data security.
HR	▪ Train and orient staff on how to avoid a data breach and how to report a suspected intrusion or breach. ▪ Handle and secure employee data appropriately.
COMMS	▪ Minimise the amount of comms data collected and stored to mitigate potential breach damage. ▪ Follow IT and HR data security policies.
FINANCE	▪ Minimise the amount of financial data collected and stored to mitigate potential breach damage. ▪ Follow IT and HR data security policies. ▪ Engage with Legal to prepare for any financial implications of a breach.
LEGAL	▪ Prepare legal precedent and requirements for handling data and reporting a breach. ▪ Negotiate data privacy requirements and data sharing agreements with partners and contractors. ▪ Determine what the organisation is willing to do in the event of a data breach.
PROGRAMMES	▪ Follow good data privacy and security principles ▪ Minimise the amount of programme and beneficiary data collected and stored to mitigate potential breach damage. ▪ Work with partners to support data privacy and security efforts to help avoid a data breach.

M&E	▪ Use 'privacy by design' principles when designing research or M&E. ▪ Follow sound data privacy and security principles. ▪ Minimise the amount of M&E data collected and stored to mitigate potential breach damage. ▪ Work with partners to support data privacy and security efforts to help avoid a data breach
ALL STAFF	▪ Adhere to IT data security recommendations. ▪ Immediately report any suspected data breaches, including loss of a device or removable drive, potential malware or spyware, or other suspected intrusions into the system.

Reporting a data breach or leak

Many national data privacy laws require that certain types of data breach are reported to a Data Protection Authority. If data was encrypted, or the data was accessed but not misused, or if data was highly aggregated and/or anonymised, potential harm and risk may be minimal, and it may not be necessary to notify any authorities or data subjects of the breach. In most cases, there is a 72-hour time limit for informing Data Protection Authorities of a breach unless the breach is unlikely to result in a risk to the rights and freedoms of natural persons. There may also be time frames for directly notifying individuals whose data has been breached. In Nigeria, for example, data subjects must be informed within 48 hours of the Data Protection Commission being notified of any breach of their personal data.

DATA PROCESSORS	▪ Notify the data controller immediately if a breach is suspected.
STAFF	▪ Inform IT immediately if a breach is suspected or a device is lost or compromised.
DATA CONTROLLERS	▪ Inform data protection authorities within 72 hours (Check national laws for specific requirements per country). ▪ Inform affected persons or organisations if there is potential for harm due to the breach.

Responding to a data breach or data leak

1. **Immediately assemble the team responsible for dealing with a data breach.**
   This will likely include a representative from IT, HR, legal, communications/PR, finance, and a member of whichever team was affected by the breach.

2. **Determine what happened and how severe it is.**
   - ☐ What was the nature of the breach or attack?
   - ☐ What was the extent of the breach or attack on the system?
   - ☐ What assets are affected?
   - ☐ What information is affected?
   - ☐ What partners or associates or other networks have been affected?
   - ☐ What are the implications of the attack on the organisation and/or its partners or data subjects?

3. **Immediately work to contain the breach. If a network was affected, disable any connections to the point of the breach to prevent further access.**
   - ☐ Clean the system of any unwanted files that might have been installed and make a detailed report of what has been removed for further analysis at a later point.
   - ☐ Run security patches and software updates.
   - ☐ Isolate critical data, especially any highly sensitive data such as financial data or data of vulnerable or sensitive groups.
   - ☐ Initiate new login procedures to any networks and/or devices.
   - ☐ Uninstall and reinstall affected files and programs.

TIP SHEET 5 (continued)

4. **If the breach was a lost, stolen, or compromised device.**
   - ☐ Determine whether there was any personal or sensitive data about individuals, groups, or partner organisations on the device that could be accessed (based on the types of security and encryption that were on the device).
   - ☐ Determine (if possible) who may have accessed the device and whether they may have an interest in any of the data on it.
   - ☐ Determine whether any harm could come to those whose data was compromised and, if so, what type of harm.
   - ☐ Determine if, based on the applicable laws, it is necessary to report the incident to authorities and/or data subjects.
   - ☐ Determine what response or mitigation efforts will be put in place for partners (clients) and/or data subjects, depending on the type of data that was breached and an assessment of potential harm.
   - ☐ Determine if there is a need for any type of damage compensation or additional support to those who have been severely affected by a data breach and how that would be managed and/or financed.
   - ☐ Determine how to reach and clearly explain the situation to individuals or partners (clients) who may be affected. This should be done in a manner that provides them with the information they need to manage any consequences but should not alarm them or make them anxious or upset.
   - ☐ Prepare and release any necessary public notification or media release about the breach and equip staff with a clear narrative about the breach.
   - ☐ Ensure that contact information is available for anyone who has further questions or requires additional support in relation to the breach.

Roles for different teams in responding to a data breach

The response team generally consists of members from all part of the organisation. Below is a chart that defines what roles would be played by different teams in the event of a breach. If working with other partners or contractors, they will need to be involved and informed as well.

CONTRACTORS AND PARTNERS	▪ Anyone who is processing data on our behalf needs to inform us immediately if a data breach is suspected.
IT	▪ Identify and address compromised data. ▪ Determine the number of records compromised and the types of personal information that they contain. ▪ Support forensic investigation and evidence preservation. ▪ Oversee deletion of malware or hacks and correct vulnerabilities that might have precipitated the breach. ▪ Monitor systems for additional attacks. ▪ Fix gaps in the IT system. ▪ Hire additional expertise that may be needed to identify cause and scope of a breach and the type and location of compromised data.
HR	▪ Keep employees updated about data breaches (depending on how severe). ▪ Decide on appropriate action if the breach is linked to a particular employee or their actions.
COMMS	▪ Draft messaging to inform different stakeholders of a breach (where necessary) and mitigate any brand or reputational damage. ▪ Respond to any media inquiries. ▪ Provide reassurance that the breach is being handled.
FINANCE	▪ Determine the financial impact of a breach and recommend budget parameters to respond to the breach. ▪ Work with any vendors affected by the breach.

TIP SHEET 5 (continued)

LEGAL	▪ Advise on response notifications to affected individuals, media, law enforcement, internal teams, government agencies, financial institutions or other third parties. ▪ Review contractual requirements in the case of a breach involving a partner or third party. ▪ Prepare for any post-breach litigation. ▪ Notify regulators and law enforcement. ▪ Alert credit card/credit reporting agencies if needed. ▪ Review contracts to understand any obligations. ▪ Help manage breach investigations and evidence preservation. ▪ Review communications for potential liability.
PROGRAMMES	▪ Work with partners in the event of a data breach.
M&E	▪ Work with partners in the event of a data breach.
ALL STAFF	▪ Keep contacts and partners informed as directed by the incident response team.

☑ Recovering from the initial data breach crisis and lessons to be learnt to improve responses to future crises

- ☐ Appoint a specific person to handle future questions or communication about the breach.
- ☐ Prepare a final report for the files as well as for local authorities or clients on the response effort.
- ☐ Use the data breach experience to improve response systems and ensure that learning is applied in the event of a future data breach.
- ☐ Continue to educate and make staff aware of data privacy and security protocols to prevent future breaches.

STAGE 4

Responsible M&E data cleaning, analysis and use

SNAPSHOT

This section offers guidelines on responsible processes and steps, such as data quality standards and data anonymisation. It includes a review on dealing with missing data and a tip sheet table for selecting analysis methods.

Data cleaning and analysis is the stage in the data lifecycle during which M&E data is prepared and processed to generate useful insights. Depending on the scale of the project, this stage may require specific expertise and experience of working with a specific type of software. You may, for example, need a data analyst, researcher or even a data scientist if you are working with more advanced approaches.

Analysing quantitative and qualitative datasets should start with a transparent analysis plan which includes steps for cleaning raw data. Raw data often includes errors or consistencies that need to be checked and adjusted during the data cleaning process. Omission of this step could result in inaccurate conclusions being drawn. The data cleaning process must be carefully planned so that raw datasets are not mislaid and crucial data is not removed in error. M&E practitioners should follow clear guidelines when cleaning of their datasets. Guidelines should be agreed upon at organisational level in the form of Standard Operating Procedures and other procedures, as outlined below.

Guidelines for cleaning and analysing M&E data

You may have heard the phrase 'garbage in, garbage out". In the context of data, this maxim is a clear reminder us that useless data will result in useless insights. It is imperative that potential risks in data quality are identified and dealt with before data is analysed. Data cleaning involves addressing incorrect, corrupt and incorrectly formatted data, and removing duplicate entries or incomplete data within a dataset. This process may be onerous if it requires a significant number of data queries to be handled before the actual analysis can take place. Data quality standards should (ideally) be established at the beginning of any M&E planning. This is the framework that will guide data quality assurance throughout M&E processes and will also highlight specific steps to be taken for data cleaning after data has been collected.

Data quality standards (USAID)

- Validity – the data is a direct measure of the intended result, and no measurement errors were made.
- Integrity – the data is protected from transcription errors or manipulation.
- Precision – the data has a sufficient level of detail to inform decision making.
- Reliability - the data reflects stable and consistent information.
- Timeliness – the frequency of the data is useful, current, and timely.

See Tableau's guidance on data cleaning and **USAID's** data quality assurance processes for additional orientation.

4.1 Establish clear processes for data cleaning and for the monitoring of data quality, validity and integrity

It is important to systematically organise and document how the data cleaning process will be conducted. Good practices include:

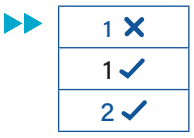
- **Back up the raw data** before starting the data cleaning process and set up a clear and logical file naming system.
- **Anonymise the data.** As the data collected may include personal details of respondents, unique data identification numbers should be assigned for each respondent prior to the analysis phase to ensure the anonymity of respondents. When data is collected from a small number of respondents and variables could allow for respondents to be identified, other measures should be taken such as aggregation of results to a level which prevents individual identification. The data being cleaned and analysed at this point should be restricted to what is needed for the intended M&E purpose. However, in cases of where secondary data sources are used, data fields unnecessary for the intended purpose must be excluded.
- **Keep a logbook** with an overview of all the variables, variable labels, and variable codes of the dataset. Use this logbook to keep track of how variables were recoded.
- **Run descriptive analyses** of data (e.g. frequency sheets, histograms) to determine where the errors in the data may lie. Look out for coding errors, missing values, duplicates, out-of-range values, and other errors. In the analyses of data, protection is also paramount, and data should routinely be presented at aggregate level to prevent identification of respondents.
- **Work with scripts/syntax files** to document all the steps taken in the data cleaning process and save the cleaned file as a separate file.
- **Devise a plan** for dealing with missing values, duplicates, or incorrect values.

See Guidelines on collecting and acquiring data in Stage 2 for additional ways to assure data quality during collection.

4.2 Create and use clean data

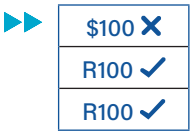
A few additional steps must be taken to prepare data for analysis and use, including:

- **Remove duplicate or irrelevant observations.** Duplicate data is data which is repeated. This may occur if an individual submits an entry twice or data is collated from various data sources which include repeat individuals, entities, or units. There are simple software methods that can help with this, for example Microsoft Excel's feature for finding and removing duplicates. However, if there is no unique identifier this is not always a simple process. Your dataset might have two entries for the same name if two individuals in the set happen to have the same name. You may need to group fields according to addresses, phone numbers, and ages to create a unique identifier. Scanning of small datasets and flagging duplicates may also be necessary. This is an important first step as it affects the true scale of your dataset.

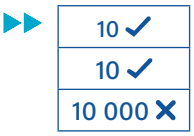


- **Fix structural errors.** Conducting a basic descriptive analysis of the data enables the identification of anomalies in the data, such as outliers or errors in the presentation of the data. Sometimes the outliers (data points which appear to far exceed the average of the rest of the data) are correct, but it is important to verify these. There might also be cases where data is recorded or saved in a format which causes a visible error, for example, US$ might be indicated instead of ZAR currency. Frequencies and ranges should be established for each variable so that structural errors can be identified.

  Many errors in datasets can be avoided by using digital survey tools instead of traditional paper-based surveys. Such tools allow survey designers to incorporate features such as mandatory answers, skip logics, ranges, and to impose limits on the number of surveys submitted. These features prevent respondents or enumerators from skipping questions, providing out-of-range answers or submitting multiple responses.

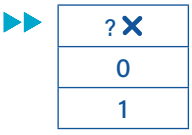


- **Filter unwanted outliers.** If an outlier/s is identified in the data it is important to assess the impact that the outlier data point will have on the final analysis. Outliers can skew the average for a particular variable or influence an incorrect relationship between variables. It is important to first identify the outliers as outlined above and then to consider creating one dataset that includes the outliers and another dataset that excludes the outlier. This could assist to provide a more relevant analysis. Often an outlier is a valid data point and should be further investigated to understand the unique factors or underlying data which influence it which, while they could be important, are not necessarily generalisable to the rest of the population or sample.



- **Handle missing data.** Datasets must be checked to determine whether missing data is a true missing field or if it represents a zero or null response. This must be clarified so that null fields (where someone responded 'no' or provided a value of zero) are documented as such and not as a missing value. True missing values must then be checked for the possibility that they have populated a field based on other observations related to it.

  For example, in the case where Question 1 asks: Have you eaten a meal in the last hour? = yes/no; and Question 2 asks: What type of meal did you eat? = list. If there is missing data for Question 2, you could identify how many responded 'no' to Question 1 and complete the field for Question 2 as 'no meal' so that it is not mistaken for missing data. In other instances, it might be more appropriate to drop entries with missing data. This may affect the representativeness of your sample, so this option must be considered carefully.



For more reading and learning on this topic, see Tableau's guidance on data cleaning .

4.3 Selecting the appropriate method of data analysis

Data analysis is the process of seeking patterns in data. This includes quantitative (numeric) and/or qualitative (text, images) data. Data analysis fields are vast and have many levels of complexity.

See Tip Sheet 6 on the next page for an overview of some data analysis methods and uses.

4.4 Other considerations for responsible data cleaning and analysis

Responsible data analysis might be overlooked if we assume that it is sufficient to focus on ethical data collection methods. Considering ethics at the collection stage, however, does not ensure that data is managed ethically during the data lifecycle. There are some important questions to consider during the analysis of data, such as:

- How have you aggregated your data (e.g. gender or race grouping)? Ensure that the data aggregations in your analysis are not exclusionary or prejudiced.
- Have you included sufficient supporting data to understand nuances? Data patterns alone cannot inform a decision. They must be stress-tested against contextual factors. Is the trend you identified applicable in all contexts? How far can the trend be extrapolated? Have you developed a spurious correlation between variables?
- Have you completed sufficient quality checks and balances to ensure that your analysis is sound? Human error or lack of data checking can result in the misrepresentation of data.
- Have you included representative stakeholders in analysis and interpretation? This is a crucial step to help build ownership of the information and validate the data, and also assists to fill any inclusion gaps.
- Have I accounted for the sample size and power in the analysis and how my data is interpreted?

Other useful resources include the Responsible Data Forum and this piece on Responsible Data Science.

TIP SHEET 6

Selecting analysis methods and uses

Method	Description	Used for numeric data	Used for text data	Describing data	Seeking relationships	Seeking trends	Exploring unknown patterns	Predictive models
Correlation	A statistical measure ranging from +1.0 to -1.0 that indicates how strongly two or more variables are related.	✓			✓			
Cross-tabulations	Using contingency tables of two or more dimensions to indicate the relationship between nominal (categorical) variables.	✓		✓	✓			
Data mining	Computer-driven automated techniques that run through large amounts of text or data to find new patterns and information.	✓				✓	✓	
Frequency tables	A visual way of summarising nominal and ordinal data by displaying the count of observations (times a value of a variable occurred) in a table.	✓		✓				
Measures of central tendency	A summary measure that attempts to describe a whole set of data with a single value that represents the middle or centre of its distribution. The mean (the average value), median (the middle value) and mode (the most frequent value) are all measures of central tendency.	✓		✓				
Measures of dispersion	A summary measure that provides information about how much variation there is in the data, including the range, inter-quartile range and the standard deviation.	✓		✓			✓	
Multivariate descriptive	providing simple summaries of (large amounts of) information (or data) with two or more related variables. Specific methods include multiple regression; factor analysis; cluster analysis.	✓		✓			✓	

TIP SHEET 6 (continued)

Method	Description	Used for numeric data	Used for text data	Describing data	Seeking relationships	Seeking trends	Exploring unknown patterns	Predictive models
Non-parametric inferential statistics	Methods for inferring conclusions about a population from a sample's data that are flexible and do not follow a normal distribution (i.e. the distribution does not parallel a bell curve), including ranking: the chi-square test, binomial test and Spearman's rank correlation coefficient.	✓			✓	✓	✓	✓
Parametric inferential statistics	Methods for inferring conclusions about a population from a sample's data that follows certain parameters: the data will be normal (i.e. the distribution parallels the bell curve); numbers can be added, subtracted, multiplied and divided; variances are equal when comparing two or more groups; and the sample should be large and randomly selected.	✓			✓	✓	✓	✓
Summary statistics	Provide a quick summary of data which is particularly useful for comparing one project to another, before and after.	✓		✓				
Time series analysis	Observing well-defined data items obtained through repeated measurements over time.	✓				✓		✓
Textual analysis	Analysing words, either spoken or written, including questionnaire responses, interviews, and documents.		✓				✓	
Content analysis	Reducing large amounts of unstructured textual content into manageable data relevant to the (evaluation) research questions.		✓	✓			✓	
Thematic coding	Recording or identifying passages of text or images that are linked by a common theme or idea allowing the indexation of text into categories.		✓	✓				

TIP SHEET 6 (continued)

Method	Description	Used for numeric data	Used for text data	Describing data	Seeking relationships	Seeking trends	Exploring unknown patterns	Predictive models
Framework matrices	A method for summarising and analysing qualitative data in a two-by-two matrix table. It allows for sorting data across cases and by theme.		✓	✓			✓	
Timelines and time-ordered matrices	Aids analysis by allowing for visualisation of key events, sequences, and results.		✓		✓	✓		

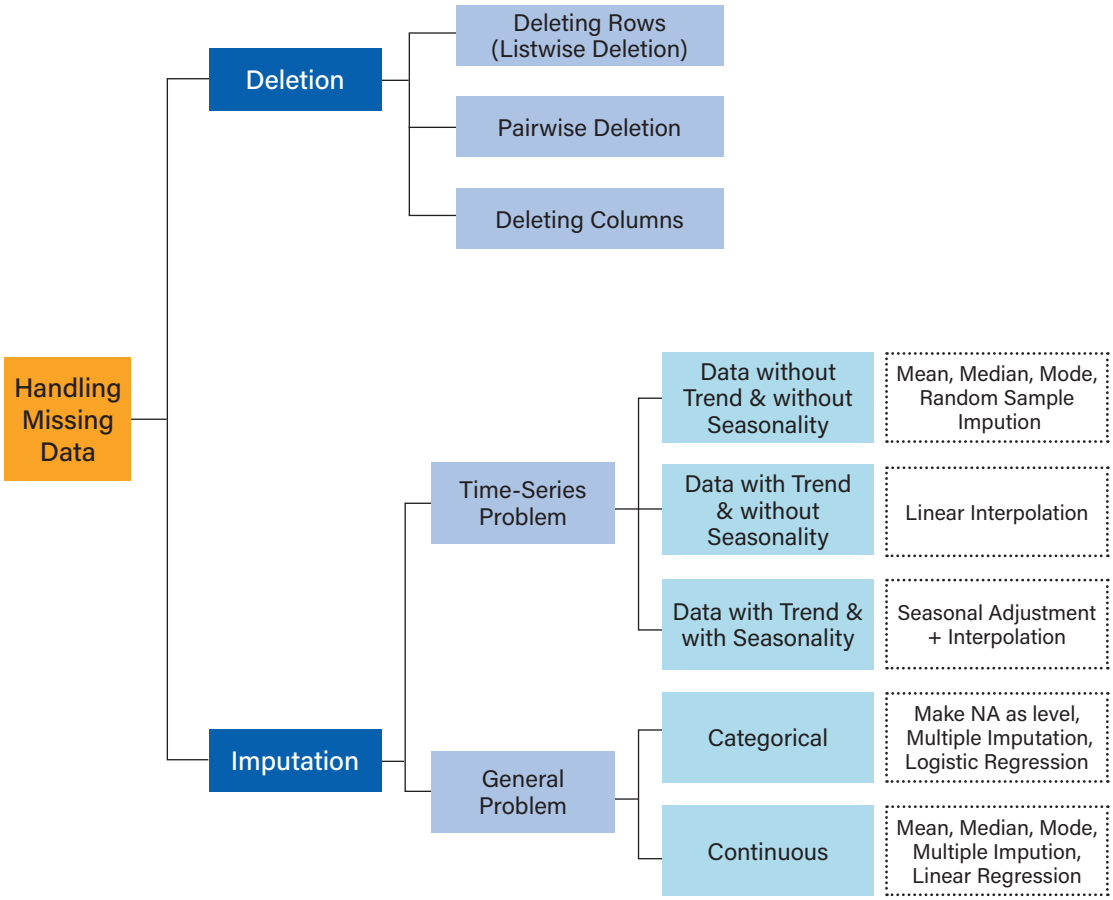
Dealing with missing data

A dataset with many missing values will affect the quality of your analysis. At the outset it is important to ask why there are missing values. Is the nature of the question too difficult or too sensitive? Is the missing value pattern related to the characteristics of the respondent (for example, age)? Understanding why the data is missing will help you decide how to deal with the missing data.

The term Missing Completely at Random (MCAR) refers to causes of missing data being unrelated to the data collection design, and data may be missing simply because of bad luck (e.g. the weighing scale ran out of batteries). The term Missing at Random (MAR) implies that there is a known factor at play that influences the missing value pattern, but this factor is random (e.g. the weighing scale works differently on different surfaces). If we know that the missing data is due to MCAR or MAR, missing data can be safely removed or replaced with a mean, mode, or by multiple imputation without introducing bias.

The term Missing Not at Random (MNAR) is more problematic, as this implies that there is a relationship between the respondent and the missing data. This may be the case, for instance, when the weighing scale is not as sensitive when people are very heavy. When MNAR situations are encountered, it is important to investigate the causes of missing data so that different sensitivity analyses can be applied.

Figure 4: Overview of techniques for handling missing data (MCAR and MAR)



See causes of missing data, flexible imputation of missing data, and how to handle missing data for more on this topic.

STAGE 5

Responsible open data and data sharing

SNAPSHOT

This section provides an orientation on safe and responsible data sharing and open data, as well as guidance on establishing data sharing agreements based on the distinct roles that involved entities play. A review on data sharing is also included.

Several scenarios exist which might involve the sharing of data internally. Within government, for example, this would be across ministries. In some cases, data might need to be shared externally with partners or with contracted third parties such as evaluation firms, survey companies, technology providers, or data processors. You might also be sourcing existing data, such as administrative data, from various places for an assessment or evaluation. Or you might also be asking partners or third parties to share data they have collected with you. In these cases, it is important to assess the degree of data privacy and security and to develop data sharing agreements or data processing agreements in cases where personal, sensitive or confidential information is to be exchanged or shared.

Data sharing agreements outline aspects such as what data will be shared, for what purposes, and for how long. They set out how confidentiality should be managed and the accountability that each partner has for the data. They must align with consent processes. Data sharing agreements are normally established when different parties are working together on an initiative and need to share data to meet their partnership and programme goals such as conducting national level surveys or needs assessments, providing services across multiple ministries or organisations, ensuring that there is no duplication of services, or meeting legal requirements.

Data processing agreements are similar to data sharing agreements. Data processing agreements are usually signed if you engage a company to process your data, for example, if you are working with a third-party survey company such as Ona, or a management information system like Salesforce. The third-party company will usually create the agreement for your review and signature as part of the contracting process. You should carefully review it before signing.

Guidelines for open data and sharing data

5.1 Open data where possible, but only after ensuring that it does not lead to harm

National government might have a mandate to open its data. Perhaps this is because the government has signed an agreement to open data, for example, the Open Data Partnership, or because a bilateral or multilateral donor requires data to be opened and shared as part of an initiative such as the International Aid Transparency Initiative (IATI).

Open data is data that can be freely used, re-used, and redistributed by anyone, and is subject only to the requirement to attribute and share alike (any products or derivative works using the data must be licensed under the same type of license).

The main aspects of open data as described in the Open Data Handbook are:

- **Availability and access.** The data must be available as a whole and at no more than a reasonable reproduction cost, preferably by downloading over the internet. The data must also be available in a convenient and modifiable form.
- **Re-use and redistribution.** The data must be provided under terms that permit re-use and redistribution including the intermixing with other datasets.
- **Universal participation**. Everyone must be able to use, re-use and redistribute the data – there should be no discrimination against fields of endeavour or against persons or groups. For example, 'non-commercial' restrictions that would prevent 'commercial' use, or restrictions of use for certain purposes (e.g. only in education), are not allowed.

A benefit of open data is that it is 'interoperable', meaning that the data is structured in a way that allows it to be used by different systems and organisations, and that different datasets can be combined because they share the same structure or are in the same format. It is important to note that not all data can and should be opened. Personal data and sensitive data should never be opened, as this puts individuals at risk. Combining datasets could also inadvertently identify individuals in the dataset. Some national security restrictions might also apply to decisions about whether data can or cannot be opened.

5.2 Put clear agreements in place before sharing data

Data protection laws and regulations assign roles and responsibilities to the different parties in a data sharing or data processing arrangements – typically data controllers and data processors. These designations are useful for assigning obligations, responsibilities, and accountability for data and these should be formalised in a data sharing or processing agreement.

- The **data controller** makes decisions about how personal data will be processed.
- The **data processor** takes instruction from the controller on which personal data is collected and how it will be processed.
- **Joint controllers** decide together on the purpose of processing personal data and the method to be used for processing.

Data controller: The entity that alone or jointly with others exercises control over the purposes and means of the processing of personal data is known as a data controller. Controllers hold the highest level of compliance responsibility for data. They must demonstrate compliance with data protection principles and other legal requirements. They are also responsible for ensuring the compliance of their data processor(s). Supervisory authorities and individuals can act against controllers in cases of a breach of controllers' obligations.

Data controllers:

- Decide to collect or process the personal data.
- Decide what the purpose or outcome of the processing is to be.
- Decide what personal data should be collected.
- Select the individuals whose personal data will be collected.
- Obtain a commercial gain or other benefit from the processing, except for any payment for services from another controller.
- Process the personal data according to the contract with the data subject.
- Make decisions about the individuals concerned as part of or because of the processing.
- Exercise professional judgement in the processing of the personal data.
- Have a direct relationship with the data subjects.
- Have complete autonomy as to how the personal data is processed.
- Appoint the processors who will process the personal data on their behalf.

Joint Controller: If two or more data controllers together determine the purposes and means of processing the same personal data, they are known as joint controllers. However, they are not joint controllers if they are processing the same data for different purposes. Joint controllers must decide who will take primary responsibility for compliance with legal frameworks. All joint controllers remain responsible for compliance with controller obligations. Data protection authorities and individuals may act against any controller in cases of a breach of controllers' obligations.

Organisations are joint controllers in cases where they:

- Have a shared objective regarding data processing.
- Are processing the personal data for the same purpose.
- Are using the same set of personal data (e.g. one database) for this processing.
- Have designed the process together.
- Have common information management rules.

Data Processor: The entity that processes personal data on behalf of, or under the instruction of, a data controller is known as a data processor. Data processors do not have a purpose of their own for processing data. Processors do not have the same obligations as controllers and do not have to pay a data protection fee. However, processors do have direct obligations. Both supervisory authorities and individuals may act against a processor regarding in cases of a breach of a processor's obligations.

Data processors are those who:

- Follow instructions from someone else regarding the processing of personal data.
- Are given the personal data by a customer or similar third party or told what data to collect.
- Do not make the decision to collect personal data from individuals.
- Do not decide what personal data should be collected from individuals.
- Do not decide the lawful basis for the use of that data.
- Do not decide what purpose(s) the data will be used for.
- Do not decide whether to disclose the data, or to whom.
- Do not decide how long to retain the data.
- Make some decisions on how data is processed but implement these decisions under a contract with another party.
- Have no material interest in the outcome of the processing.

These definitions and checklists are drawn from the Information Commissioner's Office of the UK and from the Cash Learning Partnership's Data Responsibility for Cash and Voucher Assistance Toolkit. We have minimally adapted them. See the originals for more details.

See related practices in **Guidelines for responsible transmission and storage of data in Stage 3.** When establishing data sharing or processing agreements, sufficient protection should be in place for data transmission and storage.

TIP SHEET 7

Developing a data sharing agreement

☑ Before entering into a data sharing agreement, it is important to understand the following:

What is the purpose of data sharing? What plans or requirements do you have?

- ☐ Why are you sharing data and what will be achieved? What is the specific goal?
- ☐ What potential harm could result in the short- and long-term from the sharing of data?
- ☐ What are you and your partners, third parties, or sub-contractors planning to do with the data? Do you plan to share it onward with others and, if so, for what purpose? Will it be shared with national or local governments or government ministries? Will it be shared with donors?
- ☐ Are there requirements for data sharing or opening data? How do they affect your process?
- ☐ For how long will you be sharing data? What happens at the conclusion of the agreement or initiative or if the funding runs out? What is the plan?
- ☐ When will data be anonymised, aggregated or deleted?

What kind of relationship will you have with your partners or contactors?

- ☐ Is this a grant agreement where you are receiving funding to collect, process or manage data?
- ☐ Is it a contract agreement where you are contracted by someone to collect, process, or manage data?
- ☐ Is it a contract agreement, where you are contracting someone else to collect, process or manage data?
- ☐ Is it a sub-contract in which case the original data sharing conditions need to be replicated?
- ☐ Is the data sharing between two partners or multiple partners?

TIP SHEET 7 (continued)

What exact data needs to be shared?

- [ ] Based on the purposes mentioned above, exactly what data needs to be shared?
  Note: Only share data that is absolutely necessary for the specific purpose.
  Note: If possible, only share anonymised data.

What category of data will be shared?

- [ ] Personal data?
- [ ] Sensitive personal data?
- [ ] Sensitive non-personal data?
- [ ] Sensitive data that is and currently anonymised but which could be re-identified in the future due to advances in technology?
- [ ] How will you contextualise, describe, or tag the data to avoid it being used out of context and in such a way that the assessment of the quality of the data can be documented?

What data laws do you need to follow?

- [ ] Is there a national data privacy law (or multiple laws) that should be followed?
- [ ] Based on privacy legislation, are you legally allowed to share this data?
- [ ] Does the data privacy law spell out specific roles for those who collect or handle data?
- [ ] For data that has already been collected, are the data subjects aware that their data will be shared and, if so, have they given consent?
- [ ] For data that is still to be collected, ensure that the consent makes it clear to data subjects that data will be shared and with whom.
- [ ] Will the data remain in the country where it is collected, or will it be shared or processed or stored in another country? Will the data cross borders at any point? Is cross-border data transfer of this type of data permitted? (For example, some countries do not allow health data to cross borders)
- [ ] Does the law mandate that you conduct a Privacy Impact Assessment to identify any potential harm that could come from collecting, processing, sharing, and using the data? The potential harm should be weighed against the benefits of sharing the data.

How exactly will the data be shared?

- [ ] Will the data be shared across two or more organisations? Or will one or more organisations share with one another (or a few others)? Or will one organisation dictate to the other(s) how and why the data will be collected (the data controller) and the other(s) will collect and process the data as directed? Or will all organisations collect and process data and then share with one another?
- [ ] How often will data be shared? Once only? Ongoing?
- [ ] How exactly will it be shared or transmitted and via which systems and processes?
- [ ] What security measures do the partner organisations have in place? (These should be assessed and verified).
- [ ] If you are collecting or processing sensitive information, be aware that you may be legally obligated to conduct a Privacy Impact Assessment.

Who will manage and be accountable for the data and how?

- [ ] Which organisation(s) will manage and maintain the data, the database and related systems?
- [ ] Has it been clearly defined who the data controller is and who the data processor is? (Each has specific responsibilities as defined by the law). Or will both organisations collect and use this data? (Joint data controllers and processors).

TIP SHEET 7 (continued)

- ☐ What systems are in place to secure and protect the data?
- ☐ Have you checked to ensure that your data and data systems are interoperable if both parties are collecting data, or if data is to be combined?
- ☐ How will data access be managed and controlled and by whom?
- ☐ What will happen at the conclusion of the project, data collection or partnership?
- ☐ Are there sufficient staff and financial resources to ensure data protection and security?
- ☐ How will any critical data incidents (leaks, unauthorised access, breaches, data loss) be handled and communicated and by whom?
- ☐ How will any requests for data removal, deletion, correction, etc. be handled? (e.g. if a person whose data has been collected withdraws or wants to change data or has some other question about their data)

Addressing these questions will assist in gathering the information necessary to develop and implement a data protection and/or data sharing agreement with a partner.

SEE SOME COMMON TYPES OF DATA POLICIES AND DOCUMENTS BELOW:

- *Data privacy policy* – an overall policy that an organisation follows to protect data (a high-level organisational policy that guides how data privacy will be carried out – normally accompanied by specific procedures. See The Netherlands Red Cross's Policy and Oxfam Great Britain's Policy.
- *Data sharing agreement* – an agreement that specifies how organisations will share data in secure ways, the responsibilities of each entity, and how accountability will be ensured. See this data sharing agreement between the UNHCR and the World Food Program.
- *Data processing agreement* – an agreement between a data controller and data processor that outlines how the processor will manage and secure the data. After the enactment of the GDPR, it became commonplace for survey companies to ask their clients to sign a data processing agreement. This link provides an example of a data processing agreement.

STAGE 6

Responsible M&E data visualisation and communication

SNAPSHOT

This chapter offers guidance on M&E communication, from identifying audience requirements to designing data visualisations that avoid biased interpretations and that meet specific objectives. It includes a review on developing a data visualisation and information on lessons learnt from COVID-19 visualisations. It also includes a tip sheet with specific data visualisation considerations for effective communication.

Visualising data and communicating about data are crucial components of the Data Lifecycle. If the message cannot be clearly conveyed to the target audience, there is little point in collecting the data. Visualisation is key for understanding complex issues, but it can also be the starting point for initiating debates and discussions and exploring new research questions. Visualisations can expand both the so-called 'island of knowledge' (a knowledge gain) as well as the 'shoreline of wonder' (prompting more questions).

While M&E practitioners have traditionally relied on written text, visualisations (e.g. graphs, drawings, infographics) are increasingly being recognised as valid and useful ways of summarising, presenting, and communicating development-related data. This trend has been influenced by the proliferation of user-friendly tools and greater global digitisation.

Responsible M&E is about ensuring that the data extracted from communities is fed back to different audiences, including to the community members themselves.

Guidance is provided below on the different steps that can be followed to ensure responsible and effective visualisation and communication to different audiences.

Guidelines for M&E data design and communication

6.1 Know your audience and what forms of communication are accessible and useful for them

An important first step in deciding on the type of visualisation relates to the target audience. It should be remembered that not everyone has the same level of data literacy and that cultural and other factors influence how well someone will understand a visualisation of data. As such, one of the first questions to ask when designing a visual is: "How data-literate is my audience?"

Data literacy is "the ability to read, work with, analyse and argue with data." A determination of the data literacy level of the target audience will assist to decide on the complexity of the visuals. It will also help to choose the approach – whether it should be author-driven or reader-driven.

Cultural factors also influence how data is perceived and should be considered when choosing an appropriate visual for a target audience. In particular, the use of symbols and colour is not universal and may be open to interpretation. While red can be used to emphasise an issue in one cultural context, it can be associated with a political party in another.

An **author-driven approach** displays data in a specific order, including a structured narrative with no possibility of interaction with the data. The advantage of such an approach is that the visuals are clear and there is little room for misinterpretation.

A **reader-driven approach** allows readers to interact with the data but provides no narrative or messaging.

In determining the suitability of visuals for an audience, it is also relevant to reflect on the following three contextual factors:

Not all users will be equally motivated and interested in seeing the data, so the visual needs to catch their attention and relate to their frames of reference. It is also important to keep in mind that a window of 5–10 minutes is often all the time available for capturing the attention of policymakers.[6]

See Stage 1 for more about how stakeholder considerations inform audience communication requirements.

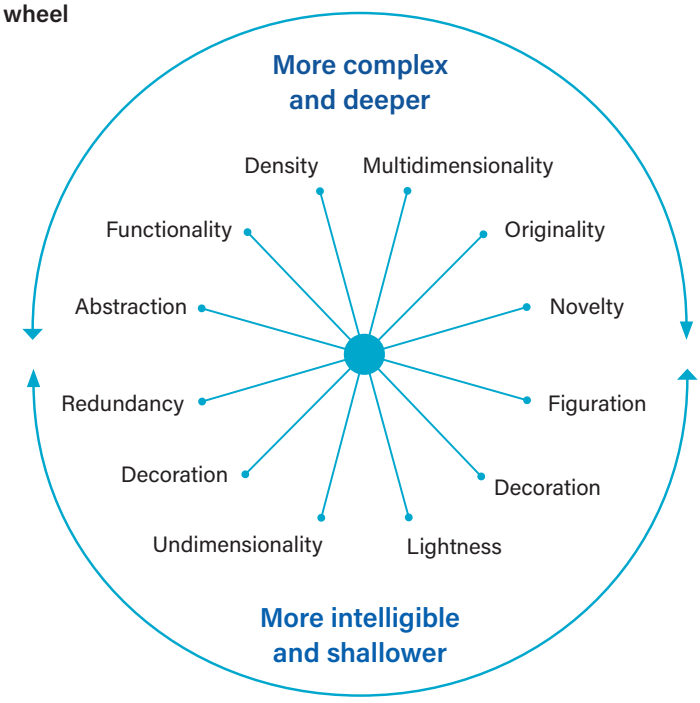
6.2 Make the 'right' design choices for your audience

When the specific interests and needs of your audience/s have been considered and identified, a decision can be made on your communication outputs. This means making visualisation and communication choices that best suit the audience (and not merely resorting to a personal preference or a generic one-size-fits-all approach).

Let us consider a scenario: *You have just shared the report of an evaluation for an education intervention that you have recently completed. The same report is shared with the executive director, funders and teachers involved in the intervention. What are the differences in the type of information they are looking for? The funder might be looking for a return on investment, the executive director may want to understand the impact of the programme and what to change, whereas the teacher may be looking for information on how his/her school compares to others. Variation in the type of information sought is common for different data outputs. It affects the actual design in the choices we make about what and how to present data. For example, the director may need a summary of key metrics and actions upfront, whereas the teacher may be looking for a comparative presentation of the metrics relevant to her role.*

Making poor choices for the data user can leave the audience confused, lead to misinterpretation of data, or cause people to lose interest in the information before the salient points are presented. The following two simple steps assist in making the right choices in data visualisation and communication: 1) determine the level of simplification of information needed (e.g. a detailed dashboard with many graphs vs. a bullet list of key findings); and 2) choose the most appropriate presentation elements to present the data (e.g. bar chart vs. tables). The visualisation wheel includes 12 dimensions of a visualisation that should be considered – Figure 5.

Figure 5: Visualisation wheel



6.3 Create an appropriate channel and medium

Following a decision on the appropriate content and design of your data visualisation or communication output, the next step is to devise the most effective way to convey the data information to your audience. Make sure to consider the varied needs of audiences. The channel to reach a management team of an NGO may be quite different to that required to reach a community member. The internet and digital technology have influenced the proliferation of channels for disseminating information, such as YouTube, Zoom, Facebook, etc. providing many choices of ways to communicate data to the person(s) that you want to influence. It is important to consider which channels your audience can access – is the internet the best channel or is sharing printed one-pagers more appropriate? Is your audience on social media? If so, how can you get your message across in a single post? The channel and medium may also place restrictions on design. You may not be able to share an interactive map on YouTube, but you could share videos of the views of the map.

The below table provides a few examples of media that might be chosen for different channels.

Channel>	**Print**	**Email**	**Website**	**Social media**
Medium>	Microsoft Word report in A4 layout	Mailchimp newsletter or report	Piktochart embedded infographic	Canva animated social media data posts

Finally, you must stress-test the sustainability of your output. How long must your audience have access to your report or infographic? What implications will this have on cost and access? To ensure long-term access to information, identify open source or low-cost media to maximise access and ensure longevity of the information's use/usefulness.

6.4 Clearly convey key lessons or messages

Stephen Few developed a useful data visualisation effectiveness profile as a means to assess different visuals based on two dimensions and seven criteria of data visualisations. The two dimensions are emotive and informative. Informative dimensions include usefulness, completeness, perceptibility, truthfulness, and intrusiveness. Emotive dimensions include aesthetics and engagement. Though all criteria are important, truthfulness is the only criterion on which no compromises should be made.

The emotive dimension is important in enhancing engagement with the information you share. However, the emotive dimension is often overplayed, thereby over-sensationalising data and serving to detract from constructive critique or a real understanding of the meaning of the data. If used correctly, aesthetic improvements can guide the reader to the key points of the data, facilitate the process of reading or mentally organising the data, and help to minimise extraneous elements surrounding the data. Prior sharing of draft visuals with the audience or with a colleague for their inputs will help to pinpoint aesthetic elements that are either confusing or missing from a visual.

See Tip Sheet 7: Visualisation improvements at the end of Stage 6

While visualisations can help to communicate data and improve our understanding of the world, they have their limitations. Some visualisations fail to provide useful data but, instead, leave the audience confused and frustrated. Moreover, false information can be visualised and communicated and thereafter may be wrongfully used for questionable purposes.

Misinterpreting findings or misunderstanding the messages depicted in visualisations is a common pitfall, especially when the audience has low levels of data literacy. It is the responsibility of M&E practitioners to support users of data – irrespective of their educational background – to clearly understand the messages being conveyed in visualisations. An important first step in this regard is to train communities in data literacy so that they can not only correctly comprehend the data but can also use it and, if they disagree, they can dispute it.

In addition, it is also necessary to create platforms that allow the audience to ask critical questions about the data and to promote discussions about the ways forward and possible recommendations in relation to the data. Examples of common platforms used to open such discussions include moderated chatrooms and online forums, conference presentations and panels, external reviews, and stakeholder feedback sessions with individuals or groups.

Some useful resources for data visualisation in the context of M&E include the Interactive Quantitative Chart Chooser, The Qualitative Chooser, this Infographic Guide, this Dashboard Guide, and other related guides found here.

TIP SHEET 8

Improving data visualisation

- [x] Simple aesthetic improvements to a visualisation can be made with aid of the following eight elements:
- [ ] **Alignment –** Generally your reader will read from left to right or top to bottom so structure or organise your data in this manner so that it is read in a logical pattern. Scattered visuals require the reader to organise and make assumptions about your storyline. If you make the flow easy it is easier to follow. Use gridlines as markers to check alignment.
- [ ] **De-clutter –** Remove any aspects of your visualisation that are not useful to the reader. For example, do not present gridlines if you have included data labels – labels are enough to understand the scale you are presenting so gridlines are just a distraction. Identify other unnecessary clutter in your visuals.
- [ ] **Font –** Use font differentiation only where it has a purpose, such as in headings or labels. As far as possible, limit your fonts to 3–4 types but keep them simple and related to the same family of fonts. Also check your organisation's brand standards and style guidelines in this regard.
- [ ] **Lines –** Lines are helpful in creating structure and grouping information and can be used as a visual paragraph to indicate the end of a thought and the start of a new one. This helps the reader to organise and process the data visualisation in a logical manner, thereby assisting the user to assimilate the data.
- [ ] **Arrows –** Where there are many components in the visualisation, arrows are useful in guiding the reader to relationships between data visualisations or to a specific point which will help interpret the visual.
- [ ] **Colour –** While colour can be very emotive, it is also subjective – e.g. generally we interpret red as bad, negative, or suggestive of danger. Check that your colour use is appropriate in emphasising a data point but does not distract the reader from more salient points. Also consider whether the colours you use can be seen by your reader. Might the audience include a colour-blind person? As with fonts, often your organisation may prescribe a certain palette which will guide your colour choices.
- [ ] **Text –** Use an appropriate amount of supporting text for your audience to provide context, guidance, or further details that are important in interpreting your visualisation. Too much text can be overwhelming, but too little could leave the reader with insufficient information to draw meaning from the visualisation.
- [ ] **Icons and images –** Include icons and images to help categorise information for the reader or convey the gravitas of the data you present. Icons and images can also provide quick insights into the context if you need to reduce the text – e.g. an image of young people is representative of the demographic that the data represents.

Results visualisation with different stakeholders

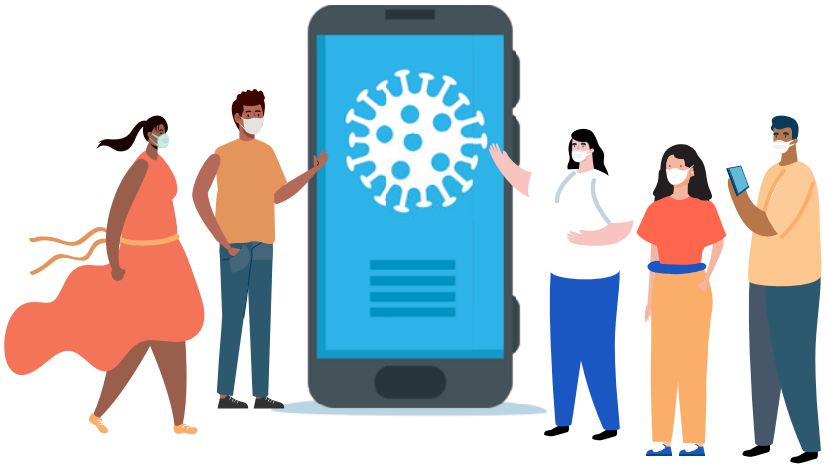
As part of the programme known as Leading, Teaching and Learning Together, VVOB Education for Development partnered with the research advisory firm Laterite to develop a digital data ecosystem, allowing different stakeholders ease of access to results on the impact of continuous professional development programmes for key education actors. In this case study, we describe how dashboards for this system were developed in accordance with the key considerations outlined on data visualisation and communication.

Knowing your audience: Users of the digital data ecosystem included a wide range of stakeholders, from lecturers currently training school leaders at the University of Rwanda, teachers, sector and district officials, and government officials. Most stakeholders were unfamiliar with using data for improvement and empowerment with the consequent risk that data could potentially be used a control mechanism.

Making the right choices for the audience: For the visualisation of data through dashboards, we opted for Power BI, a Microsoft application. The choice to use this platform was largely driven by its similarity to Excel, making it somewhat more familiar and user-friendly than other software, and also easier to adopt by VVOB and its partners. As the purpose of the data ecosystem is to make the M&E system more accessible to different stakeholders, different interactive dashboards are being created and we are still exploring whether, through protected pages, different data can be displayed for different users. Using static and interactive charts with brief explanations, a balance is being sought between an author- and a reader-driven approach.

Pilot testing the instruments and dashboards: To carefully weigh the needs of the different users of the data, the instruments are extensively pilot tested among users, and Key Performance Indicators (KPIs) are identified in collaboration with users, who include school leaders and district officers. For pilot testing, we chose a two-phased approach. Instruments and KPIs were first pilot tested internally among programme staff. After a first adjustment of the instruments and KPIs, the revised instruments, KPIs and dashboards were pilot tested among the end users through a series of workshops and online questionnaires. The aim of these workshops and questionnaires is to elicit feedback on the key data visualisation criteria (as developed by Stephen Few) – usefulness, completeness, perceptibility, truthfulness, intuitiveness, aesthetics, and engagement.

Ensuring key lessons are conveyed: It is extremely important that is that VVOB's partners can continue to monitor and evaluate future continuous professional development programmes for educators, and to ensure that data and assessments are used for improving school leadership, teaching, and learning, and are not used to control educators at a centralised level. We plan to train different stakeholders on general data literacy and, more specifically, on the use of instruments and dashboards. Different toolkits and protocols are being developed to ensure that a good data governance framework is in place and that sufficient attention is paid to data protection and privacy.



Lessons from COVID-19 data visualisation

COVID-19 has accelerated the level of public engagement in news and media. For example, during the COVID-19 lockdowns, private hospitals in South Africa shared their pandemic statistics with government for better planning and decision making. The pandemic has raised the awareness of governments on the importance of data and of representing data in ways that enable individuals and institutions to act. The popularised term 'flatten the curve' prompted the public to read data to check daily caseloads and compare the local situation to that in other countries. It emphasised the need to be sensitive to all consumers of data and, in particular, the general public – not only other technical specialists, as has tended to be the case historically.

The Data Viz Society hosted a webinar in 2020, on the topic "Visualising health data responsibly". Key lessons shared were, first, that in an effort to simplify data, we risk oversimplifying it and, in so doing, provide too little data for relevant action. All narratives based on data should be comprehensive and remain true to the individuals involved or to the nuances represented by the data. The second lesson related to visual elements such as colour, shape and scale. This is evidenced by the headlines used when presenting COVID-19 data. These have sought to create drama – large red bubbles, headlines proclaiming that 'older people are at risk' and the indiscriminate sharing of personal data the early stages of the pandemic. These tactics draw attention to information but risk creating fear amongst the readers rather than encouraging appropriate action.

COVID-19 data visualisation has emphasised the need to consider both the technical and human-centred angles as essential ethical considerations when presenting data.

Chiqui Esteban's 2015 article, A Quick Guide to Spotting Graphics That Lie, provides further examples of the potential of poorly visualised data resulting in deception and misinterpretation of the data. Esteban lists five 'lies':

- Broken scales show drama where none exists.
- Showing data on two different scales can make for an apples-with-oranges comparison.
- Showing a correlation can imply causation.
- Ignoring population size makes rates impossible to compare.
- Decoration can be deceiving.

When creating visualisations, it is important to test or self-check whether the data presented may inadvertently generate fear and confusion as has been seen in some COVID-19 visualisations, and whether the design subsequently conveys 'lies.' As misinformation and disinformation become common and are widely spread throughout digital media, the need for responsible data visualisation becomes ever pressing.

STAGE 7

Responsible data retention, maintenance and destruction

SNAPSHOT

This section emphasises the importance of managing personal and sensitive data from the 'birth' of data (during collection) to its 'death' (when we aggregate it, anonymise it, or delete it). It includes a Tip Sheet on developing a data retention policy.

A principal way to minimise the risk of data misuse is to minimise the amount of data that is collected, and to formulate a clear plan for how data will be maintained, retained, or destroyed. This may include steps to anonymise data, periodic data aggregation, or destruction of data after a particular point where it is no longer needed.

Data retention policies should reflect the fundamental principle of 'data minimisation', meaning that only data which is required for specific purpose is captured; data is kept for as short a period as necessary; and data is kept as securely as possible at all times. Data minimisation is a core principle of most data privacy regulations.

Guidelines for developing and implementing a data retention policy

A data retention policy is based on the diverse types or categories of data that are being collected, how long each kind of data should be stored, and when and how it should be anonymised or deleted.

To formulate a data retention policy, review the agreements with other entities (donors, partners, ethical committee) and any local laws that set terms for the retention of data or records (for example, you might be bound by law to keep financial records for seven years). Review how long you need to retain various kinds of personal, sensitive, and sensitive non-personal data and establish a reasonable timeframe for keeping raw data, anonymising or aggregating it, and deleting it. It is important to remember to include any datasets that are stored in cloud services like KoBo Toolbox or Survey CTO, and to have a schedule for destroying paper-based surveys and other hard-copy data.

For example:

- If you are doing longitudinal qualitative analysis, it might be prudent to keep data in its original form for a decade or more, or even in perpetuity if the data is historical data.
- You might keep contact information of survey respondents for a 3-year evaluation for a total of 5 years in case you need to return to them for a project evaluation.
- If you are collecting quantitative data, it should be aggregated quickly and the raw files destroyed.
- If you are doing web analytics, you might only need detailed data for a week or only need data in aggregate form.
- You might have to retain some personal data for audit requirements from a grantor or partner that requires data to be kept for a specific period.

How to develop a data retention policy

A data retention policy could include the following content:

- What is the purpose of the policy?
- Who must follow the policy?
- Who is responsible/accountable for administering the policy and ensuring compliance?
- A note that you might, for certain legal reasons, be forced to keep data longer.
- A list of the types of data included in the policy.
- The period for which each type of data will be retained (see more information on this below).
- Any specific orientation on how the data will be treated while retained (e.g. password-protected, only accessible by role, etc.).
- Any specific orientation on how data will be anonymised or aggregated.
- Any specific orientation on how data will be destroyed.

Based on the policy formulated, it is advisable that data in the system is flagged by category, so that you receive automatic notifications indicating when it is necessary to aggregate or delete data. Data can then be reviewed periodically based on the automated notifications (not auto deleted) and data that is no longer needed can be deleted or aggregated.

In addition to the data retention plan, processes should be put in place to ensure that the plan is followed and someone should be assigned the responsibility of stewarding the data along its lifecycle – follow the link for a sample data retention policy.

Align your data retention policy with your informed consent process and other privacy-related policies

As can be seen from the description above, data retention plans are the basis of consent processes, terms and conditions, privacy policies, or any other information provided to data subjects about the use of their data. The plans should also guide how data is handled and stored and should provide details on who is authorised to access the data and any systems that manage the data.

See Stage 2 for more orientation on consent

At the start of any new partnership or initiative involving data, it is important to establish and document appropriate data retention periods and these should be based on the following:

- Data source (from whom/where is the data collected).
- Type of data.
- Reason for keeping data.
- Personal data or sensitive data.
- Will data be aggregated or de-identified or anonymised and when will this happen? (if it is not subject to a retention period).
- Where and how will the data be stored?
- How long will the data be held?
- How will data be destroyed?

Ensure that you have a data management plan for the termination of your programme or M&E initiative.

The absence of a plan on how to terminate a data collection exercise risks that the data can be left unsecured or improperly disposed of, creating an opening for a data breach or other type of data misuse. Therefore, it is important to have a plan from the start on how a project will terminate, whether it will be sustained, and what will happen to any data collected as part of the project or M&E effort. If you do not plan to delete data at the conclusion of the project, you will need to decide on and allocate a budget for maintaining the data and keeping it secure.

See Guidelines for designing and planning for M&E in Stage 1

ENDNOTES

1 Canadian Institutes of Health Research, Natural Sciences and Engineering Research Council of Canada, and Social Sciences and Humanities Research Council. Tri-Council Policy Statement: Ethical Conduct for Research Involving Humans. 2018 Dec. Catalogue No: RR4-2/2019E-PDF. ISBN: 978-0-660-29942-6.

2 Facca, D., Smith, M.J., Shelley, J., Lizotte, D. and Donelle, L. (2020). "Exploring the ethical issues in research using digital data collection strategies with minors: A scoping review" https://doi.org/10.1371/journal.pone.0237875

3 Schwab-Reese L.M., Hovdestad W., Tonmyr L. and Fluke J. "The potential use of social media and other internet-related data and communications for child maltreatment surveillance and epidemiological research: Scoping review and recommendations". Child Abuse Negl.2018;85:187–201.

4 Clearance may be needed from the Commission for Science and Technology (COSTECH) of Tanzania.

5 The Kenyan National Commission for Science, Technology and Innovation may need to approve research.

6 Steele, J, and N. Iliinsky (eds.), (2010) "Beautiful Visualization: Looking at Data Through the Eyes of Experts" Cambridge: O'Reilly. https://archive.org/details/isbn_9781449379865

NOTES

NOTES

CLEAR-AA

Centre for Learning on Evaluation and Results – Anglophone Africa

The Oval Building
University of the Witwatersrand
2 St David's Place, Parktown,
Johannesburg

Telephone: +27 11 717 3157
Fax: +27 86 765 5860
Email: CLEAR.AnglophoneAfrica@wits.ac.za

www.wits.ac.za/clear-aa

MERL Tech

UNIVERSITY OF THE WITWATERSRAND, JOHANNESBURG

clear Centers for Learning on Evaluation and Results ANGLOPHONE AFRICA