GUIDE TO GOOD PRACTICE FOR COLLECTING AND MANAGING MEDICOSOCIAL DATA

june
2021

MÉDECINS DU MONDE

Graphic design :

Adriana Lyra

If you have any question regarding this guide, please get in touch with the Research and Knowledge Management Unit via the following address:

PoleRechercheApprentissages@medecinsdumonde.net

You can also consult the Intranet page dedicated to Data Management.

INTRODUCTION AND CONTEXT

This handbook of good practice is addressed at those Médecins du Monde actors who are involved in collecting, managing or analysing medical and social data or sensitive personal data more generally. Its objective is to help improve the data management culture within the organisation by offering practical solutions to the key questions of quality and security associated with data management.

This document is deliberately general in scope in order to remain relevant for the greatest number of scenarios. In some instances therefore, reading this guide will not be enough to resolve specific operational problems. Do not hesitate to contact staff at head office – the Research and Knowledge Management Unit, Legal Department Data Protection Officer, French Observatory and Health Advisors and Technical Advisers, for example – for solutions designed to suit every situation.

This guide and document are part of the work currently underway relating to the Medicosocial Data Management Project[1]. This project will help us employ the same language in all field situations by providing the same data management references for :

1. Ensuring the medical and social follow-up for people on MdM France programmes,
2. Facilitating and ensuring the reliability of data collection and management for project monitoring,
3. Improving care coordination, at home and abroad.

The aim of the project is to harmonise the process of data management and quality assurance. It falls within the Horizon 2025 Transformation Plan[2], which sets out the ways Médecins du Monde France needs and wishes to develop, notably by acquiring up-to-date, effective and powerful tools in order to achieve greater impact and establish a global steering mechanism across the organisation.

[1] For more information on the Medical and Social Data Management Project, see **the intranet page concerned.**

[2] For more information on the Horizon 2025 Transformation Plan, see **the intranet page concerned.**

Data quality is crucial for Médecins du Monde, as it helps improve the quality of the actions we take and the care we deliver on our projects for the benefit of the people concerned. It also strengthens and increases the impact of our advocacy aimed at securing access to rights and healthcare. The data collected must be of the highest possible quality in order to increase the organisation's credibility among service users, donors and partners for example, and to ensure we steer our projects effectively via monitoring, evaluation and capitalisation. Data quality is measured according to the following seven criteria:

- **Validity:** The data measure and describe what it is we want to measure and describe.
- **Reliability:** The data are collected in a consistent manner.
- **Accuracy:** The data are sufficiently detailed to explain the phenomena studied.
- **Completeness:** All the data scheduled for collection are collected.
- **Timeliness:** The data collected are representative of the moment at which they are collected.
- **Integrity:** The data are precise and protected from any form of manipulation designed to distort the original source.
- **Uniqueness:** A single piece of data or unit of observation must not appear more than once in the quantitative database.

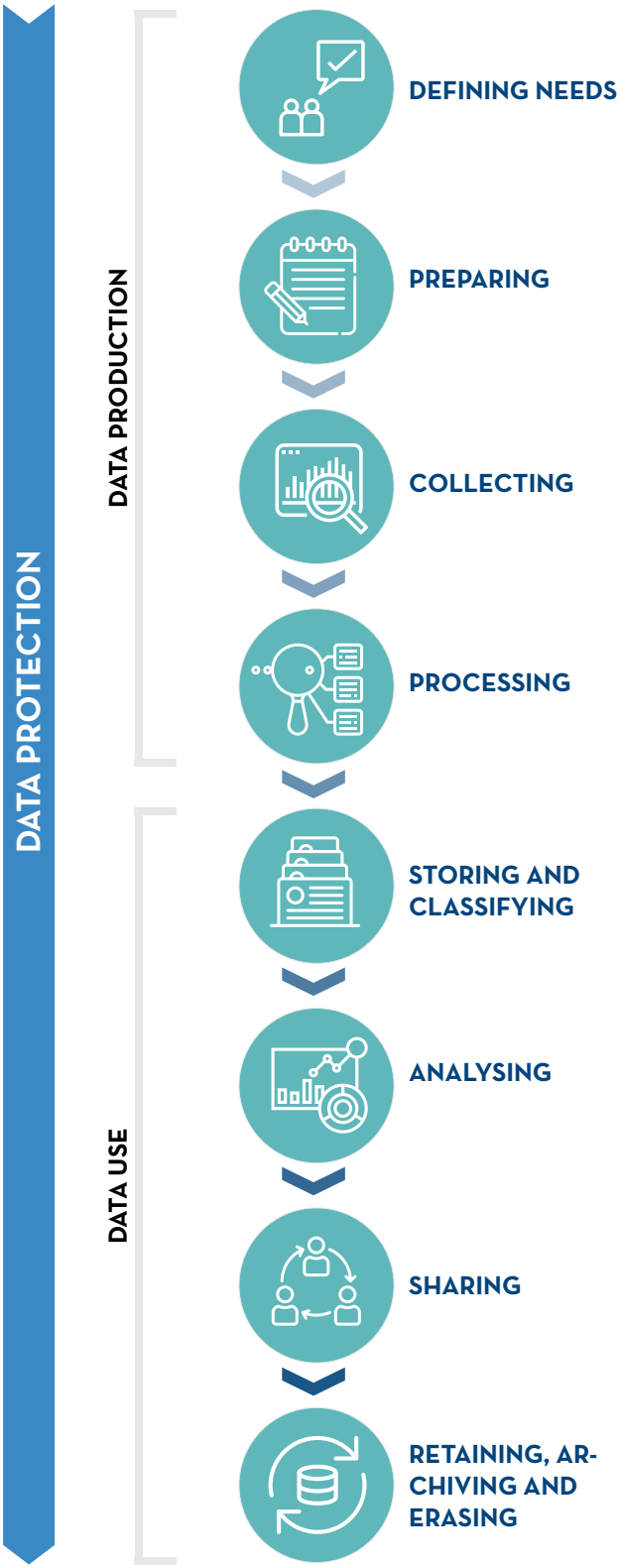
The other key issue concerns data security. This is particularly important when sensitive personal data are collected and retained. In some cases, the circulation of sensitive personal data could significantly harm the individuals concerned, and this would go against the principle we uphold of seeking to protect people. In addition, a strict legal framework governs the processing of these data (General Data Protection Regulation – GDPR[3]) and any failure to observe the legislation would expose Médecins du Monde to reputational damage and hefty administrative and financial penalties. Medicosocial data fall within the definition of sensitive data and must therefore be handled with great care and attention. Data protection is a responsibility shared jointly by everyone involved in the data-management cycle. Whatever the resources and infrastructure put in place to secure data, they are ultimately protected by data users' good practice.

The format of this document follows the different key stages in data management from defining needs to erasing data (see diagram below). The good practice aimed at achieving the desired quality and security objectives is set out in stages. Each element of good practice is categorised according to whether it is a requirement and must be respected or whether it is advisory and therefore (strongly) recommended.

At the present time, tools and practice may vary between our operations in France and international operations. Although the majority of the recommendations found in this guide apply to all fields of operation, an attempt will be made to clearly identify the sections that apply solely to international activities and vice versa.

[3] GDPR is the legal text governing the processing of personal data within the territories of the European Union. For more information, visit the **CNIL website**.

Stages of Data and Information Management



Note that **Médecins du Monde** is currently working on developing a new internal tool for managing medicosocial data. Its objectives include improving the quality and security of data collected, harmonising practice and simplifying the work of actors in the field. This guide must therefore evolve to take account of the specifics of this new tool once it becomes available.[4]

Useful resources:

- Health Project Planning
- For Ethics in the Field: Sensitive Personal Data Management
- Data Collection – Quantitative Methods
- Data Collection – Qualitative Methods
- Note clarifying the concept of accountability at Médecins du Monde
- Practical Guide to Satisfaction Study
- Guidance note : GBV Data Management
- MdM Charter for Ethical Research

[4] For more information on the Medicosocial Data Management Project, see **the intranet page concerned**.

CONTENTS

PLANNING PHASE

PLANNING PHASE

Required practice

Define needs

Only data with a **precise** and **specific** objective should be gathered.

The reasons for this are practical. Data collection is a potentially long, difficult and costly process in terms of human and material resources, and so it is better to focus on what is really useful and to avoid spreading one's resources thinly by gathering information that serves no purpose. The reasons are also legal: the GDPR purpose limitation principle (see annex) states that data can only be collected and used for a precise objective. Collecting data without a specifically identified purpose is therefore not permitted. Likewise, the principle of data minimisation (proportionality and relevance) stresses the need to collect only those data which are strictly necessary for achieving an objective.

As a result, within the context of research, it is helpful to draw up an **Analysis Plan** prior to collecting the data. Its purpose is to detail all the analyses that it is hoped to conduct of the data collected. This makes it easier to appreciate which data are really necessary to answer the questions posed.

For project monitoring, the **Monitoring Plan (Summary of Indicators and Data Flow Chart)** must be produced prior to data collection. Its aim is to describe in detail the indicators and how they will be used during the project as well as to identify the data required to calculate these indicators and how the data collection and entry will be organised. This enables only those data required for the project to be identified and collected.

In the case of personally identifiable data (surname, forename, postal and email addresses, telephone number, social security number, etc.), the recommendation is to record the minimum amount of information possible. Collect only what is strictly necessary, such as how to get in touch with an individual again for patient follow-up. The primary objective of such data collection is to enhance the quality of the service offered.

Useful resources :

- E-learning Monitoring, see Module 2.1. Developing specific Indicators using the Indicator Summary
- Data Collection – Quantitative Methods
- Data Collection – Qualitative Methods
- For Ethics in the Field: Sensitive Personal Data Management

Take account of ethical principles

When planning the data collection, it is important to consider key ethical principles such as **respect for individuals**[6], **the principle of beneficence**[7] (or generosity) and **the principle of justice** . This means ensuring that individuals are informed of their **rights** and are able to exercise them, that the data collecting does not expose them to unnecessary **risks** and that **no population group is excluded**. If any risks associated with the data collecting are identified, a system for monitoring their evolution must be put in place along with a strategy to ensure these risks are minimised. It is therefore recommended that a

[6] Respect for the autonomy and self-determination of participants and for the protection of those who are not autonomous, in particular by offering protection from dangers and abuses.

[7] Often referred to as 'Do no harm', this means the decision to collect the data must be taken after questions are posed about the risks faced by the human subjects contributing to the data collection.

The negative effects experienced by the subjects may vary widely in nature and extent. A non-exhaustive list of examples includes:

- The inconvenience of having to spend time responding to a questionnaire,
- The loss of earnings from having to spend time responding to a questionnaire,
- The psychological or emotional impact of having to respond to questions on the subject of traumatic episodes,
- The risk of suffering discrimination or violence in the event that sensitive information is leaked.

It is clear that zero risk is unattainable. However, it is the res-

data risk assessment table is completed relating to the data concerned (see the model table in the resources given below).

Depending on the circumstances – such as the type of data being collected and their purpose – approval for collecting the data may be required from a national ethics committee. It is important to find out whether an ethics committee exists in the country in question, and how it works. This must be taken into account when timetabling and budgeting. In the case of operational research, it is essential to refer to MdM's charter on research ethics as well.

Useful resources:

- For Ethics in the Field – Sensitive Personal data Management
- Charter for Ethical Research at Médecins du Monde
- Data risk assessment template

ponsibility of the team in charge of collecting the data to minimise the risks run by the subjects. Moreover, it is important to consider whether the benefits of the data collection outweigh the risks incurred. If this is not the case, then the data collection must be reconsidered.

[8] This involves ensuring that the data collected are representative of all population groups so that no potential beneficiaries, whoever they might be, are excluded, and ensuring that the benefits and costs are shared equitably. This is referred to as 'Leave no one behind'.

[9] See annex

Obtain consent

Free and informed consent on the part of the individuals concerned is **obligatory** for sensitive personal data to be collected and processed.

To be valid, individuals' consent must be:

- **Free :** This means it must not be influenced or forced. For example, the individuals concerned must not run the risk of negative consequences (such as exclusion from a project or denial of access to a form of assistance) should they refuse, nor be promised an advantage should they agree.
- **Specific :** Consent does not authorise Médecins du Monde to make free use of the data collected. The processing – analysis, storage, transfer and archiving – of the data collected must be made clear along with their purpose, so that individuals know what they are consenting to.
- **Informed :** The individual must have access to all the information required to make a decision in full knowledge of the facts. This includes the identity of the people in charge of data processing and their respective responsibilities, and the rights of the individual (the right to refuse to respond, the right to withdraw from the study at any point, and the right to access, erase or modify their data[9]). This information must be given in a language that the individual understands. Furthermore, the individual must be told how to contact a person in charge in the event of any questions.
- **Unequivocal :** Individuals must give their consent clearly and explicitly. There must be no ambiguity. For example, where a form has to be completed, the consent box must not be pre-ticked to ensure that the individual actively gives their consent.

If the individual questioned is unable to give consent – for example, as a dependant minor, person with a physical disability or learning difficulty or a person in

detention – the formal consent of their legal guardian must be obtained in addition to their own. In certain cases, the consent of a body with the required authority may replace that of the legal guardian.

Useful resources:

- For Ethics in the Field – Sensitive Personal data Management
- Charter for Ethical Research at Médecins du Monde

Recommended practice

Devise measures in advance to protect privacy

To **anticipate risks** rather than incur them, measures must be put in place to protect personal and sensitive data prior to their collection.

These measures must be **preventative** rather than corrective, meaning security failures must be anticipated, such as mislaid/stolen paper forms, intercepts of emailed data, unauthorised access on shared platforms and recovery from lost/stolen hardware. Everything must be done to avoid these situations. The use of a **data risk assessment table** is recommended to identify them.

The measures might involve, for example, keeping forms in locked cupboards, the systematic encrypting of files containing sensitive data, separating directly identifiable data from analysis data or restricting access to dossiers containing sensitive data to a minimum number of people.

In so far as is possible, the measures must also be systematic – digitised or regularly written up in the team members' diary – to avoid omissions on the part of the team in charge of data management. These measures concern every stage of data management without exception, from collection to erasure and including storage, cleaning, analysis, sharing and archiving. Unforeseen problems may arise during the data-management cycle, however. Therefore it is recommended that a seamless **communication system** be set up for alerting the data-protection staff responsible, so that they can speedily take corrective action. In the event of a confirmed data leak, Médecins du Monde is responsible for informing CNIL and/or the competent local authorities within seventy-two hours and must also inform the persons affected if the leak poses a risk to their safety or rights.

Ideally, all these measures are compiled and retained in a **Data Management Plan,** which serves as a route map for managing data. In cases where the data collected are particularly sensitive and/or involve an extreme risk, consideration should be given to producing a Privacy Impact Assessment (impact analysis relating to data protection) and to submitting it to the French information commissioner's office – the Commission Nationale de l'Informatique et des Libertés (CNIL). Contact the DPO at Médecins du Monde if in doubt.

Useful resources:

- Data risk assessment template

Establish an appropriate user-identification coding system

When individuals, whose data is to be collected, are met for the first time, they must be assigned a **code** which will identify them and their follow-up over time within an anonymised database. The code must therefore be **unique** to each individual (two service users cannot have the same code), **anonymous**, and, ideally, **easy to remember** for the service user. The code is always included in the databases containing sensitive information about the service user and therefore it

must not contain any information that would lead to an individual being identified.

The code provides the link between the **database for analysis** (containing all information needed to track service users, activities, and so on, but not directly identifying an individual) and the **correspondence register** (containing individuals' names and other identifying information such as telephone number or email address). The correspondence register is particularly sensitive and consequently must be protected. Only one or two people should be authorised to consult it, a record must be kept of each time it is accessed and the file must always be encrypted (see page 23).

Choose a suitable collection tool

Before beginning to choose and/or put in place a data-collection tool, remember that Medicosocial Data Management Project[10] is underway to respond to MdM's data-collection and analysis needs, particularly Health Data Hosting certification in France (HDS). This future internal tool for managing medical and social data will become the norm once available and thus is intended to replace all the tools currently in use. [11] It is therefore recommended that no new ad hoc tools should be introduced at this stage, unless absolutely essential, and that one of the existing tools available should be chosen instead.

Put simply, there are two types of tools: **so-called paper tools** and **mobile data collection tools** (electronic forms – such as KoboToolBox, ODK, SurveyCTO, Sphinx, etc.). **In France**, only the national electronic medical record system – the Dossier Patient Informatisé (DPI) – and Sphinx may be used to collect personal data.

Paper forms have the advantage of being easy to create and require no technical skill on the part of the user. They are more likely to produce errors, however, and considerable amounts of time are required for entering the data into an electronic database (such as Excel or the Monitool). There is also the risk of losing data should forms be mislaid.

Mobile data collection requires a bit more technical knowledge on the part of the team to develop and administer the tool over time, with attention needing to be paid to any skills loss due to team staff turnover. Likewise, service users need to be able to use a tablet or smartphone. It does allow for the creation of more efficient collection tools and ensures greater quality control of the data. Such systems also avoid the cumbersome storage of vast quantities of paper, help reduce the risk of data being lost, limit the risk of data-entry errors, enable the data to be exported directly to Excel, provide a more secure storage environment and can be directly analysed as desired.

Before launching the data collection, take the time to weigh up the pros and cons of each type of tool to be sure of choosing the best possible one. The parameters to consider include volume and type of data, purpose, sustainability of the tool used[12], data quality and security, cost, time and team skills.

In the context of routine data collection in health facilities, it is essential to prioritise use of any existing **National Health Information System (NHIS)** rather than developing and implementing a parallel data-collection tool.

Carefully programme the collection tool

A poor tool will invariably produce poor data.[13] A well designed form and well chosen, relevant questions are determining factors in collecting good quality data. A range of practice exists in relation to the tool used and type of data collected, but the following are good habits to adopt in the case of forms used to collect personal data:

[10] For more information on the Medicosocial Data Management Project, see **the intranet page concerned.**

[11] This work is already underway involving a number of colleagues and possibly readers of this guide too.

- Always begin by asking for the individual's informed consent when dealing with personal data.
- If directly identifiable data is being collected, isolate these so that they can be easily separated from the rest of the data (Paper form: Fill out on a separate sheet that can be detached and can be linked to the remaining data via an identification code. Electronic form: In the variable names, indicate their identifying characters.)
- If possible, limit the use of open questions, which require more time and resources for them to be exploited.
- Draft explicit, comprehensive questions. Data collectors are required to stick to the exact wording and must not interpret or reformulate.
- Express the questions in a language that is understandable to the individuals and avoid creating any bias.
- Number the questions.
- Be consistent in the way the replies are formulated. For example, always use the same scale of satisfaction on a form, for example 1. Very satisfied, 2. Satisfied, 3. Somewhat satisfied, 4. Dissatisfied, 5. Very dissatisfied.
- In the case of an electronic form, place constraints on the answers (though do not do so to excess). For example, prevent an age above 120 years or a negative age being entered.
- Use logical links and skip patterns. For example, do not ask a question about children if the person has previously stated they do not have children.
- For electronic forms, give explicit variable names.

More advice on developing good quality quantitative questionnaires can be found on this webpage (external source).

[12] For data collection as part of the follow-up of a service (a consultation of any type), the tools chosen must demonstrate they are sustainable and will continue to be applicable once MdM has left.

[13] GIGO (garbage in, garbage out)

Test the collection tool before use

Before deploying a data-collection tool in the field, **test it as much as possible**. This avoids encountering problems during the collection which would require urgent corrective measures.

It is recommended, therefore, that the form be tested initially by the person responsible for designing it. The objective of this first test is to verify, for example, that all the questions have been included, that the constraints and logical links are functioning where appropriate and there are no 'bugs' in the system.

A member of the project team should then test the tool. A fresh eye can often detect errors that are invisible to the person who has programmed the tool. This test pinpoints such problems as obvious errors, illogical sequencing of questions and excessive constraints. .

In the context of a survey, the form also needs to be tested in real-life conditions. This involves asking someone who will be part of the collection team to test it on a few people presenting the same characteristics as the target population. The aim is to reveal whether the tool is easy for the data collectors to use and whether the questions are properly understood and are acceptable to the respondents. When qualitative data is being collected, these tests determine whether the interview guides/observation tables are clear to the data collectors, and whether the information they produce is both faithful to its source and useable for analysis.

The feedback collected during the various test phases must be used to adjust the tool before it is generally deployed.

A particular scenario involves collecting routine data using the NHIS. It is strongly recommended that this system be used in all instances, even if it appears an inefficient tool, so that the data obtained match those of other organisations on the ground. The tool must not be adjusted, therefore.

Useful resources:

- Data Collection – Quantitative Methods
- Practical Guide to Satisfaction Surveys

Take time to train the data collectors

The quality of the data collected depends in large part on the people in charge of collecting them. Whether it concerns members of the MdM team responsible for implementing the project activities or external staff contracted to do the collecting, their training must on no account be overlooked.

The purpose of the training is:

- To familiarise the data collectors with the subject of the project.
- To familiarise the data collectors with the tool and the form. At the end of training, the data collectors must be very comfortable with the data-collection form. They must have no doubts about the meaning of any word or question.
- To make the data collectors aware of the ethical principles involved, such as the rights of the individual and informed consent.
- To teach the correct attitude to adopt towards individuals, such as respect for the person, neutrality and dealing with sensitive subjects.

In addition, a **confidentiality agreement** must be signed by the data collectors before starting to collect personal and/or sensitive data.

In the context of ongoing data collection (for monitoring purposes, for example), it is important that the training and support of data collectors is sustained over the long term, so that data quality remains consistently high throughout the project.

Useful resources:

- Data Collection – Quantitative Methods
- Data Collection – Qualitative Method

List of documents to produce during the planning phase

- **Data Collection Form:** Tool enabling primary data entry. Required for collecting quantitative data.
- **Interview guide and observation table (for qualitative data collection):** Tool enabling qualitative data entry.
- **Data Dictionary:** Preferably compiled using Excel, this database contains the name of each variable collected, their type, the list of categories of responses and associated constraints. Useful for data cleaning and analysis.
- **Data risk assessment table:** Compilation of security and confidentiality risks resulting from the data collection. Useful for anticipating risks and thus avoiding them more effectively.
- **Monitoring System (if monitoring):** Includes the logical framework, roles and responsibilities, Summary of Indicators, mapping of data sources and Data Flow Chart. See e-learning devoted to monitoring and monitoring manual for more details.
- **Data Management Plan:** Document listing the processes to follow throughout the data-management cycle. It contains the procedures to observe for processing, storing, sharing, archiving and erasing data. Its function is to ensure that good practice in managing data is adhered to throughout the project.
- **Analysis Plan:** Document detailing the analyses to be conducted on the data collected. Like the Data Flow Chart, its purpose is to ensure that sufficient data are collected to respond to the questions posed, and that each piece of data collected has a utility. It also serves to anticipate data needs so that everything required is collected.

DATA COLLECTION PHASE

DATA COLLECTION PHASE

Required practice

Protect data collection tools

Protecting the data includes protecting the collection tools. As these may temporarily or permanently house data, their loss, theft or hacking poses a major risk that the data will be leaked. To protect against this, the following are some of the good habits to adopt:

- Keep collection tools – paper forms or electronic tools – **closely monitored** at all times. When they are not being used, store them in a **locked** room/cabinet.
- Secure the data collection tools with a **strong password** (see the resource below). Configure them so that they lock automatically after a few minutes of inactivity.
- **Do not write down passwords** on a post-it, in a notebook, etc. or anywhere visible on the computer (on the desktop), where they could be seen by someone else.
- **Do not install unnecessary applications** or ones from unreliable sources on the digital devices.
- **Deactivate Wi-Fi and Bluetooth** on digital devices if not required. Avoid using public Wi-Fi as much as possible.
- **Never use an insecure USB key/external hard drive/CD** on digital devices. Never plug telephones/tablets into unknown computers. They might contain a virus that could infect digital devices.
- Ensure that the operating system and antivirus on digital devices are constantly updated.
- Save the data collected to the SIM/memory card rather than directly to the device. If possible, keep these cards separate from the device when not in use it to avoid the risk posed by theft.
- Activate GPS tracking on digital devices to locate them if lost or stolen. [14]

Useful resources:

- Guide to creating a safe password
- Best IT practice guideline

Conduct individual interviews in suitable conditions

When collecting personal data from individuals, it is important to ensure that no third party can witness or overhear the information divulged during the interview. Sensitive data could be leaked in this way. Furthermore, if a person is not comfortable with discussing sensitive subjects during an interview, the information provided might not be accurate.

Absolute confidentiality is never attainable, so try at least to favour an interview location chosen by the individuals concerned so that they feel at ease.

Similarly, the identity of the data collector can also affect the quality of the data. An individual will not necessarily feel at ease replying to certain questions if they are posed by someone of the opposite sex, of a different ethnicity or who is considerably older/younger, for example.

[14] Geolocation must never be used to track Médecins du Monde staff.

Recommended practice

Test data quality during the collection process

As data quality is a crucial issue, several stages during the data management cycle are designed for quality assurance purposes. This is the role of good programming of the collection tool and training of the data collectors during the planning phase. Regular testing of data quality is recommended during collection so that any problems revealed – such as a defective collection tool, poor understanding of the questions or poor quality work on the part of one or more data collectors – can be corrected.

This means data must be regularly entered so that they can be scrutinised. This rapid analysis enables errors in the data to be detected, such as outliers [15] and illogical elements [16] . Any errors detected can then be fed back to the data collector(s) for correction. These analyses can also help identify a data collector whose work is poor, for example if a high proportion of the collector's forms have missing values. At that point consideration must be given to retraining the data collector concerned.

The person in charge of the collection can also pay impromptu visits to data collection sites to check, for example, that the collection tools are being used properly by the collectors and that people understand the questions correctly.

It is also recommended to check regularly whether the 6 data-quality criteria are being met. Here the example is used of data collected from service users at a healthcare facility.

- **Validity:** The data measure what they are expected to measure and are in the format expected. For example, when studying the behaviour of sex workers aged between 18 and 30 years, only those individuals who correspond to this definition should be considered. In addition, dates of birth are given in number rather than text format.
- **Reliability:** The data collected are not affected by the context of the collection. For example, a woman might find it more difficult to provide information on her sexual health when speaking to a man as opposed to another woman. It is essential to ensure, therefore, that conditions are right for people when the data are being collected.
- **Accuracy** The data are sufficiently detailed. For example, rather than merely indicating that condoms have been distributed, it is often preferable to know the exact number of condoms handed out to each service user. Equally, this means ensuring that the data faithfully represent reality, and tracking outliers is very useful in this regard.
- **Timeliness:** Is the feedback of data rapid and frequent enough to inform decision-making? Be sure to enter accurately the date on which each piece of data is collected.
- **Integrity:** The data must not be manipulated, deliberately or otherwise. Information which people have not communicated should never be assumed. For example, never assume that a woman has been the victim of violence if she does not say so, and, conversely, never assume that a statement by a patient about violence is not genuine.
- **Uniqueness:** Never enter the same piece of infor-

[15] Outliers are data so far removed from the expected values that a data-entry error may have occurred. For example, an individual aged 70 is included in data collected from young drug users, or an individual is recorded with more than 15 children. It is not always possible to spot errors, but surprising, absurd or inconsistent values should attract attention.

[16] Inconsistencies between several variables. For example, a pregnant man or a child who is older than his or her parents.

mation several times and do not enter the same data from one individual on several occasions. If possible, opt for entering data in a single, centralised database. Ensure no duplication occurs (although the same individual may in some instances be observed several times, for example if attending the same healthcare service more than once).

Useful resources:

- Guide to data cleaning in Excel

DATA CLEANING AND ANALYSIS PHASE

DATA CLEANING AND ANALYSIS PHASE

Required practice

Create an input mask if the data are to be re-transcribed in Excel

When paper forms are used that need to be entered in an Excel document for subsequent analysis or disaggregation and inclusion in the Monitool, for example, an **input mask** must be prepared in advance. This often **saves time** when entering the data and also **minimises errors** and **increases consistency** between units of observation. The following principles should also be followed:

- The first line contains variable names which must be explicit. Enter one data table per Excel sheet.
- Enter one piece of information per cell.
- Do not merge cells.

Abiding by these principles facilitates data analysis and the exporting of data from Excel to other tools.

Abiding by these principles facilitates data analysis and the exporting of data from Excel to other tools. In the context of a survey, **double entry** is also recommended. [17] This involves asking two people to enter the same data before comparing their entries. Where these differ, return to check the paper form to find the correct value. This minimises the risk of data-entry errors.

Useful resources:

- Guide to creating an input mask in Excel

Isolate directly identifiable data

Data that can directly identify an individual (including name, telephone number, email address and social security number) must be kept in a separate database – the **correspondence registe**r – from the rest of the data in **the anonymised analysis database**. Any link between the analysis database and the correspondence register should only be possible via the unique code assigned to each individual.

No identifying personal data can remain in the analysis database. Only a limited number of users should have access to the correspondence register.

Clean the data

Cleaning is the final stage of data quality assurance. It involves searching for manifest or suspected errors in the data. Although this step is similar to the quality testing stage during collection, the difference is that it is often too late to contact the data collectors in the event an error is suspected. It might be necessary therefore to make choices about how to resolve outliers or extreme data which risk distorting the analysis. The decision might be to retain the data as they are, replace them with a more plausible value (**requires justification!**) or erase the data in question.

During the cleaning process, it is very important:

- **Not to overwrite the raw database (primary data).** Save the cleaned database under a different name and retain the original database intact**.**
- **To document** the changes made to the data.

[17] This is not always possible as it consumes time and resources.

Useful resources:

- Guide to data cleaning in Excel

Analyse the data

The analysis involves transforming the data into **useful information for decision-making** and for **ensuring accountability** towards all project stakeholders. The information produced is also presented to the populations studied to highlight the importance of collecting data and to maintain their involvement.

Conduct only pre-planned analyses as set out in the Analysis Plan (cf. p. 14). The analyses must respond to needs specified in advance, such as tracking how the monitoring indicators are evolving and supporting the reporting process, advocacy, observation and patient/activity monitoring. What matters is not to waste time conducting analyses with no added value. All data collected must be analysed, and so no data should be collected that are not scheduled for analysis.

In practical terms, a good analysis includes the following steps:

- **Formulate assumptions:** These must be established when the Analysis Plan is produced and must describe what the data are intended to demonstrate.
- **Conduct a systematic, descriptive analysis of the indicators:** Calculate statistics such as the average, median, variance and maximum/minimum of the different variables present in a dataset to establish a sound basis to understanding the information collected.
- **Extract relevant information and share this with the team:** The person responsible for the analysis must present the relevant information in a way that is comprehensible and accessible to the rest of the team, so it can be used for decision-making. Therefore make as much use as possible of **visual representations** of the data in the form of diagrams, tables and maps, to make the data more appealing to those consulting them. Make sure that no identifiable piece of data is revealed in the visual depictions.
- **Interpret the data:** Seeing figures is one thing; understanding what they mean in relation to a real-life situation is another. Involving all project stakeholders in interpreting the resulting data is recommended to avoid incorrect conclusions being drawn.

Analysis is a crucial stage in the lifecycle of data. It allows the data to be transformed into information and subsequently used and highlighted. An entire guide could be devoted to data analysis but that is not the purpose of this document. For more information on this subject, refer to chapters 3.3.D and 3.3.E of the guide to Health Project Planning.

Useful resources:

- Health Project Planning
- E-learning Monitoring, Module 2.3. Using the results of monitoring

DATA STORAGE AND MAINTENANCE

DATA STORAGE AND MAINTENANCE

Please note that this phase does not chronologically follow the data cleaning and analysis phase. Protecting and storing data has to be considered throughout the lifecycle of the data – from before they are collected until they are erased.

Required practice

Store sensitive data securely

Sensitive paper data, even when anonymised, must always be kept **under lock and key** and access to them must be restricted to a limited number of people.

In the case of electronic data (Excel file, Word document, audio recording, digital image, etc.), avoid storing them solely locally, such as saved directly to a computer or hard disk. Data could be lost in the event of equipment breaking down, being hacked or becoming obsolete. Storing data in this way limits the possibilities of working collaboratively on them, especially as the advice is never to share sensitive data via USB key or email.

Médecins du Monde is actively developing a tool for collecting and managing medical and social data. The intention is to replace all tools currently in use and to harmonise practice within the organisation. In particular, the new tool will be capable of storing all data securely (with French Health Data Hosting certification) and will allow collaborative working in a user-friendly, risk-free environment.

While awaiting the rollout of this solution, the main storage platform available for use on **international programmes** is SharePoint (in France, it is strongly recommended that Sphinx and/or the electronic medical record system, Dossier Patient Informatisé, be used while awaiting the new tool). SharePoint does not offer optimal levels of data security and, once data are saved, **it is everyone's responsibility to adopt the good practice described in this guide to avoid the risk of users' sensitive data on projects being leaked.** This entails limiting the number of pieces of identifying personal or sensitive data collected, encrypting all files containing sensitive data, separating identifying data from sensitive data and restricting access to folders containing sensitive data exclusively to authorised project team members.

In addition, some countries prohibit personal data concerning their citizens from being stored beyond their borders. Where this is the case, online storage platforms such as SharePoint cannot be used. Local storage solutions must be prioritised instead.

Encrypt files containing sensitive data

When files are encrypted, they **cannot be read** by anyone who does not have the password and/or relevant key file. The free software AxCrypt and 7zip, used by Médecins du Monde, employ the AES-256 encryption method, which offers the optimum level of security. The encryption of files containing sensitive data is therefore a highly effective method of avoiding leaks in the event of equipment containing stored data being lost or stolen, or in the event of files being hacked. Encrypted files can be opened and used once locally decrypted (namely on a computer). Once a file has been used, it must always be re-encrypted.

Files containing sensitive data must always be encrypted before being saved on online platforms or sent by email (emailing is generally strongly discouraged).

Please note: Once encrypted, files can never be unlocked if the password is forgotten and/or the key file lost. It is crucial therefore for decryption details

to be carefully retained. It is recommended that at least two people know the passwords at any one time to limit the risk of these being forgotten. But avoid oversharing passwords.

Useful resources:

- Guide to encrypting data with AxCrypt

Keep directly identifiable data separate

As indicated above, the analysis data must be kept separate from directly identifiable personal data (correspondence register). Both types of files must be kept in separate folders and access to the correspondence register must be restricted to the minimum number of people.

However, it must still be possible to identify the data so that individuals may, on request, access, modify or erase their personal data.

Transfer data via secure methods

Sharing data is a risky phase of the data management process when sensitive data are most likely to be leaked. Close attention must be paid therefore to the methods used for sharing data between users and between sites.

In general, all data – and particularly those that are sensitive – must only be accessible to a minimum number of people within Médecins du Monde and, in principle, not shared outside the organisation. However, there are occasions when such sharing is necessary and, in these instances, steps should be taken to ensure that only genuinely useful information is shared. Depending on what is required by the person with whom the information is being shared, the data can be anonymised by removing all identifying personal data and sending only the analysis database. Alternatively, the data can be aggregated. For example, if a partner wants to know the number of women taking part in a project, simply calculate and send the total rather than sending the list of the women concerned.

A good method of sharing is to use online storage platforms such as SharePoint. Access to the data can then be granted to the person in question and permission removed once access is no longer required.

In principle, sending data by email is not a secure solution and so is strongly discouraged. If sensitive files do have to be emailed, it is essential to ensure that both the email and the individual files are encrypted. Related key files must be emailed separately, again encrypted. It is better to share person-to-person via a phone call. Under no circumstances may data be sent via social media networks or instant messaging services, such as Facebook, WhatsApp or Messenger.

Although **strongly discouraged,** the practice of physically sharing data via a USB key or CD, for example, is nonetheless possible. Data must be handed over in person using a secure, portable device belonging to Médecins du Monde and used specifically for the purpose. Never entrust a portable device to a third party and never send it by post.

Useful resources :

- Guidance note : GBV Data Management

Archive and delete data once past their use

A legal framework applies to the length of time data can be kept. As MdM is an NGO subject to French law, French legislation must be followed **no matter**

the field of operation. Data must be deleted or archived once they are no longer in use. This means extracting them from active databases and storing them in a dedicated folder with restricted access.

Medical or patient files must be kept for a clearly determined period of time. In the case of adults and children aged 8 and over, files must be kept for twenty years. For children under 8 years, files must be kept until they reach the age of twenty-eight. In addition, the files must be stored in conditions that ensure their confidentiality and integrity, whether in paper or electronic format. [19]

In the case of personal data that are not medical files, **the general rule is not to retain these once they are no longer needed** (for monitoring and research purposes, for example). If, however, these data might be useful in the future, they may be archived and kept longer for their statistical value once they have been **fully anonymised**.[19] Whatever the circumstances, the length of time personal data are to be retained and how they are to be processed – shared and anonymised – **must be indicated to individuals** prior to collection, at the point they give their consent.

The first stage involves determining which data to archive, how to archive them and for how long. It is important to establish a formal lifecycle for the documents and to evaluate the value of keeping them temporarily or indefinitely. It is worth constructing a reference framework or table for managing the archives so that it is always clear which data are currently archived, from what date and for how long, and what ultimately is to be done with them. This table must be regularly updated, and new data users must be made aware of it when they take up post.

Note that sending documents to the recycle bin is not enough to erase them. The recycle bin must be immediately emptied and file-cleaning software, such as CCleaner, used in the case of sensitive data. In addition, check that all copies of the data have been properly deleted – local version, online version and backup version on a removable disk, for example.

Useful resources :

- Guide pratique : Les durées de conservation. (CNIL, in French)
- Référentiel : Les durées de conservation. Recherches dans le domaine de la santé. (CNIL, in French)
- Référentiel : Les durées de conservation. Traitements dans le domaine de la santé (hors recherche). (CNIL, in French)

Back up data as often as possible

Keeping a copy of the data collected is recommended. This is particularly critical when data are saved locally, such as on a computer or hard disk, as any equipment failure would mean that data not previously backed up would be permanently lost.

Data are preferably backed up on secure external devices – such as USB key or external hard drive – or, failing that, on an approved platform.

It is recommended that data be backed up as often as possible to reduce the volume of information lost in the event that files go missing.

[19] Full anonymisation involves making it impossible for an individual to be identified on the basis of the data held. Erasing directly identifiable data – name, address, social security number, etc. – is not enough; the risk of indirect identification must also be considered. For example, knowing a person's gender, number of children, home village and profession might be enough to identify an individual if those specific details apply to just one person.

Keep a check on usage permissions

Create different levels of authorisation for users of electronic data – or control access to paper data. This can be done whatever the data-storage solution chosen. Depending on the solution adopted for storing data, it is generally recommended to differentiate between

- Consultation accounts that allow the user to access files,
- Modification accounts that allow the user to access and modify data or enter new data,
- Administrator accounts authorising the user to carry out the same actions as the modification account, plus enabling the user to change the roles of other users.

The same user can be assigned different roles depending on the file in question. Access must be withdrawn from any project team member who leaves or who has no further need of the data.

This control is crucial: too wide a group with access to the data increases the risk of a data leak, whether by accident, as a result of hacking or due to unauthorised external sharing, for example. These different levels of authorisation must be established even before a data collection is set up.

ANNEX

ANNEX

Definitions

Raw databases: A raw database contains data that has not been processed or manipulated in any way.

Anonymised analysis database: The anonymised analysis database contains all the data collected on individuals apart from those directly identifying them. The latter are kept in the correspondence register. The two databases can be linked using individuals' identification codes. The analysis database contains all the information useful for analysis, monitoring and research purposes, for example.

Data culture: This refers to all data-related knowledge and practice shared by every member of an organisation. Its purpose is to make everyone aware of the value of collecting and analysing data and of the importance of applying good practice and sharing common tools that enable everyone to make optimum use of the data available.

Data: A piece of data is defined as information that is accepted, known or acknowledged and that serves as the basis of reasoning. It is therefore a raw element that has been neither interpreted nor contextualised. Data can be quite different in nature: qualitative or quantitative, structured or unstructured, and can be from different sources.

Personal data: Personal data are those that can directly or indirectly identify individuals. For example, name, email address, telephone number, postal address, social security number and identity photo are all personal data. These data can be explicit or can be extrapolated by cross-referring the information. Care should be taken with written transcriptions which may contain clues to people's identities.

Sensitive data: Sensitive data fall into a special category of personal data. They are personal data whose handling involves a risk of harm to the individuals concerned. Such harm might take the form of discrimination, embarrassment or identity theft. The following are considered to be sensitive data:

- Ethnic or racial origin,
- Religious or philosophical beliefs,
- Genetic data,
- Political opinions,
- Membership of a trade union,
- ***Health***,
- Sexual life and orientation,
- Biometrical data that can identify a person,
- Offences and convictions.

Observation: An observation is a data point within a database. In a database on individuals, each column represents a variable (surname, forename, age, profession, sex, etc.), and each line represents an observation (individual 1, individual 2, individual 3, etc.).

Correspondence register: A database containing all data that can directly identify a person – surname, forename, address, telephone number and social security number, for example – and the identification code. It must be kept separately from all other, potentially sensitive data on the individuals concerned. If sensitive data were leaked, it would therefore not be possible to identify any individual. The register must be very secure and accessible to only a few people. It is used where necessary to re-identify individuals, for example if they ask to access or modify their data.

Variable: A variable is a piece of raw data observed or measured in different individuals in a population, which is likely to change from one individual to another or to change for one individual over time. Variables enable the indicators to be calculated.

GDPR principles

The key principles of GDPR to consider are as follows:

- Purpose limitation: Personal data can only be collected and used for a clear and legitimate purpose.
- Data minimisation (proportionality and relevance): Only those data which are relevant and necessary to the stated purpose may be collected.
- Storage limitation: Personal data may be stored for a pre-determined, limited period of time only.
- Integrity and confidentiality principle (security): The data controller is the guarantor of the security and confidentiality of the data.
- Individual rights: The rights of individuals in relation to their data include the right to access, modify or erase the data. These rights must be respected.

As Médecins du Monde is an organisation subject to French law, the principles of GDPR must be adhered to in all fields of operation, both in Europe and elsewhere, provided they do not conflict with local legislation. For further information on data protection or in the event of any query, consult the Legal Department webpage, the Data Protection Officer via the Médecins du Monde intranet or contact the Data Protection Officer directly.

Individual rights

Right of access

Individuals have the right to request a copy of their personal data held by Médecins du Monde.

Right to rectification

Individuals can ask to have inaccurate or incomplete personal data rectified. This requires individuals to prove their identity and to prove that the data are indeed inaccurate.

Right to erasure

In certain circumstances, individuals may request that their data be erased. This may be the case if the data are no longer useful to the project, if the individual withdraws consent, if the processing does not match what was indicated at the time consent was given or if the individual was a minor at the point when consent was given.

However, Médecins du Monde may legally refuse to erase data if such action conflicts with a legal obligation, such as keeping a patient file for a minimum of 20 years, or if the data are of public interest or for scientific research or statistical purposes.

NB: It is extremely important to document all requests to access, rectify and erase data. If in doubt as to the correct procedure, contact the DPO.


MÉDECINS
DU MONDE