



DATA COLLECTION

Dr HMP Herath

Senior lecturer

Ph.D. (USJ), RN(SLNC), B.Sc. (Nursing) Hons(UPSL), CTHE (KDU)

Department of Nursing & Midwifery

FAHS, KDU



Co-funded by the
Erasmus+ Programme
of the European Union

Reference number: 618596-EPP-1-2020-1-SE-EPPKA2-CBHE-JP
This publication [communication] reflects the views only of the authors, and the Commission cannot be held responsible for any use, which may be made of the information contained therein.



Objectives



Summarize the significance of planning data collection in advance of an evaluation.



Explain the concept of data triangulation and how it contributes to the robustness of evaluation findings.



Compare and contrast administrative data and monitoring data regarding their uses and limitations.



Interpret the role of Theory of Change in guiding data collection for effective evaluation.



Describe the factors to consider when developing survey questions.



Differentiate between quantitative and qualitative data and their respective uses.



Co-funded by the
Erasmus+ Programme
of the European Union

Reference number: 618596-EPP-1-2020-1-SE-EPPKA2-CBHE-JP
This publication [communication] reflects the views only of the authors, and the Commission cannot be held responsible for any use, which may be made of the information contained therein.





Data collection

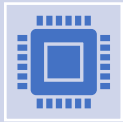
- Data collection is the process of gathering data for use in business decision-making, strategic planning, research and other purposes.
- Evaluations may be difficult, severely constrained, or excessively expensive if data collection is not planned alongside intervention development.



Data collection



Consider evaluation questions, data providers, and data access limits when arranging data collecting.



Administrative, monitoring, and large-scale (long-term) survey data are useful for evaluation. Long-term trend data prior to the evaluation can be very useful.



New research and data collecting are usually needed.



Minimising bias, verifying data collection techniques, and adding pre-planned data quality checks can improve data quality.





Introduction

- The collection of data is a crucial aspect of any evaluation and must be planned in advance.
- Planning data collection simultaneously with the development of an intervention ensures that data collection and data accessibility are incorporated into the policy's design and associated legislation.





Importance of the data collection planning

It is vital to remember that baseline data (data gathered prior to the intervention) must be obtained early, and comparative data (data from groups not affected by the intervention) needs to be discussed.

An evaluation may be impossible, severely limited, or in vain expensive if data collection or data access are not well planned.

If the data collecting process is poorly designed, wrong data may be obtained and false conclusions formed from the review.



The Theory of Change

- The Theory of Change can be used to identify data needs and gaps. A combination of the collection of existing and new data may often provide the most complete understanding and tracking of the Theory of Change and effectiveness of the intervention.
- Collecting data using several methods, such as monitoring data and bespoke surveys, builds confidence in the findings and the robustness to the evaluation; this is known as **triangulation**.
- In many cases, data relates to individual people but can relate to other units, e.g. schools, businesses, and geographical areas. Similar considerations apply in all cases.





Deciding what data is required

- The choices are not straightforward in determining the data required to answer the evaluation questions; input should be sought from those with relevant expertise.
- Next slides outlines the key questions to consider on data collection when planning an evaluation.



Deciding what data is required



Key Question	Considerations
What type of data will be required to answer each evaluation question?	What is needed to answer the evaluation questions identified during the scoping stage? Types of data include: <ul style="list-style-type: none">• Numerical data• Documentary data (data that has already been collected)• Observational data• Descriptions of people's experiences, opinions, and views• A combination / triangulation of the above evidence.
Who or what can provide this data?	<ul style="list-style-type: none">• Intervention participants• Service providers• Stakeholders• Databases• Existing surveys• Bespoke surveys



Deciding what data is required



Key Question	Considerations
Are there any issues with accessing or collecting the data?	<ul style="list-style-type: none">• Data access issues (e.g. legal issues, internal procedures, identifying the target group(s), collecting comparator data)• Data sensitivities/ethical issues (e.g. researching sensitive populations, data access issues)• The availability of necessary sampling frames• The potential for ‘data burden’ on respondents: is it proportionate to ask people to provide the data?• Who is responsible for data collection?• Is the data collection task proportionate to its value?• Are other methods possible?• Could the act of collecting the data from individuals influence their behaviour?



Deciding what data is required



Key Question	Considerations
What section of the population of interest should data be collected from?	<p>There are a number of ways to collect data from the population of interest. It is important to consider what is proportionate to answer the evaluation questions.</p> <p>Options include:</p> <ul style="list-style-type: none">• A census of all available data/populations of interest• A representative sample of the data available/population• A subset of the population of interest, purposively selected to cover a range of contexts but not statistically representative.



Deciding what data is required



Key Question	Considerations
How will the proposed data analysis method influence the data required?	<p>The type of analysis that will be applied to the data will typically dictate the nature and quantity of the data required.</p> <p>For example:</p> <ul style="list-style-type: none">• The required degree of precision• The sub-populations of interest (and therefore, the sample size required)• The nature of the method to be used and therefore, the data requirements• Whether baseline or control group data is necessary.



Sources of data



Existing administrative and monitoring data: data gathered either for the operation of an intervention, or for other functions.

Existing large-scale survey data: long-term, large-scale survey data, often managed by central governments.

New sources of data designed specifically for the evaluation: such data are typically obtained through methods such as surveys, qualitative method (interviews, observation, focus groups), web-scraping.

Social media data: this is a relatively new source of data, but potentially very rich in terms of gauging an unprompted reaction to an intervention.

Sources of data

Administrative data

- Administrative data is collected, stored and used primarily for administrative (nonresearch) purposes.
- Administrative data is typically collected on registrations, transactions and record keeping, usually during the delivery of a service.
- The availability of administrative data or data from an existing long-term survey should always be considered at the design stage of an evaluation.
- Both types of data can serve as important sources of background or explanatory data that predate the intervention giving excellent trend information.
- Where possible, the use of existing data reduces both the financial costs of the evaluation and the burden on respondents.





Issues of Administrative data

- Data may not be centred on the exact questions of interest.
- It is possible that they are not collected on a regular basis.
- It is possible that this is insufficient to provide useful evidence in response to the evaluation questions.
- Administrative data measure concepts that are similar but not identical to the measures interest..



Monitoring data

- Monitoring or performance-management data are collected during an intervention to address policy, research, and performance questions.
- Monitoring data typically includes information regarding service users, project inputs, processes, outputs, and outcomes.
- Monitoring data is often administrative and quantitative, not for evaluation.
- It is usually used to track an intervention's delivery or to determine where it is not being executed as intended and what additional action is needed to achieve its goals.



How monitoring data can be used in an evaluation



Monitoring Data	Example	How this can be used to evaluate
People accessing a service	Numbers and characteristics	Used to determine if an intervention is reaching its target population and how much, including population characteristics and non-target groups.
Inputs	Funding, resources, staff numbers	Checks if inputs meet expectations. Used to inform cost-benefit analysis and test policy implementation assumptions like cost and time.



How monitoring data can be used in an evaluation



Monitoring Data	Example	How this can be used to evaluate
Processes/activities	Referrals, waiting times	Used to determine whether the policy is being implemented as planned or whether there are any unintended consequences.
Outputs	Numbers going through a programme	Used to inform an assessment of whether the programme has delivered the target outputs to the anticipated quality.
Outcomes	Employment rates, wages	Used to measure the benefits of delivering the outputs.





Developing a monitoring system

Please read page 57,58
HMT_Magenda book



Co-funded by the
Erasmus+ Programme
of the European Union

Reference number: 618596-EPP-1-2020-1-SE-EPPKA2-CBHE-JP
This publication [communication] reflects the views only of the authors, and the Commission cannot be held responsible for any use, which may be made of the information contained therein.



Survey data



Government organizations and academic institutions collect large-scale survey data for statistical and research objectives.



Large-scale surveys, produce robust estimates due to large representative samples.



These surveys may be cheaper than new data.

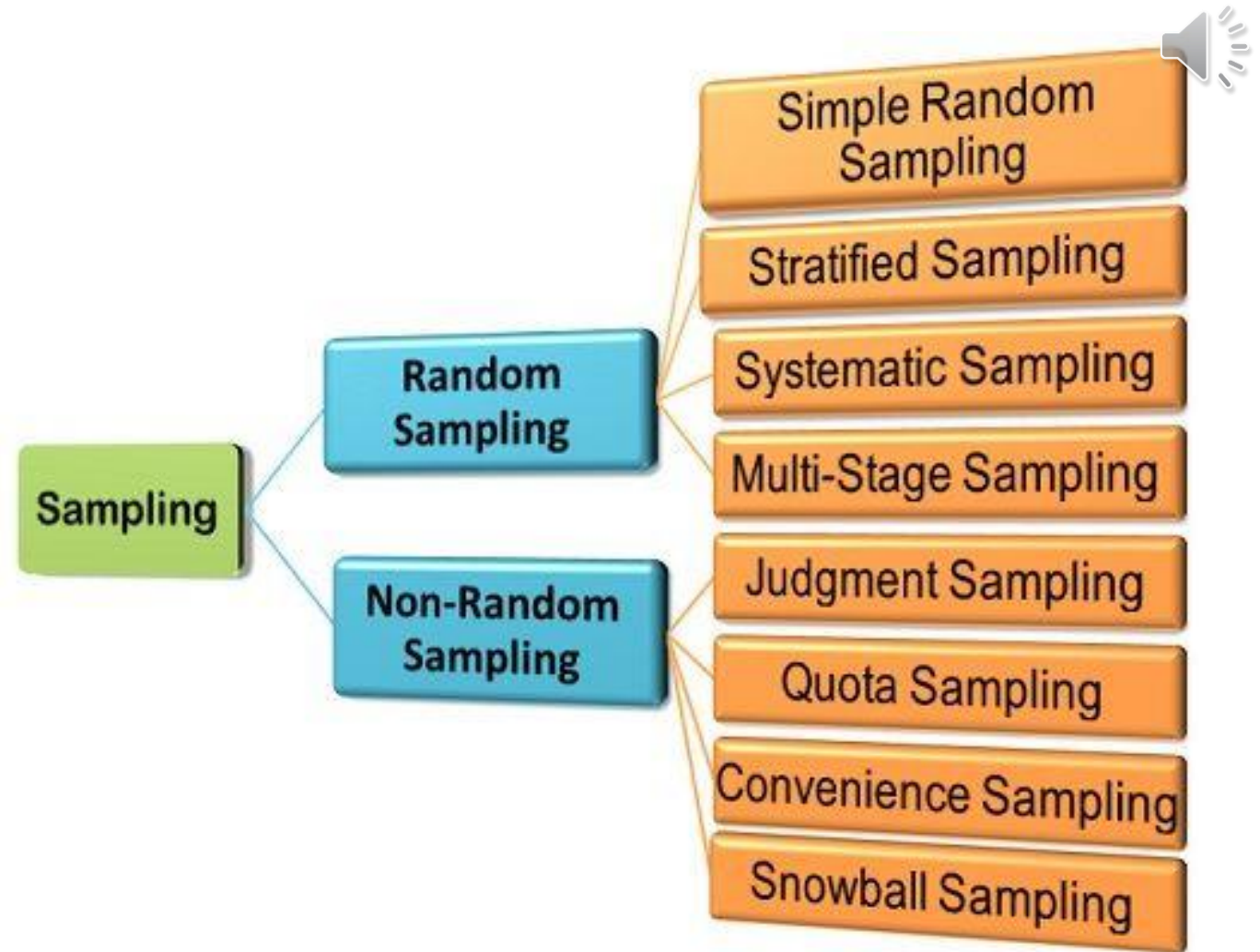


The data collection purpose will be different from the evaluation purpose, they are unlikely to provide the detail needed, such as answers to specific questions or information about the populations of interest, and the survey timing will be fixed and may not suit the evaluation's purposes.

Sampling

Sampling methods fall under two broad categories: probability sampling and nonprobability sampling.

Most surveys used in evaluations will be based on probability sampling methods, which involve selecting respondents at random from a sampling frame.



Sampling Methods (Probability)



Simple random sampling: Select people at random, usually using a computer.



Stratified sampling: Dividing a population into groups and selecting a probability sample from each group, such as by location. To assure a substantial sample from each group, selection probability can vary.



Cluster sampling: Probability selects clusters and only individuals within them. This can cut fieldwork costs, especially in face-to-face surveys if clusters are geographic locations, but it increases estimate error.



Multi-stage sampling: Involves using a combination of sampling methods

Sampling Methods (Non-Probability)



Non-probability procedures are more convenient and less expensive, but they should be used with extreme caution because the survey results cannot be used to infer anything about the population as a whole with any certainty.

- **Quota sampling** : A popular non-probability strategy for setting 'quotas' for different responders. This allows for further sample composition control, such as group representation or oversampling minority groups not in the sampling frame. Effective quota samples can replace probability sampling. The best quota sample designs use a random sample from a representative frame and group interview quotas.
- **Convenience sampling**: method of collecting samples by taking samples that are conveniently located around a location or Internet service.
- **Snowball sampling**: Existing subjects provide referrals to recruit samples required for a research study.



Survey questions



Types of questions	Type of information collected
Factual questions	Surveys often offer the only practical and affordable way of collecting such information, and in some cases, there is no other source or way of measuring the attribute of interest. This can include both objective and subjective measures.
Knowledge questions	Assess what respondents know about a particular topic and their awareness of the intervention being evaluated.
Attitudinal questions	Measures respondents' opinions, beliefs, values and feelings which cannot be verified by observation or external data sources.
Behavioral questions	Measures what people do, or intend to do, and how that has changed as a consequence of the intervention. A risk here is respondents giving socially acceptable answers.
Preference questions	Respondents provide preferences for different possible options and outcomes, including trade-offs between competing policy objectives. These can be used to elicit monetary values for different outcomes, including those not readily possessing market prices.





Four rules that are useful to consider when preparing survey questions.

1. Can the respondents understand the question? and do they understand it in the same way that you do?
2. Are respondents able to answer the question?
3. Are they willing to answer the question?
4. Will the question produce a reliable and useful response?



Co-funded by the
Erasmus+ Programme
of the European Union

Reference number: 618596-EPP-1-2020-1-SE-EPPKA2-CBHE-JP
This publication [communication] reflects the views only of the authors, and the Commission cannot be held responsible for any use, which may be made of the information contained therein.

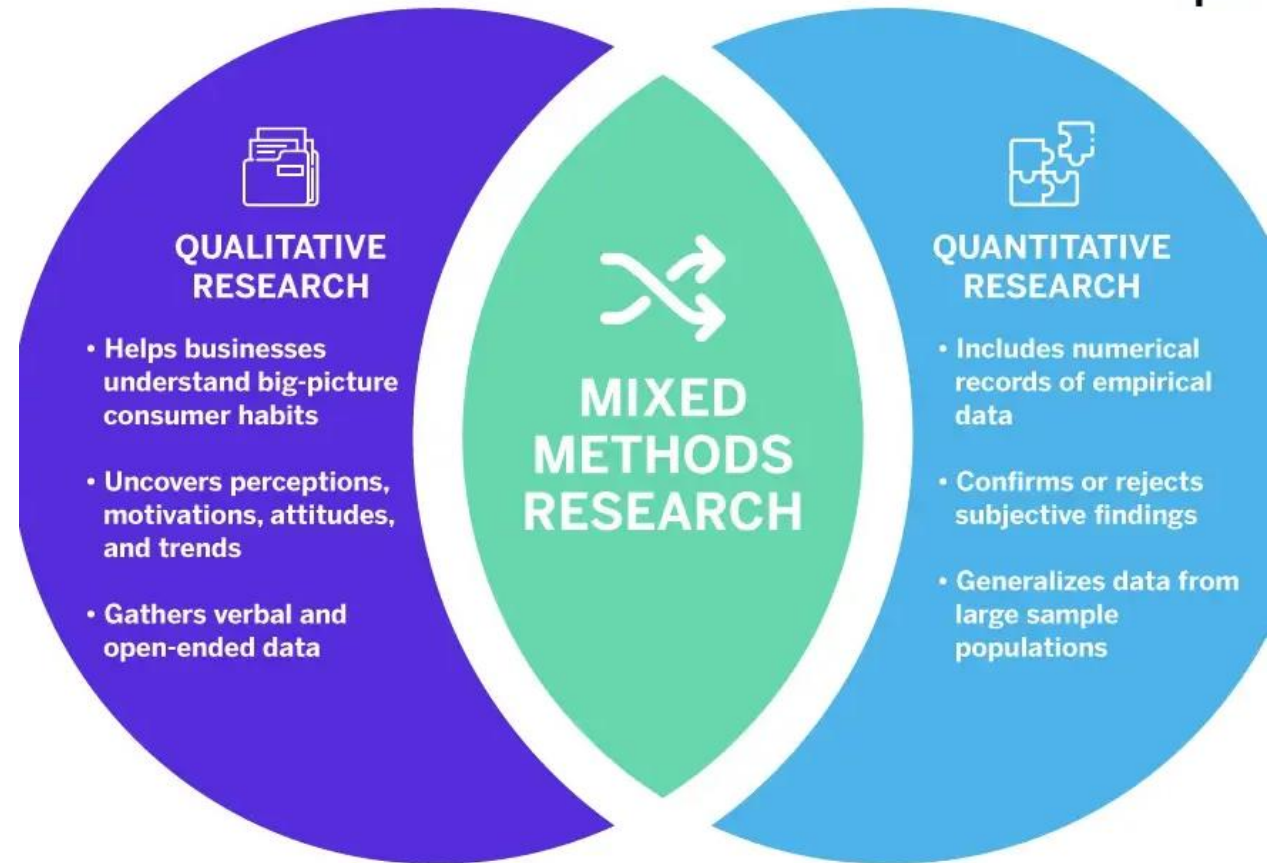


Quantitative vs Qualitative



qua

- Quantitative data is numbers-based, countable, or measurable.
- Qualitative data is interpretation-based, descriptive, and relating to language.
- Quantitative data tells us how many, how much, or how often in calculations. Qualitative data can help us to understand why, how, or what happened behind certain behaviors.



Co-funded by the
Erasmus+ Programme
of the European Union

Reference number: 618596-EPP-1-2020-1-SE-EPPKA2-CBHE-JP
This publication [communication] reflects the views only of the authors, and the Commission cannot be held responsible for any use, which may be made of the information contained therein.



Qualitative data

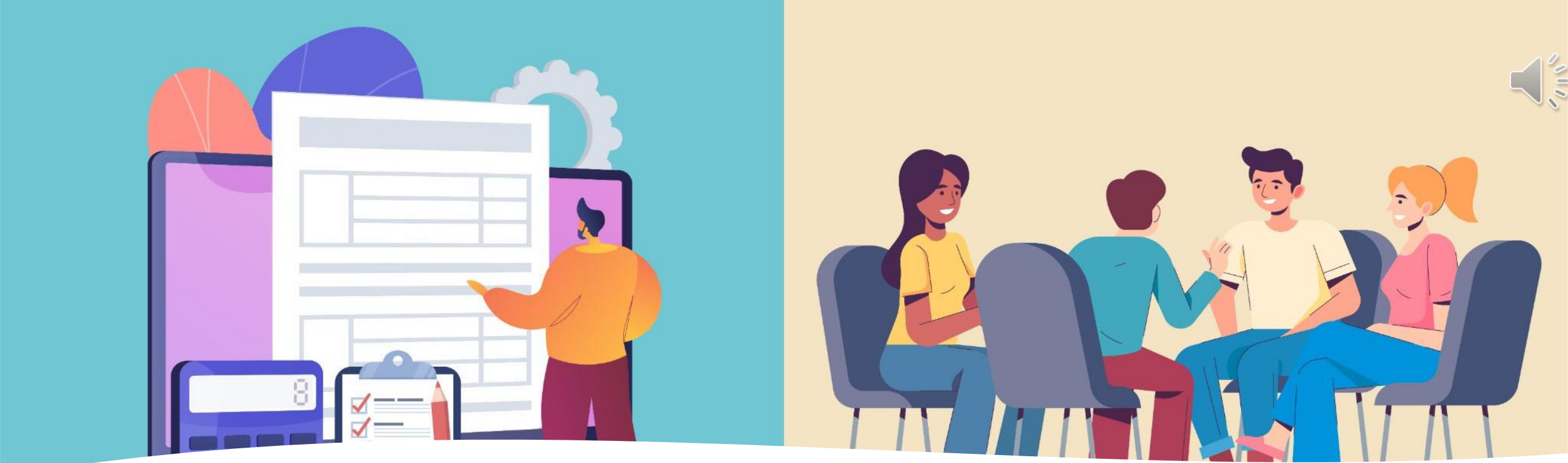


Qualitative data collection methods provide an in-depth understanding of behaviors, perceptions and underlying reasons for social phenomena. While quantitative methods are usually used to measure the 'what', qualitative methods are most often used to explore the 'how' and 'why'.

Common qualitative data collection methods include:

- **In-depth interviews:** These are used for collecting data on individuals' personal histories, perspectives, and experiences, particularly when sensitive topics are being explored, or the issue being discussed is not well understood. Questions tend to be 'open' allowing for detailed responses. In-depth interviews are typically face-to-face or by telephone.
- **Focus groups:** Focus group participants are encouraged to discuss and debate openly to understand views and experiences, allowing for a range of views to be explored. These can be used effectively for action planning and developing or improving products and services.
- **Case studies:** These are in-depth, possibly longer-term, investigations of a single issue or a small number of people, events, contexts, areas, organizations or policies.
- **Observation:** A process of watching research subjects (with their agreement and knowledge) to observe their behavior without questioning them. Can be used effectively when piloting new procedures or processes to gauge respondents' actions and behaviors.
- **Ethnography:** This includes observation, but also participation. Ethnography seeks to understand people and how they live in their cultural and physical environment. Ethnographic interviews will differ from more traditional in-depth interviews as the researcher would usually have shared time and built relationships and trust with the interviewees.





Selecting respondents for Qualitative study

- For qualitative data gathering, respondents are chosen to capture the population's rich perspectives and opinions or to illustrate unique tales or circumstances through in-depth study.
- How many respondents to choose is undefined. The goal is to understand the range of opinions on an issue and gain deeper insights into social phenomena, not to provide statistically representative findings.
- Saturation point—when additional data doesn't add anything to the findings.

Reference number: 618596-EPP-1-2020-1-SE-EPPKA2-CBHE-JP
This publication [communication] reflects the views only of the authors, and the Commission cannot be held responsible for any use, which may be made of the information contained therein.



Co-funded by the
Erasmus+ Programme
of the European Union



Data quality

- Data quality depends on the question(s) and data gathering technology.
- Research is often more reliable than monitoring systems for sensitive topics like sexual orientation and disability, where people can perceive questions differently.
- When reviewing program enrollment and termination dates, administrative or monitoring data is better.
 - Such data is more likely to be reliably recorded on monitoring systems than recalled in an interview. Data that are audited, usually as part of a payment system, are more dependable.



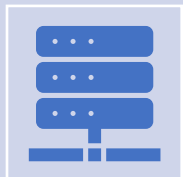
Data handling



Proper data collection, processing, and access are needed. Training everyone is necessary to manage data properly.



All assessment data must be collected, transported, stored, processed, and deleted under regulations. Data access protocols should cover user authentication, remote access, Wi-Fi access, data management, security training, and masking/encryption.



Data can be stored in a database, spreadsheet, or warehouse. Reading, writing, storing, and updating data should be restricted to legal residents with employment obligations.



Privacy in data management



Data should be anonymized immediately, removing names and addresses from the analytical file.



If temporarily kept, these should be held separately with restricted access. However, the data is usually personal and should be secured.



If possible, data should be properly anonymized at the end of a project, which may involve replacing the date of birth with month and year, or exact postcode with broader regional coding.



If this is not practicable and personal data must be kept after the project, suitable protections must be adopted.

Major data handling hazards

- Criminal data breach
- Severe harm or embarrassment
- Unlawful intrusion
- Loss of privacy
- Data loss or compromise
- Legal challenge
- Reputational damage at the department civil service, and government levels



References

HMT Magenta Book: Prudential standards in the Financial Services Bill: Policy statement, 2020. . HM Treasury, United Kingdom.





Thank you