# Course: Research Method in Software Engineering

## WEEK 6 – Data Collection Methods

**Lemlem Kassa (Ph.D.)**
**Addis Ababa Science and Technology University, Ethiopia**

# Week-6 Data Collection Methods

## Contents

# Learning Outcome

- Understand the meaning of data and the types
- Describe Importance of Data Collection Methods
- Identify different data Collection Methods
- Understand data collection process and the key steps
- Understand data sampling Strategies

# 1. Introduction to data collection

## What is date ?

- Research is a viable approach to a problem only when data can be collected to support it.
- It is a collection of facts, figures, objects, symbols, and events from different sources.
- Organizations collect data using various methods in order to make better decisions.
- **Example :** Before launching a new product, an organization must collect data on product demand, customer preferences, and competitors.


Fig. Soluciones Inába, What is Big Data and how does it work?, Soluciones Inába, https://images.app.goo.gl/B1mh9iRzoBXhoiXy7

[1] Pracfical research: Planning and design, Leedy, P. D., & Ormrod, J. E. , Global edition, 2015, Page-94

[2] Research methods: The basics, Walliman, N., Routledge , 2021 , Page-73

# 1. Introduction to data collection …Cont'd

## Primary Date Vs . Secondary Data

Data come in two main forms, depending on their closeness to the event recorded.

**i. Primary data-:** Data that have been observed, experienced or recorded close to the event are the nearest one can get to the truth. We are being shelled with primary data all day. Sounds, sights, tastes, are constantly stimulating our senses

**ii. Secondary Data :-** Written sources that interpret or record primary data **,** which tend to be less reliable. We are cascaded with secondary data in the form of news bulletins, magazines, newspapers, documentaries, advertising, the Internet, etc.

- The quality secondary data depends on the source and the methods of presentation.

[2] Research methods: The basics, Walliman, N., Routledge , 2021 , Page-77

# 1. Introduction to data collection …Cont'd

## Archival Data Vs Secondary Data

- Archival data refer to information that already exists in someone else's files.
- Archival data originally generated for reporting or research purposes, it's often kept because of legal requirements, for reference, or as an internal record.
- Archival data generally falls under the following categories: Publicly available data sets and Private data sets. Archival data is secondary data, but not all secondary data is archival data.
- Secondary data refer to research information, collected as a result of studies and similar efforts, that can then be used by others either as comparison data or as part of new research.

[2] Research methods: The basics, Walliman, N., Routledge , 2021, Page 92

# 1. Introduction to data collection …Cont'd

## Importance of data collection methods

Data collection methods play a crucial role in the research process as they determine the quality and accuracy of the data collected.

- **Maintain quality and Accuracy:** Properly designed data collection methods help ensure that the data collected is error-free and relevant to the research questions.
- **Relevance, Validity, and Reliability:** Effective data collection methods help ensure that the data collected is relevant to the research objectives, valid (measuring what it intends to measure), and reliable (consistent and reproducible).
- **Bias Reduction and Representativeness:** Carefully chosen data collection methods can help minimize biases inherent in the research process, such as sampling or response bias.

[3]. QuestionPro, Data Collection Methods: Types & Examples, Adi Bhat, https://www.questionpro.com/blog/data-collection-methods

# 1. Introduction to data collection …Cont'd

## Importance of data collection methods …Cont'd

- **Informed Decision Making:** Accurate and reliable data collected through appropriate methods provide a solid foundation for making informed decisions based on research findings.
- **Achievement of Research Objectives:** Data collection methods should align with the research objectives to ensure that the collected data effectively addresses the research questions or hypotheses.
- **Support for Validity and Reliability:** Selecting appropriate methods is critical for ensuring the credibility of the research.

[3]. QuestionPro, Data Collection Methods: Types & Examples, Adi Bhat, https://www.questionpro.com/blog/data-collection-methods

# 2. Types of Data Collection Methods

- The choice of data collection method depends on the research question being addressed, the type of data needed, and the resources and time available.
- Data collection methods can be categorized into **primary** and **secondary** methods.

## 1. Primary Data Collection Methods

- It is collected from first-hand experience and is not used in the past.
- The data gathered by primary data collection methods are highly accurate and specific to the research's motive.

- Primary data collection methods can be divided into two categories:

  **a) Quantitative**

  **b) Qualitative**

  **c) Mixed (Qualitative and Quantitative )**

# 2. Types of Data Collection Methods …Cont'd

## a) Quantitative Methods

- This method focuses on collecting numeric data that can be statistically analyzed.
- It is ideal for measuring the use and satisfaction with products or services and identifying behavioral patterns.
- Common methods of quantitative primary data collection include:
  - **Online Surveys**: Surveys conducted over the internet that have a wide reach. Participants can join from anywhere with an internet connection.
  - **Observational Studies**: Measuring human behavior in a natural environment to collect data on their actions and interactions.

[3]. QuestionPro, Data Collection Methods: Types & Examples, Adi Bhat, https://www.questionpro.com/blog/data-collection-methods

# 2. Types of Data Collection Methods …Cont'd

**b) Qualitative Methods**

- Qualitative primary data collection gathers information in an unstructured manner.
- This method is helpful for gaining an in-depth understanding of a topic or phenomenon.
- It is often used to understand the opinions and feelings of the target group and to explore a topic from different perspectives.


Fig. AIMultiple, 11 Online Survey Challenges, Cem Dilmegani, https://images.app.goo.gl/TKG4WLnLDu8AXqGd9,2024

[3].QuestionPro, Data Collection Methods: Types & Examples, Adi Bhat, https://www.questionpro.com/blog/data-collection-methods

# 2. Types of Data Collection Methods …Cont'd

## Qualitative Data Collection Methods

### 1. Surveys

- Used in interviews and focus groups to gather information from customers.
- Typically distributed in the form of questionnaires with a combination of close-ended, demographic questions and open-ended research questions on a particular topic.
- The major advantage of this method is that it's **less time-consuming** than others. Also allow to gather information from a large population of customers quickly and effectively.

**2. Interviews:** In face-to-face interviews, the interviewer asks a series of questions to the interviewee in person and notes down responses.

# 2. Types of Data Collection Methods …Cont'd

## 3. Focus Groups

- In a focus group, a small group of people, around 8-10 members, discuss the common areas of the research problem.
- Each individual provides his or her insights on the issue concerned.
- A moderator regulates the discussion among the group members.
- At the end of the discussion, the group reaches a consensus.

# 2. Types of Data Collection Methods …Cont'd

## 4. Questionnaire

- It is a printed set of open-ended or closed-ended questions that respondents must answer based on their knowledge and experience with the issue.
  - **Closed-ended questions** provide respondents with a limited set of predefined answers from which to choose such as a simple "yes" or "no," or by selecting from multiple choice options.
  - **Open-ended questions** allow respondents to answer in their own words, providing more freedom and flexibility in their responses- encourage detailed and feedback.

# 2. Types of Data Collection Methods …Cont'd

## C) Mixed Primary Data Collection

- Combines elements of qualitative and quantitative methods. This hybrid approach leverages the strengths of both methods to achieve comprehensive and meaningful results.
- **Qualitative data** -This information helps to understand the "**why"** behind certain behaviors.
- **Quantitative data** It answers the "**how much"** and "**how often"** and is often collected through surveys and observational studies.

[3]. QuestionPro, Data Collection Methods: Types & Examples, Adi Bhat, https://www.questionpro.com/blog/data-collection-methods

# 2. Types of Data Collection Methods …Cont'd

## 2. Secondary Data Collection Methods

- Data that has been used in the past. The researcher can obtain data from the data sources, both internal and external, to the organizational data. Involve quantitative and qualitative techniques

**Internal sources of secondary data:**

- Organization's health and safety records
- Mission and vision statements
- Financial Statements
- Magazines
- Sales Report
- Executive summaries

**External sources of secondary data:**

- Government reports
- Press releases
- Business journals
- Libraries
- Internet

- Secondary data is easily available, less time-consuming, and expensive than primary data.

# 3. Data Collection Process

- The data collection process typically involves several key steps to ensure the accuracy and reliability of the data gathered.

## Data collection process

- **Define the Objectives:** Clearly outline the goals of the data collection. What questions are you trying to answer?
- **Identify Data Sources:** Determine where the data will come from. This could include surveys, interviews, existing databases, or observational data.
- **Choose Data Collection Methods:** Select appropriate methods based on our objectives and data sources such as *questionnaires, Interviews, etc.*

# 3. Data Collection Process …Cont'd

## Data collection process …cont'd

- **Develop Data Collection Instruments:** Create or adapt tools for collecting data, such as questionnaires or interview guides. Ensure they are valid and reliable.
- **Select a Sample**: If you are not collecting data from the entire population, determine how to select sample. Consider sampling methods like random, stratified, or convenience sampling.
- **Collect Data**: Execute data collection plan, following ethical guidelines and maintaining data integrity.
- **Store Data**: Organize and store collected data securely, ensuring it's easily accessible for analysis while maintaining confidentiality.

# 3. Data Collection Process …Cont'd

## Data collection process ….Cont'd

- **Analyze Data:** process and analyze it according to our objectives, using appropriate statistical or qualitative methods.
- **Interpret Results:** Conclude our analysis, relating them back to original objectives and research questions.
- **Report Findings:** Present findings clearly and organized, using visuals and summaries to communicate insights effectively.
- **Evaluate the Process:** Reflect on the data collection process. Assess what worked well and what could be improved for future studies.

# 4. Data Sampling Strategies

- When we conduct research about a group of people, it's rarely possible to collect data from every person in that group. Instead, we select a **sample**- (the group of individuals who will actually participate in the research).
- To draw valid conclusions from results, we have to carefully decide how to select a sample that is representative of the group as a whole. This is called a **sampling method**.

**Sample size**

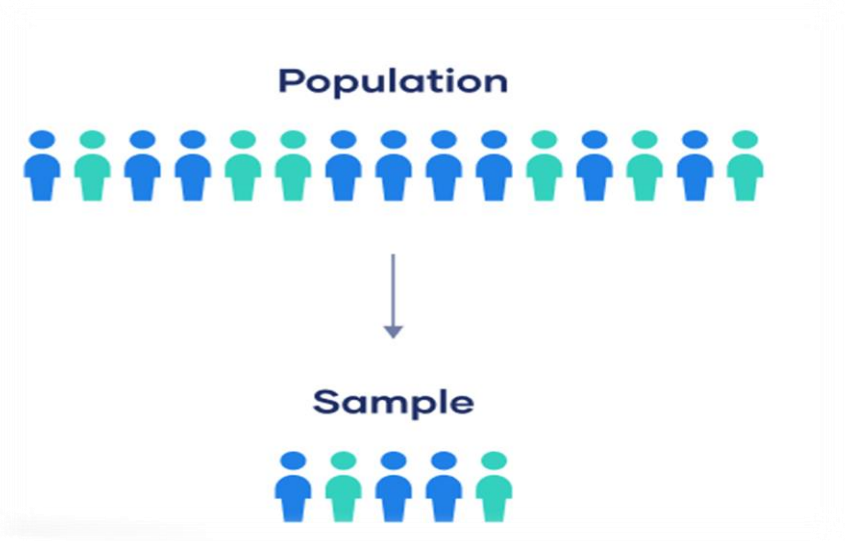
- The number of individuals we should include in our sample depends on various factors, including the size and variability of the population and research design.
- There are different sample size calculators and formulas depending on what we want to achieve with statistical analysis.

[2] Research methods: The basics, Walliman, N., Routledge , 2021, Page-106

# 4. Data Sampling Strategies …Cont'd

## Population Vs. Sample

- First, understand the difference between a population and a sample, and identify the target population of the research.
- The population is the entire group that we want to draw conclusions about.
- The sample is the specific group of individuals that we will collect data from.
- The population can be defined in terms of geographical location, age, income, or many other characteristics.




**Population vs. sample**
**https://www.scribbr.com/methodology/sampling-methods**

[4]. Scribbr, Sampling Methods | Types, Techniques & Examples, McCombes, Oct. 2024, https://www.scribbr.com/methodology/sampling-methods

# 4. Data Sampling Strategies …Cont'd

- A lack of a representative sample affects the validity of research results, and can lead to several research biases, particularly sampling bias.

**Sampling frame**

- The sampling frame is the actual list of individuals that the sample will be drawn from. Ideally, it should include the entire target population
  - **Example**: Assume we are doing research on working conditions at a social media marketing company.
    - **Population:** all 1000 employees of the company.
    - **Sampling frame**: the company's HR database, which lists the contact details of every employee.

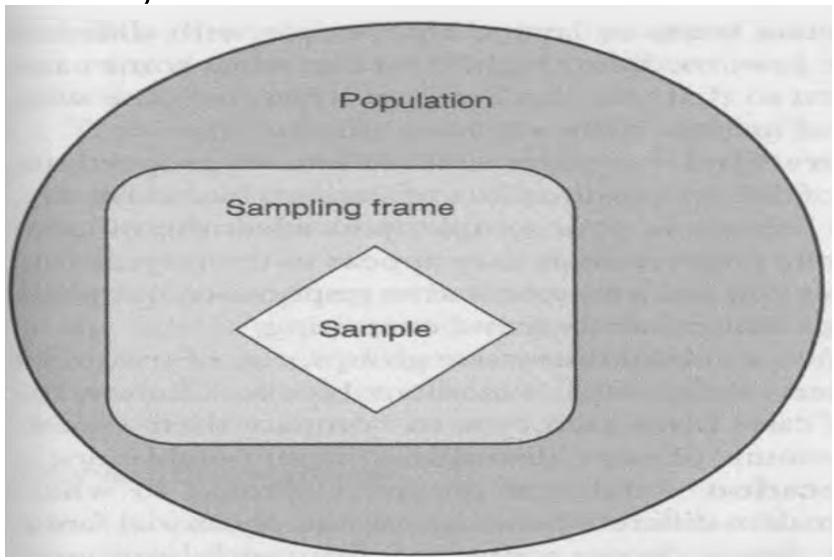



Fig. Sampling frame in relation to population and sample

[2]. Research methods: The basics, Walliman, N., Routledge , 2021, Page-107

# 4. Data Sampling Strategies …Cont'd

- There are basically two types of **sampling procedure**: **Probability sampling and Non-probability sampling**

**A) Probability sampling methods**

- Probability sampling means that every member of the population has a chance of being selected.
- It is mainly used in quantitative research.
- If you want to produce results that are representative of the whole population, probability sampling techniques are the most valid choice.

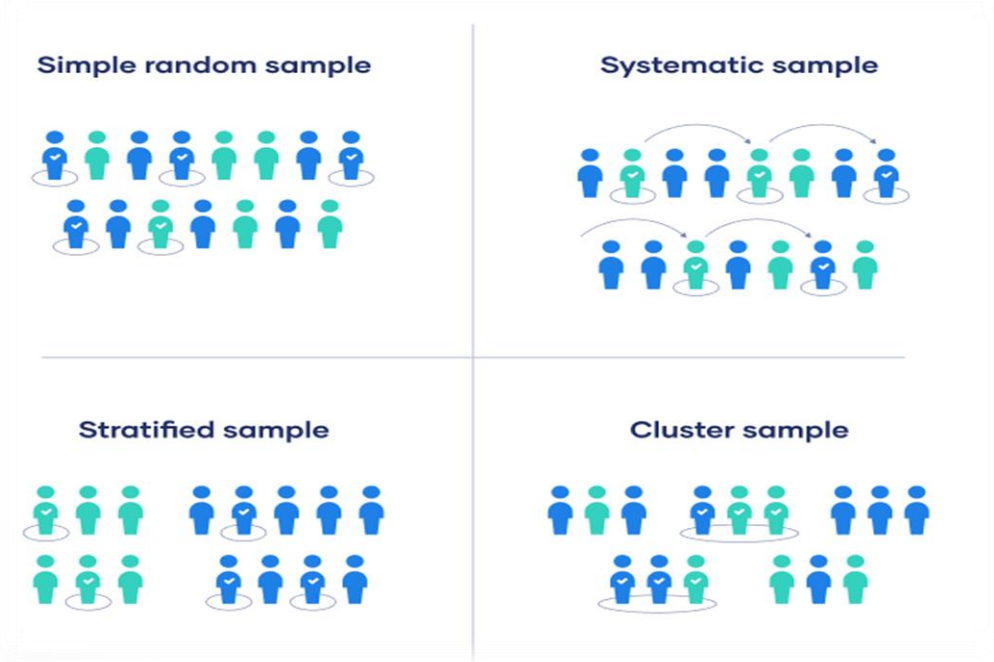
[2]. Research methods: The basics, Walliman, N., Routledge, 2021,Page -109

# 4. Data Sampling Strategies …Cont'd

There are **four main types of probability sample**

## 1. Simple random sampling

- Every member of the population has an equal chance of being selected.
- To conduct this type of sampling, we can use tools like **random number generators** or other techniques that are based entirely on chance.




[4]. Scribbr, Sampling Methods | Types, Techniques & Examples, McCombes, Oct. 2024, https://www.scribbr.com/methodology/sampling-methods

# 4. Data Sampling Strategies …Cont'd

## 2. Stratified sampling

To use this sampling method, we divide the population into subgroups (called strata) based on the relevant characteristic (e.g., gender identity, age range, income bracket, job role).

- Based on the overall proportions of the population, we calculate how many people should be sampled from each subgroup. Then use random or systematic sampling to select a sample from each subgroup.
- **Example:** The company has 800 female employees and 200 male employees.
- To ensure that the sample reflects the gender balance of the company, We sort the population into two strata based on gender.
    - Then use random sampling on each group, selecting 80 women and 20 men, which gives a representative sample of 100 people.

# 4. Data Sampling Strategies …Cont'd

## 3. Systematic sampling

- Similar to simple random sampling, but it is usually slightly easier to conduct, but instead of randomly generating numbers, individuals are chosen at regular intervals.
- We should Note that if we use this technique, it is important to make sure that there is no hidden pattern in the list that might skew the sample.

**Example:** All employees of the company are listed in alphabetical order.

- From the first 10 numbers, randomly select a starting point: number 6.
- From number 6 onwards, every 10th person on the list is selected (6, 16, 26, 36, and so on), and end up with a sample of 100 people.

[4]. Scribbr, Sampling Methods | Types, Techniques & Examples, McCombes, Oct. 2024, https://www.scribbr.com/methodology/sampling-methods

# 4. Data Sampling Strategies …Cont'd

## 4. Cluster sampling

- Involves dividing the population into subgroups, but each subgroup should have similar characteristics to the whole sample. Instead of sampling individuals from each subgroup, we randomly select entire subgroups.
- This method is good for dealing with large and dispersed populations, but there is more risk of error in the sample, as there could be substantial differences between clusters. It's difficult to guarantee that the sampled clusters are really representative of the whole population.
- **Example**: The company has offices in 10 cities across the country (all with roughly the same number of employees in similar roles).
- awe don't have the capacity to travel to every office to collect data, so we use random sampling to select 3 offices – these are your clusters

# 4. Data Sampling Strategies …Cont'd

**B) Non-probability sampling methods**

- Individuals are selected based on non-random criteria, and not every individual has a chance of being included.
- This type of sample is easier and cheaper to access, but it has a higher risk of sampling bias.
- If we use a non-probability sample, we should still aim to make it as representative of the population as possible.
- This sampling techniques are often used in exploratory research in which the aim is not to test a hypothesis about a broad population, but to develop an initial understanding of a small population.

# 4. Data Sampling Strategies …Cont'd

## 1. Convenience sampling

- Simply includes the individuals who happen to be most accessible to the researcher.
- This is an easy and inexpensive way to gather initial data, but there is no way to tell if the sample is representative of the population, so it can't produce generalizable results.
- Convenience samples are at risk for both sampling bias and selection bias.
  - **Example:** Gathering feedback on software usability from users who are easily accessible, such as team members or students in a course.
  - This is a convenient way to gather data, but as only surveyed students in a course, the sample is not representative of all the students.

[4]. Scribbr, Sampling Methods | Types, Techniques & Examples, McCombes, Oct. 2024, https://www.scribbr.com/methodology/sampling-methods

# 4. Data Sampling Strategies …Cont'd

## 2. Purposive sampling

- Also known as judgement sampling, involves the researcher using their expertise to select a sample that is most useful to the purposes of the research.
- It is often used in qualitative research, where the researcher wants to gain detailed knowledge about a specific phenomenon rather than make statistical inferences, or where the population is very small and specific.
- We should always make sure to describe our inclusion and exclusion criteria and beware of observer bias affecting our arguments.
- **Example:** For gathering opinions on software usability, researchers may choose participants who represent specific demographic or professional backgrounds.

# 4. Data Sampling Strategies …Cont'd

## 3. Snowball sampling

- If the population is hard to access, snowball sampling can be used to recruit participants via other participants. The number of people we have access to "snowballs" as we get in contact with more people.
- The downside here is also representativeness, as we have no way of knowing how representative our sample is due to the dependence on participants recruiting others. This can lead to sampling bias.

**Example: Snowball sampling**

- It is a valuable technique in software engineering research, particularly when studying specialized or hard-to-reach populations. By leveraging existing networks, researchers can gather rich qualitative data and insights that might difficult to obtain.

# 4. Data Sampling Strategies …Cont'd

## 4. Quota sampling

- Relies on the non-random selection of a predetermined number or proportion of units. This is called a quota.
- First divide the population into mutually exclusive subgroups (called strata) and then recruit sample units until we reach our quota. These units share specific characteristics, determined by prior to forming our strata. The aim of quota sampling is to control what or who makes up our sample.

**Example:** Quota sampling method to study the impact of remote work on software development productivity.

- Determine the characteristics that are important for the study (e.g., role in the organization, years of experience, type of software developed).

- Establish quotas for each subgroup:- 30% junior developers ,40% senior developers, 30% project managers

# Summary

- Research is a viable approach to a problem only when data can be collected to support it.
- Data come in two main forms, depending on their closeness to the event recorded. Such as Primary data and secondary data.
- Data collection methods play a crucial role in the research process as they determine the quality and accuracy of the data collected.
- The choice of data collection method depends on the research question being addressed, the type of data needed, and the resources and time available. Data collection methods can be categorized into primary and secondary methods.
- The data collection process typically involves several key steps to ensure the accuracy and reliability of the data gathered.
- Effective data sampling strategies improve the validity of study results and reduce costs and time.

# References

1. Pracfical research: Planning and design, Leedy, P. D., & Ormrod, J. E. , Global edition,  2015, Page-94

2. Research methods: The basics, Walliman, N., Routledge , 2021 , Page-73

3. QuestionPro, Data Collection Methods: Types & Examples, Adi Bhat, https://www.questionpro.com/blog/data-collection-methods

4. Scribbr, Sampling Methods | Types, Techniques & Examples, McCombes, Oct. 2024, https://www.scribbr.com/methodology/sampling-methods

.

# Thank You!