

TYPES AND COLLECTION OF DATA

Data collection is the process of gathering and measuring information on targeted variables in an established systematic fashion, which then enables one to answer relevant questions and evaluate outcomes. The data collection component of research is common to all fields of study including physical and social sciences, humanities and business. While methods vary by discipline, the emphasis on ensuring accurate and honest collection remains the same. The goal for all data collection is to capture quality evidence that then translates to rich data analysis and allows the building of a convincing and credible answer to questions that have been posed.

Quantitative data can be collected using four main types of surveys:

Census	data collection about every unit in a group or population, e.g. if a VET provider collects data about the age of trainees in a class, that would be considered a “class census”
Sample Survey:	only part of population is approached for data, e.g. if a VET provider collects data about the age of female trainees in a programme, that would be considered a “sample survey of the programme”
Administrative data:	collected as a result of an organization’s day-to-day operations, e.g. data on enrolment
Tracer studies:	usage of a) a regular survey as the core tool b) combining it with in-depth discussions with a sample of those surveyed and c) interviewing key informants on particular key issues

Qualitative data can be gathered in a variety of ways, for example:

Questionnaires:	Series of questions and other prompts for the purpose of gathering information from respondents, i.e. VET stakeholders
Interviews:	Conversation between two or more people where questions are asked by the interviewer to obtain information from the interviewee(s)
Focus groups:	A group of people are asked about their attitude towards a product, service, concept, an idea
Observation:	A group or single learners are asked to perform a specific task or action. Observations are then made of behaviors, processes, workflows etc, in a real-life situation (e.g. the workplace)

Case study:

Often used to provide context to other data (such as outcome data), offering a more complete picture of what happened and why

Census

A census refers to data collection about every unit in a group or population. If you collected data about the height of everyone in your class, that would be regarded as a class census. There are various reasons why a census may or may not be chosen as the method of data collection:

Advantages (+)

Sampling variance is zero: There is no sampling variability attributed to the statistic because it is calculated using data from the entire population.

Detail: Detailed information about small sub-groups of the population can be made available.

Disadvantages (-)

Cost: In terms of money, conducting a census for a large population can be very expensive.

Time: A census generally takes longer to conduct than a sample survey.

Response burden: Information needs to be received from every member of the target population.

Control: A census of a large population is such a huge undertaking that it makes it difficult to keep every single operation under the same level of scrutiny and control.

Sample survey

In a sample survey, only part of the total population is approached for data. If you collected data about the height of 10 students in a class of 30, that would be a sample survey of the class rather than a census. Reasons one may or may not choose to use a sample survey include:

Advantages (+)

Cost: A sample survey costs less than a census because data are collected from only part of a group.

Time: Results are obtained far more quickly for a sample survey, than for a census. Fewer units are contacted and less data needs to be processed.

Response burden: Fewer people have to respond in the sample.

Control: The smaller scale of this operation allows for better monitoring and quality control.

Disadvantages (-)

Sampling variance is non-zero: The data may not be as precise because the data came from a sample of a population, instead of the total population.

Detail: The sample may not be large enough to produce information about small population sub-groups or small geographical areas.

Administrative data

Administrative data are collected as a result of an organization's day-to-day operations. Examples include data on births, deaths, marriages, divorces and car registrations. For example, prior to being issued a marriage license, a couple must provide the registrar with information about their age, sex, birthplace, address and previous marital status. These administrative files can be used later as a substitute for a sample survey or a census.

Advantages (+)

Sampling variance is zero: There is no variability attributed to the statistic because it was calculated using data from the entire population.

Time series: Data are collected on an ongoing basis, allowing for trend analysis.

Simplicity: Administrative data may eliminate the need to design a census or survey and the associated work.

Response burden: Since the data are already collected, there is no additional burden on the respondents.

Disadvantages (–)

Flexibility: Data items may be limited to essential administrative information, unlike a survey.

Population: Data are limited to the population on whom the administrative records are kept.

Change over time: Definitions are created to serve specific purposes, but often change and evolve over time. The statistician must understand that there is a possibility of change to the definitions of these files.

Concepts and definitions: The definitions are established by those who create and manage the file for their own purposes. For example, income definitions may not include everything a user expects to see.

Data quality: The emphasis placed on data quality may differ from organization to organization. This may be evident when someone relies on data collected from another organization.