**7. Using Secondary Data**

**Secondary Data**

They are data that already exist and do not have to be collected by the researcher. Examples of secondary data: statistical bulletins, government publications, information published or unpublished and available from either within or outside the organisation, data available from previous research, case studies and library records, on-line data, web-sites and the internet.

Advantages of Secondary Data

1. They are available.
2. Save time and money.
3. Essential in instances when data simply cannot be obtained using primary data collection procedures.

Disadvantages of Secondary Data

1. May not match your need.
2. Access can be costly or difficult.
3. Aggregations and definitions may be unsuitable.
4. No real control over data quality.
5. Initial purpose may affect how data are presented.

Evaluating secondary data sources

You need to be sure:

1. The secondary data will enable you to answer your research question(s) and to meet your objectives.
2. The benefits associated with their use will be greater than the costs.
3. You will be allowed access to the data.

**Overall suitability**

Pay attention to measurement validity and coverage including unmeasured variables.

Measurement validity - the secondary data provide you with the information that you need to answer your research question(s) or meet your objectives. Sometimes the measures used may not match those that you need then the secondary data are invalid.

Coverage and unmeasured variable - ensure that the secondary data cover the population about which you need data, for the time period you need, and contain data variables that will enable you to answer your research question(s) and to meet your objectives.

Coverage will concern two issues:

1. Ensuring that unwanted data are or can be excluded.
2. Ensuring that sufficient data remain for analysis to be undertaken once unwanted data have been excluded.

**Precise suitability**

This refers to the secondary data being reliable and valid and no biasness.

Reliability - is assessing the reputation/authority of the source.

Validity - refers to the consistency and accuracy of the data, the method by which the data are collected and the precision needed by the original collector of the data. This could relate to the sampling technique used, the sampling error associated with it and the response rates.

The reliability and validity of secondary data are related to the method by which the data are collected and the source.

The source from which the secondary data re obtained can be assessed by its authority or reputation. Well known organisations are noted for their reliability and trustworthiness because their existence is dependent on their credibility.

The reliability and validity of secondary data are dependent on a detailed assessment of the method(s) used to collect the data. The following aspects should be examined:

1. Who were responsible for collecting or recording the information and the context in which the data were collected?
2. What are the possible error/biases?
3. The process by which the data were selected and collected.
4. The sampling design used to extract the data and the associated sampling error and response rates. E.g. survey gives high response rate and is more reliable than low response rate. For dairies, transcripts of interviews or meetings are unlikely to have formal methodology of how the data were collected and therefore reliability would be difficult to assess. Letters and memos are personal point of view and should only be considered as writer’s perceptions and view but not an objective account of reality. In other words, reliability and validity of secondary data will be easier to assess, if the methodology used to collect the data is explained, the sampling technique used and the response rates are explained.
5. Where data are compiled as in a report, pay particular attention to how these data are analysed and how the results are reported. If percentages are given without the totals being given, examine the data very carefully.
6. Similarly where quotations of figures or data are given without supporting evidence of references or authorities again you should be very careful. These data may not be reliable.

**Measurement bias**

This can occur for two reasons:

1. Deliberate or intentional distortion of data.

2. Changes in the way data are collected.

In organisations, distortion of records occurred deliberately by managers and employees. However, measurement bias resulting from deliberate distortion is difficult to detect. This can be detected by triangulate the findings i.e. cross-checking verification, with other independent data sources. Where two or more conclusions are similar you can be confident that the data are not distorted.

Changes in which the data are collected can also cause distortion of data. There is no consistency in the collection of the data. Once the method is changed to collect the data, the bias also changes. You have to discover the way data are recorded has changed.

**Costs and Benefits**

According to Kervin (1999) the final criterion for assessing secondary data is a comparison of the costs of acquiring them with the benefits they will bring.

Costs include time and financial resources that are needed to obtain the data. Some data can be obtained free while others may need to be paid to get them. Data from consultant organisations may be costly to obtain.

Benefits may be assessed in terms of the extent to which data will enable you to answer your research question(s) and meet your objectives. The benefits can be assessed from the overall and precise suitability. This assessment is summarised in the checklist below.

|  |
| --- |
| Checklist to evaluate secondary data sources |
| Overall suitability   * Does the data set contain the information you require to answer your research question(s) and meet your objectives? * Do the measures used match those you require? * Is the data set a proxy for the data you really need? * Does the data set cover the population that is the subject of your research? * Can data about the population that is the subject of your research be separated from unwanted data? * Are the data sufficiently up to date? * Are data available for all the variables you require to answer your research question(s) and meet your objectives? |
| Precise suitability   * How reliable is the data set you are thinking of using? * How credible is the data source? * Is it clear what the source of the data is? * Is the source of the data likely to be reliable? * Do the data have an associated copyright statement? * Do associated published documents exist? * Is the methodology clearly described? * If sampling was used what was the procedure and what were the associated sampling errors and response rates? * Who were responsible for collecting or recording the data? * (For surveys) Is a copy of the questionnaire or interview checklist included? * (For compiled data) Are you clear how the data were analysed and compiled? * Are the data likely to contain measurement bias? * What was the original purpose for which the data were collected? * Who was the target audience and what was their relationship to the data collector or compiler (were there any vested interests)? * Have there been any documented changes in the way the data are measured or recorded including definition changes? * How consistent are the data obtained from this source when compared with data from other source? * Are you happy that the data have been recorded accurately? |
| Costs and benefits   * What are the financial and time costs of obtaining these data? * Have the data already been entered into a computer? * Do the overall benefits of using these secondary data sources outweigh the associated costs? |

Questions

1. Give three examples of different situations where you might use secondary data as part of your research.