2

DATA QUALITY ASSESSMENT

Emelia Akashah Patah Akhir

2.1 DATA QUALITY MANAGEMENT

Data quality management is a set of practices to maintain the data quality. It is an essential process to be carried out from data acquisition to data distribution process. It involved several steps in making sense of data which includes measurement and assessment of data before data pre-processing can be done. These steps in data quality management put quality in context to improve the data to be used for decision-making and analysis.

2.2 DEFINITION OF DATA QUALITY

The Data Management Body of Knowledge [4] defines data quality as,

> *"The planning, implementation, and control of activities that apply quality management techniques to data, in order to assure it is fit for consumption and meet the needs of data consumers."[4]*

Data is always one of the valuable assets in any organisation. Thus, ensuring their quality is very important especially in assisting them to make decision for the organisation.

2.3 IMPORTANCE OF DATA QUALITY

Data quality assessment is very crucial to deliver business value. Sound and informed decisions cannot be made on poor quality data or the data for which the quality is unknown. Analysis results from poor quality data can lead to a wrong prediction or decision making.

Thus, ensuring and maintaining the defined data quality according to identified parameters or dimensions is necessary to minimise risks while facilitating decisions.

In the context of any project or decision-making process in an organisation, the data quality status for the dataset that will be used for the said activities need to be measured prior to the execution of the activities to eliminate any unfit data to be used for analysis and decision-making. These datasets will be measured according to relevant parameters identified during the data understanding and assessment phase. The below section will explain further on data quality parameters or dimensions.

2.4 DATA QUALITY DIMENSIONS

Data quality dimensions are a common term that is widely used to describe the data characteristics to be used in measuring the data quality. In practice, these datasets will be measured against defined standard or rules to maintain its quality.

Data quality dimensions from the Data Management Association (DAMA) will be adopted in measuring data quality for this dataset. DAMA introduced six (6) primary dimensions to be used for data quality assessment as in Figure 2.1.



Figure 2.1 DAMA data quality dimensions