

SCHOOL OF ECONOMICS
DEPARTMENT OF DATA SCIENCE AND ECONOMIC POLICY
UNIVERSITY OF CAPE COAST

MSC. (DATA MANAGEMENT AND ANALYSIS) SANDWICH PROGRAMME
DMA 820: DATA CURATION AND MANAGEMENT

GABRIEL KWASI ABOAGYE
SE/DMD/23/0003

LECTURER: DR. RAYMOND ELIKPLIM KOFINTI

QUESTION 1: OUTLINE SIX (6) DATA QUALITY PERFORMANCE INDICATORS RECOMMENDABLE FOR MEETING THE VISION OF YOUR INSTITUTION

Quality Performance Indicators are specific and measurable elements used to determine the quality of data. These indicators include:

1. **Completeness:** A set of data should be wholesome or comprehensive to allow users to be able to fully utilise the data for their intended purposes. A complete data are not “half-baked” data. Rather, they are data that contains all the essential elements or variables upon which the user (researcher) can perform meaningful analyses and draw sound conclusions. For example, a researcher collecting data on Maternal Health should ensure that a comprehensive data is obtained from all aspect of Maternal Health as a way of ensuring the comprehensiveness of the data.
2. **Unique:** One dataset should be distinct from the other. Users of such data should be able to draw distinctions among different datasets even if they are related. In order to ensure the uniqueness of data, labels or specific identifiers (words, letters, symbols) should be assigned to each of the data sets or data points to enable for a clearer differentiation from one dataset to another. For example, a dataset on Maternal Health will be labelled as such while a dataset on Child Mortality will also be labelled as such. Therefore, users of such data will be able to differentiate between these two different datasets based on their uniqueness stemming from the differences in their labelling.
3. **Freshness:** These are up-to-date data and they are available on time or as and when needed. Thus, adequate timely disclosure. Conversely, data is not fresh when it is stale and unreliable. Freshness can be associated with various levels including metadata. The freshness of data is also ascertained by the length of time between data availability and time of the event or phenomenon they describe. For example, in Time series data, if an additional data has been obtained in 2024, the databank or database needs to be updated to keep it current worthy for analysis.
4. **Valid:** A valid data is a sound data that enables researchers and other end users to achieve their intended purposes for which they want to use the data for. Thus, a valid data is the extent to which the resulting data actually represents what it intends to measure. In order to ensure that, such data should be collected according to pre-defined rules and parameters, and should conform to the right format and be within the right range. For instance, the research process, and for that matter, data collection is a scientific process. This implies that certain protocols must be followed, such as sampling technique and from whom the data were obtained, if these protocols or procedures are flouted, then it renders the data invalid, hence affecting the quality of the data.
5. **Accuracy:** As one of the most complex, yet critical data quality indicators, accuracy is the degree to which the data item correctly describes the “real world” object or event being described. In order to collect accurate data, the data collector must put in several interventions in order to ensure that the data obtained is accurate. An accurate data is the data that really reflect the realities of the real-life situations. This will enable users of such data to analyse and come out with policy driven conclusions and recommendations.

6. **Consistency:** Consistency is key when it comes to quality data assurance. Consistency, therefore refers to ensuring that data values are consistently represented within a dataset and between datasets, and consistently associated across datasets. Or the degree to which data values of two sets of attributes within a record comply with a rule. Or the degree to which the data values of a set of attributes of a dataset at different points in time comply with a rule. A consistent data will assist users to develop patterns and relationships within the dataset. A consistent data is a reliable data.

QUESTION 2: IDENTIFY AND JUSTIFY FIVE (5) ADMINISTRATIVE DATA POINTS YOUR ORGANISATION SHOULD PRIORITISE FOR ARCHIVING FOR THE NEXT FIVE YEARS

Meaning of Data Points

A data point is a piece of information that describes one unit of observation, at one point in time, at the data collection level. They are identifiable elements of a data set. They could also be described as units of observations. In a statistical or analytical context, it is the factual information derived from a measurement or research and can be represented as a numerical data point, a statistical display or a graph. In my organisation, some crucial administrative data points I will recommend for archiving in the next five years are:

1. **Staff Biographic Information:** This data point solicits information from a person about their biological and demographic characteristics including name, date of birth, sex, educational qualification and others. The justification for the essence of such information is that it will provide the organisation with relevant biological characteristics that will be significant for improved job performance. For instance, certain personality traits such as either being an introvert or extrovert is directly related to certain job specification and job description. This will facilitate effective job placement knowing certain personal information about certain staff.
2. **Staff Employment History:** This is the chronological arrangement of a person's working life over a period of time. It is important to archive such data on an employee since it gives the organisation the opportunity to gain access to the working experiences of an individual in relation to the person's current job portfolio. This will enable my organisation identify the training needs, challenges or areas of higher performance to enhance staff productivity with its resultant effect on the overall organisational performance.
3. **Staff Continuous Professional Development Activities:** These include training workshops, conferences, seminars as well as any other relevant activity undertaken by the staff which has the tendency of enhancing their career progression, development and advancement. This will give the organisation more insights on how exposed the staff is with regard to what has been learnt or experienced off-the-job and how the staff will be able to ensure effective transfer of learning to the job to augment their performance on the job. This will also give my organisation the opportunity to roll-out effective training packages for its staff.

4. **Number of Publications by Staff:** These include publications in refereed journals, books, articles, magazines, modules, dissertations, theses and any other forms of research activities having been undertaken. This gives the organisation information on how research-oriented the staff is and this also describes your capabilities and research capacity. This is important because research has been identified as one of the key drivers or performance indicators of an organisation. For instance, the organisation will be informed the kind of task, especially, when it is research-related to be assigned to a particular staff based on their research capacity or what kind of training workshop should attended by a particular staff based on their research experiences? Books written will also inform the area of specialisation or interest of the staff who is an author.
5. **Data on Staff Engagements in Community Services, Projects, Committees and Boards:** Even though, these activities are not directly related to the Job Description of the staff, the organisation should be interested in knowing the extent of engagement of the staff in activities not directly related to their work-related activities. This will give my organisation the opportunity to know how impactful the staff is on their community. Also, the organisation will know how to re-design their jobs to cater for individual staff differences in relation to their specific jobs in terms of their interests. This has the tendency to benefit the organisation since the interests of the staff has a bearing on their performance.

Concluding Thoughts: All these data/information has human resources implications for the organisation when it comes to recruitment and selection, placement, orientation and induction, training and development, performance appraisal, compensation, succession plan as well as developing employee retention policies.