**AFRICA CENTER FOR PROJECT MANAGEMENT**

**POSTGRADUATE DIPLOMA IN MONITORING AND EVALUATION**

**ERNEST GODSON**

**MODULE 5 ASSIGNMENT**

**Q1.** Explain the difference between data collection and data capture (10mrks)

Data capture is about getting a "snapshot" of a data stream from one source at any discrete point. Ways of capturing data can range from high end technologies (e.g. Synchrotron, sensor networks and computer simulation models) to low end paper instruments used in the field. Data with good metadata attached at the point of capture can expedite data sharing, publishing and citation. While data collection involves gathering quantitative and qualitative information on specific variables with the aim of evaluating outcomes or gleaning actionable insights. Good data collection requires a clear process to ensure the data you collect is clean, consistent, and reliable.

**Q 2:** Explain the benefits of correctly interpreting data in an M&E process. (5 mrks)

**Data interpretation** refers to the implementation of processes through which data is reviewed for the purpose of arriving at an informed conclusion. The interpretation of data assigns a meaning to the information analyzed and determines its signification and implications.

**Informed decision-making:** A decision is only as good as the knowledge that formed it. Informed data decision making has the potential to set project/program leaders apart from the rest of the market pack.

**Anticipating needs with trends identification:**data insights provide knowledge, and knowledge is power. The insights obtained from market and consumer data analyses have the ability to set trends for peers within similar market segments.

**Cost efficiency:** Proper implementation of data interpretation processes can provide businesses with profound cost advantages within their industries.

**Q3.** Explain the main concerns for a data analyst while undertaking the task of data analysis. (10 mrks)

Data analysis is an internal arrangement function done by data analysts through presenting numbers and figures to management. It involves a more detailed approach in recording, analyzing, disseminating and presenting data findings in a way that is easy to interpret and make decisions for the business.

* Collecting and interpreting data
* Analyzing results
* Reporting the results back to the relevant members of the project/program
* Identifying patterns and trends in data sets
* Working alongside teams within the business or the management team to establish business needs
* Defining new data collection and analysis processes

**Q4.** Describe key measures that are mandatory for data quality assurance at program level and explain the value of data quality assurance. (15 mrks).

Data quality is the planning, implementation, and control of activities that apply quality management techniques to data, in order to assure it is fit for consumption and meets the needs of data consumers. Data quality is important to any organization because it provides timely and accurate information to manage accountability and services. It also helps to ensure and prioritize the best use of resources. Thus, high-quality data will lead to appropriate insights and valuable information for any organization. We can evaluate the quality of data in certain aspects. They include accuracy, relevancy, completeness, and uniqueness.

**Main criteria used to measure data quality:**

**Accuracy:**for whatever data described, it needs to be accurate.

**Relevancy:**the data should meet the requirements for the intended use.

**Completeness:** the data should not have missing values or miss data records.

**Timeliness:**the data should be up to date.

**Consistency:** the data should have the data format as expected and can be cross reference-able with the same results.

**Validity:** Do the data entries make sense?

## ****Metrics to Measure Data Quality****

**The ratio of data to errors**

This is the most obvious type of data quality metric. It allows you to track how the number of known errors – such as missing, incomplete or redundant entries – within a data set corresponds to the size of the data set. If you find fewer errors while the size of your data stays the same or grows, you know that your data quality is improving.

Number of empty values

Empty values – which usually indicate that information was missing or recorded in the wrong field — within a data set are an easy way to track this type of data quality problem. You can quantify how many empty fields you have within a data set, then monitor how the number changes over time.

### Data transformation error rates

Problems with data transformation – that is, the process of taking data that is stored in one format and converting it to a different format – are often a sign of data quality problems. By measuring the number of data transformation operations that fail (or take unacceptably long to complete) you can gain insight into the overall quality of your data.

### Amounts of dark data

[Dark data](https://blog.syncsort.com/2015/01/big-data/dark-data-deal/?utm_source=Blog-Post&utm_medium=Blog-In-Text) is data that can’t be used effectively, often because of data quality problems. The more dark data you have, the more data quality problems you probably have.

### Email bounce rates

If you’re running a marketing campaign, poor data quality is one of the most common causes of email bounces. They happen because errors, missing data or outdated data cause you to send emails to the wrong addresses

### Data storage costs

If the [data storage](https://blog.syncsort.com/2017/07/big-data/data-storage-best-practices-data-types/?utm_source=Blog-Post&utm_medium=Blog-In-Text) costs rising while the amount of data that you actually use stays the same is another possible sign of data quality issues. If you are storing data without using it, it could be because the data has quality problems. If, conversely, your storage costs decline while your data operations stay the same or grow, you’re likely improving the data quality front.

### Data time-to-value

Calculating how long it takes to derive results from a given data set is another way to measure data quality. While a number of factors (such as how automated your data transformation tools are) affect data time-to-value, data quality problems are one common hiccup that slows efforts to derive valuable information from data.

**Value of data quality assurance**

**Data confidence**

With high data quality comes trust and understanding in the data. While this might be something that is already present within the team that has produced and validated the data, this is not always the case for everyone within an organization

**Scalability**

With a quality infrastructure in place, it becomes possible for organizations to scale up while still ensuring that the data remains trustworthy and reliable

**Data consistency**

Basically, with a strong infrastructure in place – it becomes simple to ensure that data is consistent throughout the entire organization

**Customer satisfaction**

Benefits of high data quality extend beyond your organization and can also be reaped by your partners and/or customers.

**Save costs**

This ties in neatly with the next benefit- saving costs. It’s not very complex – by maintaining high quality you no longer have to pay the price that comes with poor quality. Mistakes will be detected early and business decisions will be based on trustworthy data.

**Save time**

In line with this, you will also be able to save heaps of time. No more backtracking because of mistakes, no more double-checking your results, no more wasting time due to bad data

**Q5:** In about 350 words, describe the main challenges to effective data interpretation and analysis. (10 mrsk)

### The amount of data being collected

With today’s data-driven organizations and the introduction of big data, risk managers and other employees are often overwhelmed with the amount of data that is collected. An organization may receive information on every incident and interaction that takes place on a daily basis, leaving analysts with thousands of interlocking data sets.

### Collecting meaningful and real-time data

With so much data available, it’s difficult to dig down and access the insights that are needed most. When employees are overwhelmed, they may not fully analyze data or only focus on the measures that are easiest to collect instead of those that truly add value. In addition, if an employee has to manually sift through data, it can be impossible to gain real-time insights on what is currently happening. Outdated data can have significant negative impacts on decision-making.

### Visual representation of data

To be understood and impactful, data often needs to be visually presented in graphs or charts. While these tools are incredibly useful, it’s difficult to build them manually. Taking the time to pull information from multiple areas and put it into a reporting tool is frustrating and time-consuming.

### Data from multiple sources

The next issue is trying to analyze data across multiple, disjointed sources. Different pieces of data are often housed in different systems. Employees may not always realize this, leading to incomplete or inaccurate analysis. Manually combining data is time-consuming and can limit insights to what is easily viewed.

### Inaccessible data

Moving data into one centralized system has little impact if it is not easily accessible to the people that need it. Decision-makers and risk managers need access to all of an organization’s data for insights on what is happening at any given moment, even if they are working off-site. Accessing information should be the easiest part of data analytics.

### Poor quality data

Nothing is more harmful to data analytics than inaccurate data. Without good input, output will be unreliable. A key cause of inaccurate data is manual errors made during data entry. This can lead to significant negative consequences if the analysis is used to influence decisions. Another issue is asymmetrical data: when information in one system does not reflect the changes made in another system, leaving it outdated.

### Lack of support

Data analytics can’t be effective without organizational support, both from the top and lower-level employees. Risk managers will be powerless in many pursuits if executives don’t give them the ability to act. Other employees play a key role as well: if they do not submit data for analysis or their systems are inaccessible to the risk manager, it will be hard to create any actionable information.

### Confusion or anxiety

Users may feel confused or anxious about switching from traditional data analysis methods, even if they understand the benefits of automation. Nobody likes change, especially when they are comfortable and familiar with the way things are done.

### Budget

Another challenge risk managers regularly face is budget. Risk is often a small department, so it can be difficult to get approval for significant purchases such as an analytics system.

### Shortage of skills

Some organizations struggle with analysis due to a lack of talent. This is especially true in those without formal risk departments. Employees may not have the knowledge or capability to run in-depth data analysis.

### Scaling data analysis

Finally, analytics can be hard to scale as an organization and the amount of data it collects grows. Collecting information and creating reports becomes increasingly complex. A system that can grow with the organization is crucial to manage this issue.

**REFERENCES**

Box,G. E. P., Hunter, W. G., and Hunter, J. S. (1978), Statistics for Experimenters: An Introduction to Design, Data Analysis and Model Building, John Wiley and Sons.

Pope, C., Ziebland, S., and Mays, N. (2000). ‘Analysing qualitative data’. British Medical Journal, 320 : pp. 114–116 [Online]. Available at