How is data collected? What are the different entry points of data and how do you improve the quality of the data you receive?

Organizations and many businesses nowadays collect data from various sources such as their clients, employees, vendors, customers. This data can be used to understand their association and to improve and expand their operations. Organizations use several methods to collect the data ranging from the traditional face-to-face interviews, surveys to highly sophisticated methods. Few of the data collection methods include interviews, questionaries and surveys, observations, documents, and records, focus groups and oral histories. These data collection methods can be classified into Qualitative and Quantitative methods. Questionnaires, Surveys, documents, and records comes under quantitative methods which deals with something that can be counted or measured numerically. These quantitative methods answer the question “how much?” But not “why?”. This is where Qualitative methods comes into picture. Qualitative methods include interviews, focus groups, observations, and oral histories. These methods provide a depth in understanding the raw data. But quantitative methods are the most popular ones because they are relatively straightforward. Data can be valuable but too much of it can be cumbersome and no data is better than wrong data. Collecting right data helps us in understanding and creating useful insights.

Different entry points of Data into an organization include data files that one can receive from different sources, Manual data entry through the UI screens, data received from multiple databases maintained by different teams within an organization. Preprocessing this data received is an important step to improve the quality. There are different ways to improve the quality of data that we receive or enter into the business systems. Few of them are listed below.

**Setting data entry standards:**

Setting data entry standards such as setting mandatory fields, hierarchy data and processes to delete the data or making any records inactive help in reducing missing data. Data requirements should be established by the corresponding business teams to complete the entry of a new record into the system.

**Providing options:**

If the fields defined have a set of standard set of values, instead o providing the user a free text format, a dropdown should be provided there by reducing the scope of having the same name in different formats. For example: Country Name field. United states can be written in various ways and providing a dropdown value as USA will help in removing any redundancy.

**Key data:**

Determining the minimum fields needed to proceed with processing a record can reduce the need to add any extra data. To make sure that the key data is entered these fields should be made mandatory.

**Duplicate detection & cure:**

Duplicates in the data should be identified and an action should be taken such as merging these records into a single record or purging one record that has old data to eliminate the duplicates.

**Duplicate prevention:**

Everybody wants to prevent duplicates from entering into their systems. This can be done by setting up the rules that identify any potential duplicate data there by preventing it from entering into the system.

**Integration:**

If there are multiple systems in the organizations that consume the data, Instead of sourcing data into multiple systems we can enter the data into one system and then integrate it with other systems so that there’s a less chance of errors.

**Reviewing Data Quality**:

Standards set for the data entry such as the mandatory fields needs to be reviewed periodically to ensure that they still make sense for the data that may change over time.

**Training and reminding:**

Everyone involved in the data entry needs to be made aware of the process. New persons join in the team should be trained on these procedures and a constant feedback needs to be taken to improve any of the procedures.