Alyssa Basile

15 May 2022

Southern New Hampshire University

DAT 300 Module 2 Journal

Data Quality Assessment

When merging between 2 companies, the data set must match for it to be done properly. The data also must be accurate, viable, consistent, current, and valid. In the Data Analytics Life Cycle, there are 6 steps: Objective, Understanding the Data, Data Cleaning and Data Transformation, Data Enhancement, Data Analytics, and Data Visualization (Sood, 2017). They all play a vital role in data quality assessment.

To detect a problem in the data set and determine whether it represents an organizational problem, one must understand the objective. Understanding the purpose, the why, and what outcome to expect. Beginning to understand the data set is important because when we know the purpose of the data, know what data policies/governance policies the company follows, and detect any data anomalies, we can figure out whether the data set will portray a problem or not. Using data quality SLAs (Service Level Agreements) is ideal because it specifies the organization’s expectations for response and remediation (Technics Publications, LLC). This is a good step to ensuring the data set is usable. When starting an assessment of a data set, it helps to expect the data is dirty. Detecting for anomalies is a good start in detecting whether it is usable or if it needs to be cleaned, transformed, and enhanced. One of the tools that will be helpful in ensuring the data set matches before merging is using a data dictionary that contains the header, the description, the data type, and sometimes, the formula used for quantitative data.

To verify that all accurate data is present to complete a data quality assessment and whether it is consistent and complete, I would use a tool, such as Excel, to detect any rows that is missing a value. If it is missing a value, determine whether the value can be obtained. If it cannot be obtained, determine whether it will affect the overall analysis and it needs to be removed or if it does not affect the analysis and it can be kept. Using data dictionary, determine if the data types are correct and ensuring each value is in the right format and in the right columns. Which leads us to decide what data to keep and what to discard. Some of the things to look out for when making the decision are:

* Does it serve a purpose for the analysis?
* Does it comply any of the data regulations set up by data governance and the company?
* Is it missing the data it needs? Can it be obtained?
* Is the data current?
* Is it a duplicate?
* Is it accurate?

(Whyte, 2014)

The obstacles that can interfere with providing an accurate data quality assessment are trying to assess big data can be difficult and time-consuming. Big data contains wide variety of data, contains high data volume, and is complex (*Big Data: What It Is and Why It Matters*). Another obstacle is, if it is international, each area contains different regulations/laws, and it can be difficult to comply (Cai, 2015). To safeguard sensitive data, using DAMA DMBOK guide and skimming through chapter 7 (data security management) and knowing the data privacy policy the company follows would provide different ways to ensure sensitive data is protected (Technics Publications, LLC).

References:

*Big Data: What It Is and Why It Matters*. SAS. Retrieved May 15, 2022, from <https://www.sas.com/en_us/insights/big-data/what-is-big-data.html#:~:text=Big%20data%20is%20a%20term,with%20the%20data%20that%20matters>.

Cai, L., & Zu, Y. (2015, May 22). *The Challenges of Data Quality and Data Quality Assessment in the Big Data Era*. Data Science Journal. Retrieved May 15, 2022, from <https://datascience.codata.org/articles/10.5334/dsj-2015-002/#:~:text=At%20present%2C%20big%20data%20quality,as%20sales%20and%20inventory%20data>.

Gupta, A. (2021, June 25). *The 7 Most Common Data Quality Issues*. Collibra. Retrieved May 15, 2022, from <https://www.collibra.com/us/en/blog/the-7-most-common-data-quality-issues>

Sood, S. (2017, May 10). *Data Analytics Life Cycle: What is It? How to Approach?* VokseDigital. Retrieved May 15, 2022, from <https://www.voksedigital.com/data-analytics-life-cycle/>

Technics Publications, LLC. The DAMA Guide to The Data Management Body of Knowledge (DAMA-DMBOK Guide). Retrieved May 15, 2022, from <https://learn.snhu.edu/d2l/lor/viewer/view.d2l?ou=6606&loIdentId=24443>

Whyte, A. (2014, October 31). *Five Steps to Decide What Data to Keep*. DCC. Retrieved May 15, 2022, from <https://www.dcc.ac.uk/guidance/how-guides/five-steps-decide-what-data-keep>