

Data Quality with Continuous Monitoring

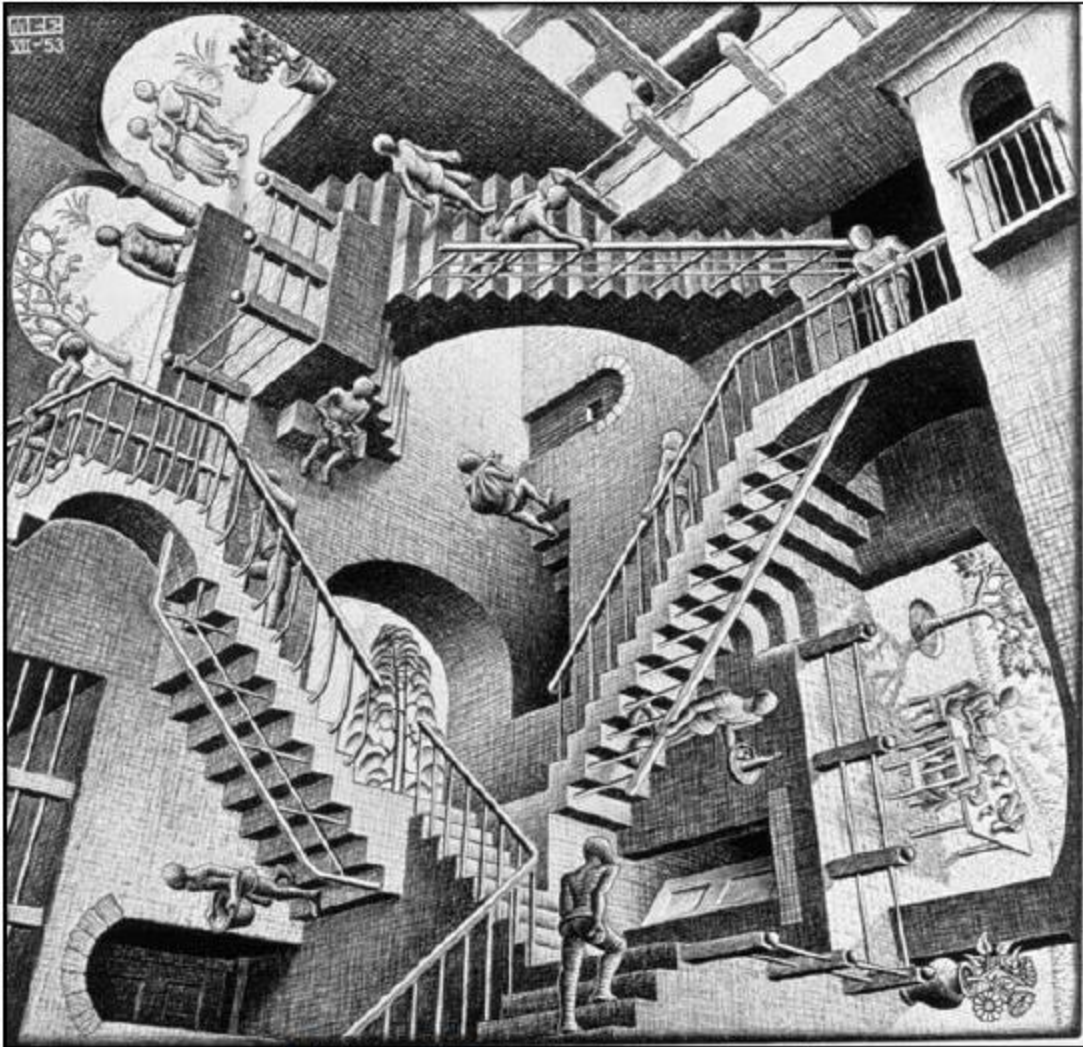
Saran Ketharam, Core Architecture, 2018

“In God we trust. All others bring data,” said the multi-talented William Edwards Deming. Data is a key ingredient for business-driven initiatives. The bar for quality data is set very high right from the outset for consumers. Poor data quality puts organizations at risk, jeopardizes key operational systems, and leads to distrust of key individuals who use it to make critical business decisions. Hence, managing the quality of enterprise data is extremely important.

Where does the problem start? It is easy to blame the source systems or the initiation point for being the cause of bad data quality. While this may hold true in some cases, there are several other reasons for bad data in your enterprise:

- **Merger/Acquisition:** When data is brought in from old systems, focus is given to current data points while others are simply moved for use at a later time. The old data points may not make sense in the new environment.
- **Decay:** Studies show that name and address records decay at 2 percent per month. When the warehouse is used for customer information, there is a good chance of missing data, such as changed addresses, names, etc.
- **Migrations/Consolidations:** Not all data is equal. Therefore, when data from disparate sources merge, overall data quality suffers. Similar issues occur when data sets are moved from one system to another.

Are these the main sources of data quality issues? While they cover the majority of problems, a larger pain point is in the perception of data use. The picture below depicts the challenge of business relativity; in this case the deemed fitness for use of the data. While a myopic vision may seem perfect in each fiefdom, the overall picture shows the state of union.



Relativity by M.C. Escher (<http://www.mcescher.com/>)

As we start the solution path, it is important to understand the data. The following metrics help us understand the problem in our data sets:

- **Consistency:** The degree to which a unique piece of data holds the same value across multiple data sets
- **Correctness:** The degree of conformity of a data element or a data set to an authoritative source that is deemed to be correct or the degree the data correctly represents the truth about a real-world object
- **Completeness:** The degree to which all required occurrences of data are populated
- **Uniqueness:** The extent to which all distinct values of a data element appear only once
- **Timeliness:** The degree to which data is available when it is required

Data quality can also be measured for business satisfaction, opportunity vs. risk, and compliance. Which measure to use depends on factors such as organization maturity and criticality of the data sets. The assessment should quantify the extent to which data quality will affect the organization. This will also provide a benchmark that can be used when measuring the data sets on a frequent basis.

When you understand the problems in the data, the first step is to tackle the low hanging fruit. The quality of the data of some elements can be improved with techniques like standardization, cleansing, and transformation. Several software tools are available to assist in writing business rules. The tools help deduce, validate, and create consistency for each data element. It is also advisable to measure the quality after the transformation to evaluate the worthiness of the rules.

For long-term solutions, root cause needs to be determined. The repeated application of the same correction to the same data is not viable in the long run. This does not address issues like changes in source data and new additions of data elements. So, the answer lies not in treating the symptom but in identifying and addressing the root cause.

Data quality measurements and rules must be applied only to critical and important data sets in the organization. A transaction record may have 30 columns but only four may be candidates for quality checks.

Finally, existing and new data sets must be maintained through continuous monitoring. Monitoring data quality is not a one-time exercise. Continuous monitoring can provide the basis for action to manage the business process for accuracy and to improve the data quality practices. These solutions should act independently of a system to monitor data quality filtering on defined parameters to match control tolerances that effectively balances operational efficiencies against financial and compliance risk.