

## Is Data Quality the Same As Data Accuracy?

### Accurate data can be of poor quality if it doesn't suit its intended purpose

We are living in the digital age, and data have a universal presence that directly or indirectly affects our lives, even when we're not aware of it. Hence, data quality is an important topic of discussion. Data quality isn't only an aspect of data that determines their fitness for use, but is also a function or sub-discipline of data management.

Data quality can be defined as evaluating data's fitness to use (that is, serve their purpose) in a given context. Sustaining high quality data is a challenge that most organizations face, and the data quality arena is surrounded by its own set of myths. This misleads people when it comes to making data quality management-related decisions. These myths can slow down, hinder, or put a stop to an organization's data quality management efforts or the deployment of data quality projects or initiatives.<sup>1</sup>

"Data quality is data accuracy" is one of the most common myths of data quality. The general misconceptions are that data quality is synonymous to data accuracy, or that data quality is *only* about data accuracy. When people think about high quality in relation to data, they tend to think about the accuracy aspect only. When an organization is under the influence of this myth, data accuracy becomes its only data-quality improvement goal.

#### What is data accuracy?

Data accuracy refers to how closely or how well the data stored in a system reflect reality. It is the degree to which data correctly describe the characteristics of the real-world object, entity, situation, phenomena, or event. Measuring data accuracy requires that an authoritative source of reference be identified and available to compare the data against. If the data show that Arvind Kumar lives in Australia but he actually lives in the United States, then the data are inaccurate. However, without an authoritative source of reference, such as a utility bill that contains the home/office address, it is not possible to ascertain where Arvind Kumar actually lives.

Data must not only reflect reality, they must also be complete, valid, and consistent. For data to be accurate, they need to be *complete* in the first place (that is, values need to be present). For data to be *valid*, they must conform to some sort of standard. As a validity example, as per ISO's list of country codes, AU is a *valid* country code, but AAA is not. Data can be valid but not accurate. For example, if a person's postal address records "AU" as the country code when the person is actually residing in the United States, then the data are valid (because AU is a valid code) but fail the accuracy test. *Consistency* means that exactly the same data appear the same way across different data sets. As a consistency example, if one data set records a name as Arvind Kumar, but the other data set reports this person's name as Arvind Kumar, then the data are inconsistent; at least one of the sets is inaccurate. If data are accurate, then they meet all the tests above.

### What is data quality?

Although data accuracy is one of the important characteristics or dimensions of data quality, and therefore shouldn't be overlooked, accuracy alone doesn't completely characterize the data quality. Data quality has several dimensions, known as data quality dimensions, that enable the measurement of the quality of data. These dimensions include but are not limited to completeness, uniqueness, granularity, precision, consistency, accessibility, security, traceability, conformity/validity, timeliness, integrity, currency, volatility, and so forth.

For example, if data are accurate but not delivered in time for reporting purposes, the data wouldn't be considered of high quality because the intended purpose won't be served. Data might also be accurate but not granular enough to serve the business need. If data are accurate but not accessible to authorized people, they are also not of much use and, thus, the data quality is poor.

Undeniably, data are normally considered of poor quality if erroneous values are associated with the real-world entity or event. However, data quality is about striking a balance between all data quality dimensions. Depending on context, situation, the data themselves (e.g., master data, transactional data, reference data), business needs, and the industry sector, different permutations and combinations of data-quality dimensions would need to be applied.

To learn more about data quality and its myths, challenges, critical success factors, strategy, DQ dimensions, data profiling, and more, including how to measure data quality dimensions, implement methodologies for data quality management, and data quality aspects to consider when undertaking data intensive projects, please read *Data Quality: Dimensions, Measurement, Strategy, Management and Governance* (Quality Press, 2019). This article draws significantly from the research presented in that book.

Readers may please note that D. L. Shah Trust brings out two e-journals on a fortnightly basis. These are mailed to those persons or institutions who are desirous of receiving them:

These two e-journals are:

1. Safety Info
2. Quality Info

If you or your friends or colleagues wish to receive these journals, you may send us an email requesting for the same. There is no charge for these journals. Our e-mail address is:

[dlshahtrust@yahoo.co.in](mailto:dlshahtrust@yahoo.co.in) or [haritaneja@hotmail.com](mailto:haritaneja@hotmail.com) or [dlshahtrust@gmail.com](mailto:dlshahtrust@gmail.com)

You can also access these journals on our website: [www.dlshahtrust.org](http://www.dlshahtrust.org)

Published by : D. L. Shah Trust,  
Room No. 16, 1<sup>st</sup> Floor, Gool Mansion, Homji  
Street, Mumbai 400 001  
email: [dlshahtrust@yahoo.co.in](mailto:dlshahtrust@yahoo.co.in)  
Ph: 022-22672041  
Subscription: Free on request  
(soft copy only)

Edited by : Hari K Taneja, Trustee,  
D. L. Shah Trust  
email: [dlshahtrust@gmail.com](mailto:dlshahtrust@gmail.com)  
Phone: 022-2309 6529  
Subscription: Free on request  
(soft copy only)