

[Strategy](#) [Governance](#) [Architecture](#) [Data Mgmt](#) [Quality](#) [Modeling](#) [Operations](#) [Metadata](#) [MDM](#)  
[BI](#) [Applications](#) [Big Data](#)

**Subscribe to The Data Administration Newsletter Today - it's FREE!**

**REGISTER HERE!**

# Current State Analysis of Your Data – Part 1

Published: January 19, 2022 | 12:09 am · Vanessa Lam



Whether you're stepping into a new organization as a data lead or trying to overhaul your data infrastructure, the first step in the process is to understand how your organization currently uses data. While that may sound simple, it can be an intimidating process to start. This is the beginning of a series of articles meant to give you a jumping-off point to understanding your data organization. In this series, I will be outlining the important categories to consider as well as some example questions to ask to better understand the current state of your data organization.

There are four main categories to consider when doing a current state analysis of an organization's data: Data Quality, Data Freshness, Data Culture, and Data-Driven Outcomes. This first article will give you definitions of these four terms and why they are important, and also address the first term, Data Quality. In future articles, I will provide some questions (and some of their potential answers) to better understand these categories.

## Data Quality

Data Quality refers to the accuracy and precision of your data. Is the data being input properly? Is there high enough coverage on data fields to trust the data? Good data quality is essential in order to know that insights drawn from this data can be trusted.

## Data Freshness

Data Freshness refers to whether or not data is available where and when you need it to be. It also used to assess the fit of the tools and processes to the business needs. Practically, we will recognize these as the

cadence of data input and data refreshes as well as the reliability of processes and tools.

## Data Culture

Data Culture refers to the way that people in the organization interact with data. Who is able to interact with it? How do they feel about it? How do they speak about it? This is critical to understand if employees are willing and interested to work with data or if increased education and discussions are required to improve adoption of healthy data practices.

## Data Outcomes

Data Outcomes refers to an organization's main goal for their data. Is it a revenue driver? Does it dictate decision making? Do they need it for fundraising? These questions will indicate how data is being used now, but also potentially areas for it to grow.

These are the four major categories that we will be focusing on when doing a current state analysis on the data. As you can see, this is a wholistic look at data; we are not merely looking at the data itself, but the treatment and sentiment toward data across the organization. It is essential to consider all these factors when trying to build a sustainable data organization. Follow along for future articles that will dive deeper into the questions we can ask to better understand these four categories!

This section of this article will focus on Data Quality: what it is, why it is important, and what questions to ask to determine its current state.

The questions are organized by stakeholder group to facilitate usability; hopefully you can use this as a template to start your Current State Analysis journey. A few definitions before we begin – note that these groups are not mutually exclusive:

*People who Input Data:* These are people who collect and/or input data into the system. For example, salespeople inputting their sales numbers, or survey creators.

*People who Manipulate and Analyze Data:* These are people who organize the data and create analyses. This includes Data Engineers, Business Intelligence Professionals, and Data Analysts.

*People who Make Decisions based on Data:* These are the people who use the data to make decisions. This may be a sales manager deciding where to invest resources, a product manager understanding product use demographics, or an executive trying to cut costs.

## What is Data Quality?

Data quality is a measure of the condition of your data, including accuracy, consistency, and completeness. Data quality can be affected in any part of the process, from collection to ETL to analysis. Accuracy refers to how well your data reflects the truth while consistency refers to how well individual data elements match each other. An example of accuracy and consistency in data collection is when a Sales Representative fills out their sales region (e.g., Pacific Northwest). If the input is a manual text field, it is possible for them to spell the region incorrectly (Pacific Nortwest) or even capitalize the region incorrectly (pacific northwest). This is a failure of data consistency. A possible correction here would be to supply a drop-down list of the possible regions (e.g., Pacific Northwest, Midwest, South, East). However, this scenario still provides opportunities for inaccuracy in choosing the incorrect region. An example of incomplete data here would be the Sales Representative leaving this field blank or not filling out the sales record entirely.

Potential solutions to these issues could be making the Region field mandatory or even pre-filling the field if Sales Representatives are assigned to specific regions. However, be aware that putting these rules on data collection may slow down data collection or even disincentivize people from completing forms if they think that it is too difficult or restrictive. When assessing data quality rules, it is important to evaluate whether guardrails and processes can be added to make positive change, or if the restrictions will cause unintentional side-effects.

## Why is Data Quality Important?

Good data quality is essential to know that insights drawn from the data can be trusted. Most analysts have the experience of their manager questioning the numbers from the data, especially if they have some anecdotal knowledge that implies that the data is completely incorrect. Better data quality will reduce the number of times that this happens as the data will be more accurate and complete, meaning that it is a better reflection of the truth. In the cases where there are still disagreements, they may be more substantive and provide opportunities to talk about where anecdotal evidence and collected data may differ.

## Questions to Determine Current State of Data Quality

### *To Those Who Input Data*

These questions are designed to better understand the guardrails around data collection. It is important for data to have strong validations, such as choosing from a dropdown list instead of allowing free text; however, be aware that this could backfire if the validations prevent accurate data entry (for example, missing categories in a dropdown) or make it too difficult to fill out (making long forms with many mandatory fields can make a form unappealing to fill out). Often, data quality issues are rooted in poor data collection.

- What is your process for inputting data?
- Are there specific fields for inputting commonly collected information or is it largely free text?
- Are there selection options for appropriate fields? (e.g., A dropdown with “True” or “False” selections or a free text field where you can type “True” or “False”)

- Are there automatic or pre-loaded values in the fields? Are you typically using the pre-loaded value, or do you frequently need to change it?
- What is your biggest frustration when inputting data?

### ***To Those Who Manipulate and Analyze Data***

These questions are designed to understand how the data looks to someone who is in charge of cleaning, understanding, and verifying the numbers to be presented to business stakeholders. They are often the ones who feel most impacted by poor data quality, and therefore likely have a good idea of where the data is weak.

- When doing analyses, how confident do you feel about the outcome based on the data?
- Do you feel like you need to put caveats on your data based on low confidence intervals?
- Are there data fields that are being used for multiple purposes? Are there data fields that still exist in the data, but everyone says to avoid?

### ***To Those Who Make Decisions based on Data***

These questions are designed for people who are using data to make decisions. Often, they are not the ones who have done the analysis themselves, so they may not know the intricacies of the data as well. However, they are usually keenly aware of the business implications of their decisions and have a good sense of what aggregate numbers make sense. For example, a sales executive will likely have a good idea if the number presented is approximately correct based on their experience with the industry, company, and knowledge of their sales representatives. It is important to hear their opinion because ultimately these are the customers of the data. If they aren't happy with the quality, then some changes may need to take place.

- Do you find yourself seeing the data and overriding data-backed decisions because you don't think that what you're reading is a reflection of reality?
- What parts of the data are you confident in? What parts are you concerned by?

## **Conclusion**

Data quality is a bedrock of your data systems. Without good quality data, analyses cannot be trusted, decisions cannot be made with data, and eventually people will stop relying on data altogether. Before attempting to implement a strong data ecosystem, it is important to have a plan to improve data quality throughout the organization, from data collection and data processes to analyses. This article is the first in a series discussing the important considerations when assessing your Current State of Data. Follow along for the next article about Data Freshness – measuring whether the data is up-to-date and reliable!

---