

ASSESSING YOUR DATA QUALITY AND SEVEN STEPS TO IMPROVING IT.

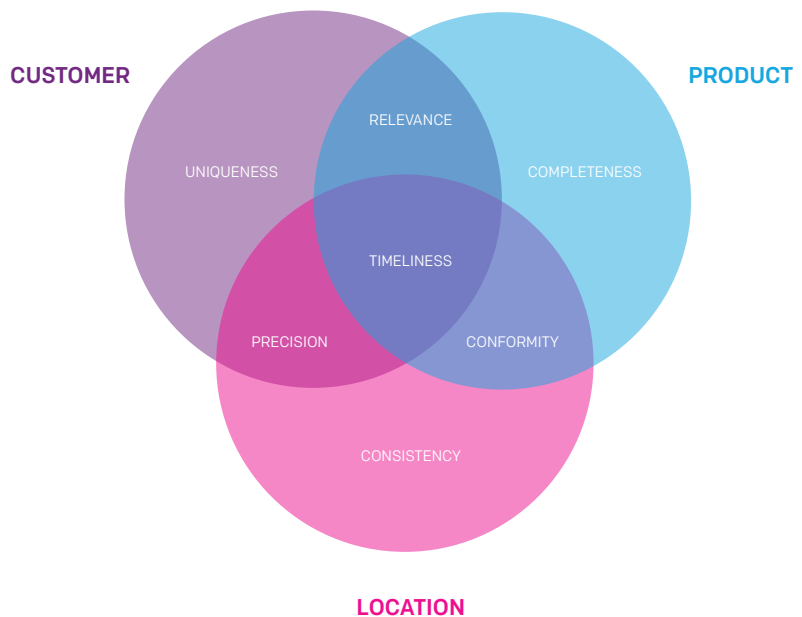


Assessing your data quality.

Most businesses these days are drowning in data. Yet data isn't about quantity. It's about quality. Because data quality is at the heart of successful business operations.

Assessing and improving your data quality starts by understanding that data is multi-dimensional. You can't just improve one thing and expect great results. You need to measure and improve a range of data quality dimensions. And it starts by understanding the dimensions themselves.

CAUSES OF LOW-QUALITY DATA



UNIQUENESS:

This is the most commonly addressed data quality dimension for customer data. That's because customer data is often riddled with duplicates, where two or more contacts, leads, accounts, or any set of rows describe the same real-world entity.

Fortunately, there are several remedies available, from intercepting the duplicates at the edge (such as your marketing automation systems), to bulk **de-duplication** at the core (your CRM data. And doing this across multiple databases).



COMPLETENESS:

This dimension often poses problems in product master data. One reason is that completeness means different requirements for different categories of products across multiple systems. Enriching this data **at the edge in real-time, or at the core in batch** as data decays, will create an environment of trust and efficiency, while improving **customer experience and conversion rates**.

CONSISTENCY:

When working with location or address data, consistency can be a significant challenge. So it's imperative to normalize and segment address data, such as the different postal address formats around the world, into a common structure.

Consistency can also be a challenge at the intersection of the local domain and the customer domain. That's because different use cases require different precision for territory management. It could be as granular as postal codes or states segmented into territory groupings. Or in logistics, where ship-to addresses vary significantly from corporate locations' bill-to addresses.

And then there's conformity of product data, which is also critical in locations. Take, for example, unit measurement for shipping and logistics. In the United States, imperial measurements are required, while for overseas organizations, you need to use metric units.

TIMELINESS:

If your data isn't available when you need it, it's often useless. So timeliness is the overarching data quality dimension to measure and improve. Timeliness can be measured as the time between when information is expected and when it is readily available for use.

Data Conformity, at the intersection of Data Consistency and Data Completeness, is how well data adheres to standards and how well it's represented in an expected format.

Data Precision, at the intersection of Data Uniqueness and Data Consistency, is the depth of knowledge encoded by the data.

Data Relevance, at the intersection of Data Completeness and Data Uniqueness, is the closeness between data needed and data output provided.

Other important data quality attributes to measure on an ongoing basis are:

- **Data accuracy:** The real-world alignment or alignment with a verifiable source
- **Data validity:** Ensuring data is within the specified business requirements
- **Data integrity:** Making sure the relations between entities and attributes are technically consistent

Data quality management

Now that you understand the many dimensions of data quality, it's time to understand how to manage and improve it.

The goal is to create a balanced set of always-on remedies that will help prevent future data quality issues. Ultimately, it will be important to purge data that does not meet the data quality Key Performance Indicators (KPIs) needed to achieve immediate and ongoing business objectives.

These data quality KPIs will typically be measured on the core business data assets within the dimensions outlined above. And of course, these KPIs must relate to the KPIs used to measure the business performance in general.

Seven disciplines to higher quality data and management


To help you assess, improve, and manage data quality, we've outlined seven data disciplines:

- Data Governance
- Data Matching
- Data Profiling
- Digital Asset Management (DAM)
- Data Quality Reporting
- Master Data Management (MDM)
- Customer Data Integration (CDI)

Data Governance

A data governance framework must lay out the data policies and standards that set the bar for data quality KPIs—including which KPIs are needed and which data elements should be addressed. This also includes the business rules that must be adhered to and underpinned by data quality measures.

The framework must encompass the organizational structures needed to achieve the required level of data quality. This includes the creation of a data governance



committee with roles such as data owners, data stewards, and data custodians—depending on what makes sense in a given organization.

A business glossary is another valuable outcome from data governance. The business glossary is a primer to establish the metadata used to achieve common data definitions within an organization, and eventually in the business ecosystem where the organization operates.

Data Matching

Data matching technologies can be used to de-duplicate a single database, to find matching entities across several data sources, or to find an object within a single data source such as leads and contacts in Salesforce.com.

It's often based on data parsing, where names, addresses and other data elements are split into discrete data elements. For example, an envelope-type address is split into building name, unit, house number, street, postal code, city, state/province, and country. This may be supplemented by data normalization and segmentation, such as using the same value for street, str and st.

There are many examples of data matching, all of which are based on match-codes. For example, there's soundex, a phonetic algorithm for indexing names by sound. There's also fuzzy logic, which uses degrees of truth rather than true or false. And machine learning, where algorithms are used to determine if two or more data records are describing the same real-world entity—typically a person, a household or an organization.

Data Profiling

Data profiling directly measures data integrity and can be used as another input in the measurement of other data quality dimensions. It's often supported by dedicated technology to understand the data assets involved in data quality management.

When you profile data, how it's structured is just important as the data element itself. Many times, data assets have been populated over the years by different departments operating under varying business rules and objectives, so the completeness and structure of the data may be all over the board. When you have consistent and relevant structure levels, you're able to access and distribute your data more efficiently.



Digital Asset Management (DAM)

Digital assets are images, text documents, videos and other files often used in conjunction with product data. When viewed through the data quality lens, the challenges for this kind of data is making sure the tags (metadata) are tagged with correct, detailed, and relevant information. For example, it's much more efficient if a photo is tagged with the content and context, such as "a man walking on the beach," rather than just "man."

LEARN MORE ABOUT METADATA.

Data Quality Reporting

To effectively manage and improve data quality, you need good reporting. That's why it's essential to operate from a dashboard that tracks measurements of your data quality KPIs—using the findings from data profiling as input.

These findings are especially helpful for measuring data uniqueness. Free apps like [Dupe Dive](#) and [Field Trip](#) can give you even more insight into duplicate data entries and how your fields are being used.

Tracking quality issues is important as well. This can be done with a data quality issue log, where known data quality issues are documented, and the preventive and data cleansing activities are monitored.

Master Data Management (MDM)

Master Data Management and Data Quality Management (DQM) are tightly coupled together. MDM and DQM will be a part of the same data governance framework and fall under the roles of data owners, data stewards, and data custodians.

Master data assets are often where data profiling activities happen. And when doing data matching, the results must be kept in master data assets that control the merged and purged records, as well as the survivorship of data attributes relating to those records.

Having a DQM framework in place is a sustainable way to prevent data quality issues and reduce the need to launch data cleansing activities over and over again.



Customer Data Integration (CDI)

In many organizations, customer master data is sourced from both internal platforms and 3rd party acquired data. This can include a range of applications such as self-service registration sites, Customer Relationship Management (CRM) applications, ERP applications, customer service applications, and perhaps many more.

Integration requires setting up a technical platform for compiling the customer master data, and ensuring the quality of that source of truth. This involves all of the data quality disciplines to ensure the best data consistency and accuracy of the data.

Quality data leads to quality operations

Quality data is core to successful business operations—because it impacts almost every initiative. When you understand the many dimensions of data quality, and follow the disciplines to improve and manage it, you create a solid foundation that can drive everything else.

RINGLEAD CAN HELP YOU ASSESS YOUR DATA QUALITY AND GUIDE YOU IN WAYS TO IMPROVE IT. VISIT OUR [RINGLEAD.COM](https://www.ringlead.com) OR CALL 888-240-8088 TO LEARN MORE.