

h Accelerator

data management and why is it important?



Search the TechTarget Network

[Home](#) > [Data quality tools](#) > [Database management](#) > [data quality](#)


DEFINITION

data quality

Posted by: [Margaret Rouse](#) [WhatIs.com](#)



Contributor(s): [Jack Vaughan](#)


 Data quality is a measure of the condition of data based on factors such as accuracy, completeness, consistency, reliability and whether it's up to date. Measuring data quality levels can help organizations identify data errors that need to be resolved and assess whether the data in their IT systems is fit to serve its intended purpose.

The emphasis on data quality in enterprise systems has increased as data processing has become more intricately linked with business operations and organizations increasingly use data analytics to help drive business decisions. Data quality management is a core component of the overall [data management](#) process, and data quality improvement efforts are often closely tied to [data governance](#) programs that aim to ensure data is formatted and used consistently throughout an organization.

Why data quality is important

Bad data can have significant business consequences for companies. Poor-quality data is often pegged as the source of operational snafus, inaccurate analytics and ill-conceived business strategies. Examples of the economic damage that data quality problems can cause include added expenses when products are shipped to the wrong customer addresses, lost sales opportunities because of erroneous or incomplete customer records, and fines for improper financial or regulatory compliance reporting.

THIS ARTICLE IS PART OF

 **What is data management and why is it important?**

Which also includes:

- [Top database cloud migration considerations for enterprises](#)
- [Evaluating data warehouse platform options and your need for one](#)
- [Why organizations need a solid data governance strategy](#)

An oft-cited estimate by IBM calculated that the annual cost of data quality issues in the U.S. amounted to \$3.1 trillion in 2016. In an article he wrote for the *MIT Sloan Management Review* in 2017, data quality consultant Thomas Redman estimated that correcting data errors and dealing with the

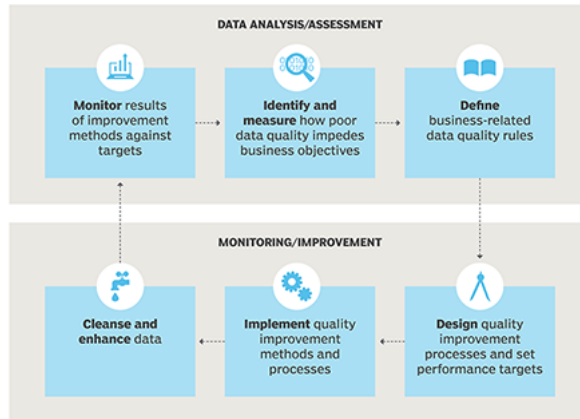
business problems caused by bad data costs companies 15% to 25% of their annual revenue on average.


In addition, a lack of trust in data on the part of corporate executives and business managers is commonly cited among the chief impediments to using business intelligence (BI) and analytics tools to improve [decision-making](#) in organizations.

h Accelerator

data management and why is it important?

The virtuous cycle of data quality management



 Steps in the data quality improvement process.

What is good data quality?

Data accuracy is a key attribute of high-quality data. To avoid transaction processing problems in operational systems and faulty results in analytics applications, the data that's used must be correct. Inaccurate data needs to be identified, documented and fixed to ensure that executives, data analysts and other end users are working with good information.

Other aspects, or dimensions, that are important elements of good data quality include data completeness, with [data sets](#) containing all of the data elements they should; data consistency, where there are no conflicts between the same data values in different systems or data sets; a lack of duplicate data records in [databases](#); data currency, meaning that data has been updated as needed to keep it current; and conformity to the standard data formats created by an organization. Meeting all of these factors helps produce data sets that are reliable and trustworthy.

How to determine data quality

As a first step toward determining their data quality levels, organizations typically perform data asset inventories in which the relative accuracy, uniqueness and validity of data are measured in [baseline](#) studies. The established baseline ratings for data sets can then be compared against the data in systems on an ongoing basis to help identify new data quality issues so they can be resolved.

Another common step is to create a set of data quality rules based on business requirements for both operational and analytics data. Such rules specify required quality levels in data sets and detail what different data elements need to include so they can be checked for accuracy, consistency and other data quality attributes. After the rules are in place, a data management team typically conducts a data quality assessment to measure the quality of data sets and document data errors and other problems -- a procedure that can be repeated at regular intervals to maintain the highest data quality levels possible.

Various methodologies for such assessments have been developed. For example, data managers at UnitedHealth Group's Optum healthcare services subsidiary created the Data Quality Assessment Framework (DQAF) to formalize a method for assessing its data quality. The DQAF provides guidelines for measuring data quality dimensions that include completeness, timeliness, validity, consistency and integrity. Optum has publicized details about the framework as a possible model for other organizations.

The International Monetary Fund (IMF), which oversees the global monetary system and lends money to economically troubled nations, has also specified an assessment methodology, similarly known as the Data Quality Assessment Framework. Its framework focuses on accuracy, reliability, consistency and other data quality attributes in the statistical data that member countries need to submit to the IMF.

Data quality management tools and techniques

Data quality projects typically also involve several other steps. For example, a data quality management cycle outlined by data management consultant David Loshin begins with identifying and measuring the effect that bad data has on business operations. Next, data quality rules are defined, performance targets for improving relevant data quality metrics are set, and specific data quality improvement processes are designed and put in place.

Those processes include data cleansing, or [data scrubbing](#), to fix data errors, plus work to enhance data sets by adding missing values, more up-to-date information or additional records. The results are then monitored and measured against the performance targets, and any remaining deficiencies in data quality provide a starting point for the next round of planned improvements. Such a cycle is intended to ensure that efforts to improve overall data quality continue after individual projects are completed.

Software tools specialized for data quality management can match records, delete duplicates, validate new data, establish remediation policies and identify personal data in data sets; they also do [data profiling](#) to collect information about data sets and identify possible outlier values. Management consoles for data quality initiatives support creation of data handling rules, discovery of data relationships and automated data transformations that may be part of data quality maintenance efforts.

h Accelerator

data management and why is it important?

data repositories to data quality
management processes are often

incorporated into data governance programs, which typically use data quality metrics to help demonstrate their business value to companies, and master data management ([MDM](#)) initiatives that aim to create central registries of master data on customers, products and supply chains.

Benefits of good data quality

From a financial standpoint, maintaining high levels of data quality enables organizations to reduce the cost of identifying and fixing bad data in their systems. Companies are also able to avoid operational errors and business process breakdowns that can increase operating expenses and reduce revenues.

In addition, good data quality increases the accuracy of analytics applications, which can lead to better business decision-making that boosts sales, improves internal processes and gives organizations a competitive edge over rivals. High-quality data can help expand the use of [BI dashboards](#) and analytics tools, as well -- if analytics data is seen as trustworthy, business users are more likely to rely on it instead of basing decisions on gut feelings or their own spreadsheets.

Effective data quality management also frees up data management teams to focus on more productive tasks than cleaning up data sets. For example, they can spend more time helping business users and data analysts take advantage of the available data in systems and promoting data quality best practices in business operations to minimize data errors.

Emerging data quality challenges

For many years, the burden of data quality efforts centered on structured data stored in [relational databases](#) since they were the dominant technology for managing data. But the nature of data quality problems expanded as [big data](#) systems and cloud computing became more prominent. Increasingly, data managers also need to focus on the quality of unstructured and semistructured data, such as text, internet clickstream records, sensor data and network, system and application logs.

The growing use of artificial intelligence (AI) and [machine learning](#) applications further complicates the data quality process in organizations, as does the adoption of real-time data streaming platforms that funnel large volumes of data into corporate systems on a continuous basis. In addition, data quality now often needs to be managed in a combination of on-premises and cloud systems.

Data quality demands are also expanding due to the implementation of new data privacy and protection laws, most notably the European Union's General Data Protection Regulation ([GDPR](#)) and the California Consumer Privacy Act ([CCPA](#)). Both measures give people the right to access the personal data that companies collect about them, which means organizations must be able to find all of the records on an individual in their systems without missing any because of inaccurate or inconsistent data.

Fixing data quality issues

Data quality managers, analysts and engineers are primarily responsible for fixing data errors and other data quality problems in organizations. They're collectively tasked with finding and cleansing bad data in databases and other data repositories, often with assistance and support from other data management professionals, particularly data stewards and data governance program managers.

However, it's also a common practice to involve business users, [data scientists](#) and other analysts in the data quality process to help reduce the number of data quality issues created in systems. Business participation can be achieved partly through data governance programs and interactions with data stewards, who frequently come from business units. In addition, though, many companies run training programs on data quality best practices for end users. A common mantra among data managers is that everyone in an organization is responsible for data quality.

Data quality vs. data integrity

Data quality and [data integrity](#) are sometimes referred to interchangeably; alternatively, some people treat data integrity as a facet of data accuracy in the data quality process. More generally, though, data integrity is seen as a broader concept that combines data quality, data governance and data protection mechanisms to address data accuracy, consistency and security as a whole.

[Margaret Rouse](#) asks:

In what ways do business users work with data quality tools in your organization?



Join the Discussion

In that broader view, data integrity focuses on integrity from both logical and physical standpoints. Logical integrity includes data quality measures and database attributes such as referential integrity, which ensures that related data elements in different database tables are valid. Physical integrity involves access controls and other security measures designed to prevent data from being modified or corrupted by unauthorized users, as well as

h Accelerator

data management and why is it important?

This was last updated in November 2019

Continue Reading About data quality

- 5 things data managers can do to help boost data quality
- Tips on best practices for data quality improvement efforts
- More on why good data quality matters to businesses
- Key features and capabilities that data quality tools provide
- Common data quality issues and advice on how to resolve them

Related Terms

data stewardship

Data stewardship is the management and oversight of an organization's data assets to help provide business users with ... See complete definition ⓘ

What is data governance and why does it matter?

Data governance (DG) is the process of managing the availability, usability, integrity and security of the data in enterprise ... See complete definition ⓘ

Dig Deeper on Data quality management software

Data governance metrics: Data quality, data literacy and more

By: Craig Stedman

Align business and IT drivers through data quality best practices

By: Evelyn Bleed

operational efficiency

By: Margaret Rouse

GDPR mandates push data quality improvement into IT spotlight

By: Andy Hayler

h Accelerator

data management and why is it important?

 Join the conversation


 2 comments


Share your comment

☒ Send me notifications when other members comment.

Add My Comment

Oldest ▾





 Margaret Rouse

- 4 Nov 2005 3:04 AM

In what ways do business users work with data quality tools in your organization?

Reply



 anvithavja

- 2 Nov 2018 1:43 PM

The importance of quality data has been well highlighted. The industry is slowly moving towards standardizing data quality management as a service. Dattamza has just announced its new suite of products – Izana, that can establisg service based operationalization in organizations

Reply

-ADS BY GOOGLE



[BUSINESS ANALYTICS](#) [AWS](#) [CONTENT MANAGEMENT](#) [ORACLE](#) [SAP](#) [SQL SERVER](#)

Search[BusinessAnalytics](#)

U.K. agency uses Oracle analytics to ID annual savings

Mandated to reduce spending by a billion pounds over five years, a U.K. health agency that supports the national healthcare ...

Analytics critical to decisions about how to return to work

Data plays a critical role in the decisions organizations make as they bring employees back to work, for both increasing ...

[About Us](#) [Meet The Editors](#) [Contact Us](#) [Privacy Policy](#) [Advertisers](#) [Business Partners](#) [Media Kit](#) [Corporate Site](#)

[Contributors](#) [Reprints](#) [Answers](#) [Definitions](#) [E-Products](#) [Events](#) [Features](#)

[Guides](#) [Opinions](#) [Photo Stories](#) [Quizzes](#) [Tips](#) [Tutorials](#) [Videos](#)

All Rights Reserved,
Copyright 2005 - 2020, TechTarget

[Do Not Sell My Personal Info](#)