

Data Quality Assessment: A Methodology for Success

Data: The Good, the Bad and the Money



- 2 INTRODUCTION
- 3-4 PRE-ASSESSMENT PLANNING
- 4 CONDUCTING THE DATA ASSESSMENT
- 5 FINDINGS PRESENTATION
- 6 NEXT STEPS

DATA: THE GOOD, THE BAD AND THE MONEY

“The world’s most valuable resource is no longer oil, but data,” stated *The Economist*, a weekly publication that focuses on current affairs, international business, politics and technology. And that is not the only source that claims data is a goldmine—year after year, newspapers, business experts and economists follow the trends and benefits of quality data, singing its praises and what it can do for small and large businesses across the globe.

Data—from contact information to buying patterns, and everything in between and beyond—is the crux of a successful business, its marketing strategies, market trend analysis and customer behavior predictions. But quality is key. “In a digitally powered economy like ours, only those with the right form of data can successfully navigate the market, make future predictions, and adjust their business to fit market trends,” said *techjury*.

Data integrity issues make managing critical information assets extremely challenging. Not knowing what data in the expanding collection of information is correct erodes an organization’s confidence in using the data. Whole systems applications such as CRM, ERP, or BI go under-utilized because of the perception of defects. Even worse, bad data can be used to make poor business decisions.



AN INTRODUCTION TO DATA QUALITY ASSESSMENTS

One solution to solving data-related issues is to ensure that an organization has a good understanding of what those issues are, and how they affect their enterprise’s computing systems. In order to gain that understanding many organizations turn to data quality assessments as a way to get a “snapshot” of data issues and identify ways to begin correcting them.

A data quality assessment is the act of inspecting data, measuring the data defects, analyzing the cause and impact of those defects, and then reporting the results of the analysis to key stakeholders.

Data Quality Assessment Methodology is comprised of three stages:

1. Pre-Assessment Planning
2. Data Assessment
3. Assessment Findings Presentation

As described in this guide, each stage of the data quality assessment process includes a number of steps. Like any information technology project, a data quality assessment requires a certain amount of planning to ensure it is successful. It’s during this time that an organization establishes well-defined requirements for those involved in the assessment project. The next few pages include some critical requirements organizations often consider during pre-assessment planning.

PRE-ASSESSMENT PLANNING

Establishing the team

It's during pre-planning that general personnel and project participation decisions are made. Oftentimes, the scope of the project will begin to take shape so that key decision makers can determine who in the organization would best be able to supply the critical insights and access to the data being studied. This can include subject matter experts (SMEs) who are familiar with the data being assessed.

Typically, the assessment team is a virtual (ad hoc) one that is comprised of a project manager, an assessor, SMEs, IT system support, and an IT DBA or analyst who is responsible for maintaining the data.

Sampling the data

Whenever possible, a sample set of the problematic data can be shared with the assessment team before the formal assessment is to take place. This allows the team to perform an initial evaluation of data structure, field composition, and business rule structures to aid in the formal assessment. Reviewing sample data prior to the actual assessment allows the opportunity to uncover and identify data singularities and rules that need to be explored in depth during the assessment.

Diving into documentation

Because every data set has particular contexts it is important to review any documentation that may exist that pertains to the identified data sets. Anything of this nature that can be delivered to the assessment team members will speed the assessment process.

Identifying the assessment goals

Having an idea of how "success" of an assessment project would be defined is a good precursor to the actual assessment itself. Answering questions like these is a good place to start:

1. What are the business drivers behind the assessment? For example, is the organization seeking better data for decision support or preparing to build a new data warehouse?
2. What are the perceived data quality problems?
3. What assumptions or expectations does the organization have of the assessment?
4. Did a critical operational failure occur? If so, what were the conditions and symptoms?

Establishing the project requirements

Interviews of the key project stakeholders, consumers of the information, and custodians responsible for maintaining the information will provide great insight into the specific data sets to assess, business rules to measure against, and reports to generate.

Defining the Data Assessment Scope

Perhaps the most critical part of the assessment involves defining its scope—the amount and types of data to be assessed. The scope must be as narrow and limited as possible to deliver assessment feedback to the organization quickly. Multiple assessments can always be conducted if the organization's requirements span broad segments of the myriad data sets within the enterprise.

In defining the scope, the assessment team works to identify:

1. Specific databases, including location and platform.
2. Specific tables within the databases.
3. Specific columns within the tables.

Reviewing the Data Sample Set

To facilitate implementation of business rules and queries the assessment team performs a review and walk-through of sample data with the key stakeholders. Reviewing sample data prior to the on-site assessment allows the opportunity to uncover and identify data singularities and rules that need to be explored in depth during the assessment.

Reviewing the Data Documentation

This step includes the review of any:

- Data definitions
- Data standards
- Database schemas
- Written policies and procedures
- Written business rules and validation criteria

Once the scope is determined, and pre-planning is complete, a formal assessment can begin on a foundation of confidence!

CONDUCTING THE DATA ASSESSMENT

To begin the formal data assessment, it's important to set the context under which projects can be most successful. The person responsible for actually performing the physical assessment (noted as the "assessor" in the section discussing establishing the team earlier) will perform the following analyses and tests on the data being studied:

- Data Profile—including frequency distributions, uniqueness and completeness tests
- Column Analysis—single column validity tests
- Referential Integrity —tests for primary and foreign key integrity
- Business Rules Compliance —cross column, cross table, cross database analysis



Also, depending on the project's scope, the assessor may evaluate street address quality and even the level of redundancy of records found in the data set.

During the assessment itself, the assessor will generally utilize an assessment tool to aid in automating the process, and establish a baseline that can be repeated at regular intervals to understand progress made on quality initiatives at later dates.

During the use of such a tool, the column validation is performed and business rules are assessed as defined during the requirements phase. The results of the queries will be saved in a metadata repository for use in preparation of the findings presentation.

The actual time it takes to conduct an assessment is highly dependent on the amount and scope of data to be evaluated, along with the number and complexity of business rules to be matched against the data. The process involves building a data profile of the required tables and columns, checking for referential integrity, and then building each of the required business rules. The assessment continues until each validation and business rule has been compared against the data and the results stored in the metadata repository. During this period access to SMEs is key to tuning the business rules and reviewing intermediate findings.

FINDINGS PRESENTATION

Once all tests and queries have been conducted, the assessor will compile the results and ask for a preliminary review by the SMEs. Once the feedback from the SMEs has been collected, the final findings will be prepared. Final findings are typically reported via written report and verbal presentation as required.

In general terms, upon completion of the data quality assessment the project manager presents the findings to the stakeholders in the organization. Depending on the requirements and the assessment's outcome, the presentation will include topics such as:

- Examples of specific data defects
- Anecdotes of the impact defects have on the organization's day-to-day business
- Reports of domain measurements and analysis tests
- A view of changing metrics over time

- Recommendations for both data capture and process improvements
- Suggestions regarding how to clean existing data elements

Often organizations realize that the findings from the assessment generate new questions about the data. By the very nature of keeping scope size to a minimum, additional data elements are often identified for investigation during final findings. This is why it is so important to make data quality assessment practices repeatable and extensible to other data.

Additionally, if not part of the initial requirements, the assessor can offer to establish an environment where the business rule tests are run periodically so the organization may perform trend analysis over time.

It is important to note that a data quality assessment is not a one-time activity. Data ages and changes faster than most organizations are aware, and a continual monitoring and reporting process on key data elements is essential for optimal operations management.



NEXT STEPS: SOLUTIONS TO LEVERAGE THE POWER OF DATA

A data quality assessment is just one piece of building a comprehensive information quality strategy for an organization. It uncovers hidden and unobvious problems early in the project lifecycle when projects can be affected without fear of complete project failure. It quantifies the number and type of defects found, along with identifying the affected records. This is critical to know where the highest priority data quality problems exist so that an organization can implement the appropriate processes and technology solutions to leverage the power of its information assets.

Melissa offers a full spectrum of solutions that not only profile your data to identify problem areas but also:

- Cleanse and update records—verify identity, name, address, email and phone in real time, or batch update inaccurate or outdated information. Parse and structure data into a usable format.

- Enrich with demographics, firmographics, geographics, social media, property attributes, and missing email and phone information to support analytics, personalization and omnichannel marketing efforts.
- Match and merge duplicate records to create a single, accurate customer profile you can trust.
- Monitor your data across the entire data lifecycle to prevent bad data from entering your database and keep it clean over time.

An accurate and organized database is vital for analytics, insights, and better business practices that will help grow your company to its desired level of success. To quote an article in Forbes, “We know one thing is proven: the benefits of leveraging Big Data will outweigh IT investment.”





www.melissa.com

About Melissa

Our 35 years of address expertise started with ZIP+4 and turned into so much more. Melissa is a single-source vendor of global address management, data quality and identity verification solutions that help organizations harness accurate data for a more compelling customer view. Our industry-leading solutions have processed over 1 trillion address, name, phone and email records, making it clear why thousands of businesses worldwide trust Melissa with their data quality needs. For more information, visit www.melissa.com or call 1-800-Melissa.

US

**22382 Avenida Empresa
Rancho Santa Margarita, CA 92688-2112**

800 MELISSA (635 4772)

GERMANY

+49 (0) 221 97 58 92 40

INDIA

+91 (0)80 4854 0142

AUSTRALIA

+61 02 8091 6000

CANADA

800 635 4772

UK

+44 (0)20 7718 0070

SINGAPORE

+65 8 2997442