WHITE PAPER

# Improve Self-Service Analytics Credibility with Data Quality

PYRAMID ANALYTICS

# Contents

# Introduction

### Data is at the heart of every organization

In the journey to become more data-driven in decision making, we are seeing unprecedented democratization of data and adoption of self-service analytics. Rigid data collection and reporting processes of the past have given way to rapid gathering of raw, unstructured and crowdsourced data. As a result of that change, there are inevitable trade-offs with data quality.

Self-service visual analytics solutions often quickly expose data quality issues that you may not even realize exist. Unfortunately, inaccurate data undermines the powerful value of self-service analytics. If people don't trust your reports, they won't use them. Since self-service analytics credibility, adoption and success hinges on accurate data, data quality should be given more attention as you implement these solutions.

### When data quality becomes mission critical

Furthermore, as organizations advance in analytics maturity by adding predictive and prescriptive algorithms to automate decisions, data quality becomes more mission-critical. While most organizations indicate data supports business objectives, according to an Experian Data Quality Survey in 2016 [1] an average of one third believe their data to be inaccurate or incomplete, undermining their ability to automate decisions. In this white paper, we will explore common data quality issues and provide practical recommendations to help your organization enhance data quality and improve confidence in self-service analytics.

> "Inaccurate data undermines the powerful value of self-service analytics.

[1] Experian https://www.edq.com/white-papers/2016-global-data-management-benchmark-report/

# Common Data Quality Issues

## Human error still #1

System migration efforts, application changes and errors frequently are the root cause of data quality problems. Yet the top challenge for data quality is still the age old problem of human data entry error. When most of the data originates from within the organization or is acquired via a controlled method, data quality tools, cleansing and correction processes can be applied to address those issues. Ideally data entry and system errors get resolved at the source with validation or added references to master data.

Newer sources of data such as social media are highly prone to human error. When data originates outside the organization, data quality becomes a bit more challenging. Deciding how to decipher or standardize values, maintaining consistent values across sources for traceability, storing original and assumed values, and applying scalable techniques to larger volumes of incoming data in varied formats, generally need to be discussed and decided upon by the consumers of that information.

## Handling data changes for usability

For self-service analytics, non-technical users will often need to integrate and enhance data from a wide variety of internal and external sources. To relate disparate data sources, these users will need to make changes, add identifiers or combine fields. As changes are made, different copies of the data might exist if no governance for those use cases is in place. Ideally, governance and data quality processes will increase data usability for popular data blending scenarios while maintaining a logical repository of personalized data changes, meaning and context.

## Getting users to care

Let's face it. Data quality is not an exhilarating topic for most business users until it embarrasses or prevents them from accurately reporting or making a decision. Thus you will need to creatively find ways to encourage the business to care. Data catalogs, logical data warehousing, personalized data prep combined with data quality and cleansing service technologies can all be sprinkled into your overall self-service BI strategy. These related solutions should be appealing to reporting users. Helping users find available data for reporting is far more exciting than approaching them with the dry data quality angle. In the recommendations section, we will share a few tips and tools to help you with this universal human challenge.

# Recommendations For Improving Data Quality

You can't improve what you don't measure. Ultimately improving data quality starts at the source. Laissez-faire attitudes toward the quality of data won't change if no one is held accountable or measured on it. Success often means collecting data quality metrics, establishing baselines, goals, and holding new areas of an organization accountable for their data – not just IT or the data warehouse team.

**Unaware**

- No understanding of data quality impact



**Reactive**

- No data quality roles
- Tactical data correction
- Limited tools

**Proactive**

- Sponsors
- Data quality roles
- Data quality metrics
- Tools and processes for data quality issue discovery and resolution

**Optimal**

- Chief Data Officer
- Data quality roles
- Data quality metrics
- Platform approach combining monitoring, tools and processes for data quality issue discovery and resolution

## Data quality metrics

What types of metrics should be collected? Data quality covers a wide gamut of approaches and solutions ranging from basic data cleansing and profiling to services for data validation, standardization and monitoring for data policy compliance. Traditional methods to measure data quality assign the degree to which data is "fit for purpose" using metrics relative to what are referred to as dimensions of data quality. The six core dimensions of data quality are:

1 **Completeness:** Specifies the data elements that must have values.

2 **Uniqueness:** Represented entities once and only once within the data set.

3 **Validity:** Range of valid values.

4 **Accuracy:** The degree to which the data values are correct.

5 **Consistency:** Related data values across different instances, or known reference domain.

6 **Currency:** Maintaining values that are up to date.

For metrics to be meaningful to groups that own the data source such as sales or marketing, area subject matter experts should define additional goals and measures to motivate their teams.

> The most common objection to overcome is our data seems "good enough."

## Building a case for data quality tools

If businesses are going to increase confidence in data-driven decision making, they should budget for enterprise-wide data quality programs and tools. To ensure that business leaders see data quality as an integral component of an ongoing data management program, it's likely that stakeholders will need to build a business case for it.

How can you estimate the cost of data quality issues? Honestly, it is not easy. Although leaders may agree there are data problems, they usually don't understand the impact data quality has on the bottom line. The most common objection to overcome is our data seems "good enough."

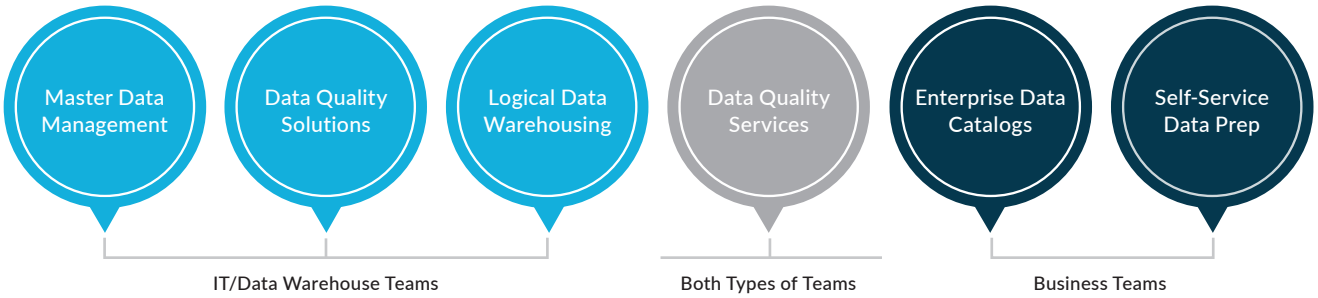
If you need to gain sponsorship for improving data quality, look specifically at how low quality data hurt operations. Were there delays in decision making? Did incorrect data reduce process efficiency such as contacting customers, manually re-entering corrected information or experiencing returned shipments? Were there any critical missed deadlines due to data issues or compliance penalty fees? Do project or finance teams spend evenings and weekends scrambling to clean up data in Excel spreadsheets for recurring reports? Usually reporting data preparation accounts for 80% of the work effort. It is an expensive problem to ignore.

The key to building a case for data quality is to link it directly to business problems departmental leaders in an organization are held accountable for today. You can start by mapping out how data quality processes link to the business problems. Quantify the time, amount of people or processes affected, estimates costs and other less tangible qualities. When putting your business case together, be sure to provide proposed solutions with a cost-benefit analysis. Leaders want you to suggest solutions - not just bring them problems.

# Data Quality Solutions

Large companies often combine enterprise master data management (MDM), a proven technology that has been too expensive for most organizations to afford, along with niche data quality services and tools. Mid-market groups simply don't have the time, budget or resources for those approaches. Recently several business-user driven data quality solutions have been entered the market. New enterprise data catalogs and self-service data preparation offerings align to service both data quality and self-service analytics needs.

### Spectrum of Data Quality




The following is a brief listing of popular solutions that can be combined with Pyramid Analytics to enhance data quality for self-service analytics.

- **Claritas Connect** offers profiles of consumer and business markets.
- **comScore Networks** provides comprehensive information on Internet users, including buying behavior.
- **Melissa Data** offers data validation, cleansing and data enrichment services.
- **Dunn & Bradstreet** master data services for universal business identification
- **Equifax** credit information services, payment services, software, modeling, analytics, consulting and direct-to-consumer services.
- **Experian QAS** leading provider of verification solutions for names, addresses, emails and phone numbers.
- **Experian Pandora** leading data quality services provider for business users to conduct profiling and relationship discovery.
- **Syncsort Trillium Software** offers data quality tools and professional expertise to help you deliver accurate, timely data.
- **Open Studio for Data Quality** is an open source data quality tool.
- **Perspective ILM Data Governor** provides business users functionality to quickly create and automate execution of reconciliation rules between source and target.
- **Profisee Maestro Platform** is a leading Microsoft SQL Server ecosystem Master Data Management solution.
- **Azure Data Catalog** is a cloud enterprise data catalog used with self-service analytics tools to discover data sources and a crowdsource metadata management.
- **Alation** is the industry leading self-service enterprise data catalog powered by machine learning to continually improve data quality.

## Getting proactive with data quality

Once you gain sponsorship for a data quality program, dedicated roles and/or tools. You will want to empower business users and provide guidelines. Much like self-service BI governance, data quality is an ongoing, continual improvement initiative. You will want to actively monitor and reconcile high priority data sets to validate their compliance with defined data quality expectations. Many solutions provide basic alerting for assigned data stewards or trigger prescriptive actions when data quality thresholds are not met.

Modern self-service BI and data catalog solutions provide a collaborative approach for sharing data source metadata to help ensure consistency for common reference domains. They also provide an environment for sharing business-friendly terminology and contextual semantics. Ensuring consistent reference data for use in reporting can help fast-track general metadata standards. While metadata standards may be assumed when working with structured sets, metadata definitions are becoming more important when dealing with unstructured content such as images, videos, audio artifacts, or content from logs and text documents.

## Data quality culture

Despite being tied to clear business objectives, most current approaches to data quality in terms of people, processes, and technology are lagging in comparison to analytics initiatives. As of 2016, only 1 in 5 companies is operating at the most sophisticated level of data quality management that includes have a data quality role and data quality executive sponsorship.  Clearly there is much room for improvement.

In all aspects of data management, and especially in reporting to outside entities, there is an expectation for data quality. Deliberate negligence cannot be used as a foundation for plausible denial. Data management practices should be proactive.

## Top five tips

To get started with data quality and overcome the human, cultural obstacles to data quality that are holding you back, here are our top five tips.

1. Start small with a small pilot project on high value data. Involve key leaders, data subject matter experts and IT in a kick-off session to review how each party uses the data to be improved. This will help reduce resistance to participation.

2. Data quality is everyone's responsibility...not just IT or the data warehouse team. It should be a partnership between business and IT. All persons and systems that capture or use data should be held accountable for data quality much like data privacy or compliance initiatives. Finance, marketing, sales and other departments should share responsibility to own the quality of their own data sources.

3. Consider assigning departmental and IT data steward subject matter experts that can review records that exceed data quality thresholds, identify root causes of data quality issues, and implement approaches to correct or improve those issues.

4. Always try to improve data quality at the source if you can with validation, master data reference services and user interface aids in applications. Human error entering data into free-form text fields is still the most common source of data quality issues.

5. Consider adding data audit metrics to measure consistency and accuracy in operational health reporting. Leaders at the C-level want to see clear metrics relating your data quality program to specific priorities for the business. Examples of priorities include operational performance, financial performance, customer satisfaction, and regulatory compliance.

> "As of 2016, only 1 in 5 companies is operating at the most sophisticated level of data quality management.

# Conclusion

Every organization today depends on data to understand its customers and employees, design new products, reach target markets, and plan for the future. Accurate, complete, and up-to-date information is essential if you want to optimize your decision making, avoid constantly playing catch-up and maintain your competitive advantage. For more detailed information on improving self-service analytics governance and data quality, here are several other excellent resources.

- Delivering Governed Self-Service BI Across the Enterprise
- Experian Data Quality Library
- Data Quality Pro community
- The Practitioner's Guide to Data Quality Improvement by David Loshin
- Data Quality: The Field Guide by Thomas Redman PhD

PYRAMID ANALYTICS

**CONTACT US**

+ 1 800 385 6704

www.pyramidanalytics.com

info@pyramidanalytics.com