

The importance of data quality: A sustainable approach

By Carol Newcomb, Senior Data Management Consultant at SAS

Nobody likes unpleasant surprises. When it's time to look objectively at what has been happening in terms of customer activity, business productivity or progress toward targets, everyone wants to be able to trust the reports they're given. And no one wants to be embarrassed by delivering inaccurate reports, no matter what the underlying reason. In situations like this, the importance of data quality is undisputed.

But how much control does an individual have over the quality of data used in reports? Who is accountable for that data? Who understands where it originated, or how and why it may have been altered? Who gets to write the business rules or the quality standards? There are usually many different business constituents for each set of data, each needing to tell their own story. Do you draw straws to decide who gets to tailor the data to meet their specific business needs? And who gets to say whether the data is wrong?



Of course, there are software tools that can help with data correction and error analysis. But tools alone won't fix the problem. Business users first need to have a plan to help them identify quality issues, track down underlying sources, develop mechanisms to resolve problems, and then set up a process to monitor and flag any new issues that arise.

Analysts spend from 20 percent to 60 percent of their valuable time trying to understand and fix poor data. Any ROI analysis will take into account the squandered time among analysts across the enterprise who are caught up in these unproductive but necessary activities.

A sustainable plan

Managing data quality is not necessarily simple. When you consider the data life cycle – from data creation/collection to archival – there are many steps along the way, including:

- Rules for collecting/creating data.
- Data quality standards, thresholds and rejection criteria.
- Data standardization and summarization rules.
- Data integration rules with other sources of data.
- Hierarchy management (relationship management).
- Ongoing triggers to detect outliers during updates.
- Data correction rules.
- An effective, sustainable data quality plan will resonate with business users and should include the following five elements.
- Elevate the visibility and importance of data quality

Poor data quality has a significant business cost – in time, effort and accuracy. Quantify the cost of poor data and build a credible business case that demonstrates the negative impact of current data quality problems. Illustrate how data quality affects different parts of the business. This becomes a key part of your justification for why a plan that comprehensively encompasses the importance of data quality is a business imperative.

Formalize decision making through a data governance program

Data correction should not happen in a vacuum, nor should each analyst have his/her own rules for correcting errors. Avoid allowing too many people to make one-off data quality decisions that don't meet a shared business purpose. Grant authority for developing business rules and standards with a decision-making data governance group that has perspective across business areas. These rules need to be vetted and approved to ensure they're valid and reusable. Only then should the data quality process be applied.

Document the data quality issues, business rules, standards and policies for data correction

A boss once told me: "If it's not documented, it didn't happen." To battle a fire-fighting culture of data fixes – and to prevent ongoing inefficiencies caused by individuals who correct data inconsistently – you'll need to document each issue, publish it and communicate the remedy. This encourages users to avoid costly and time-consuming adventures in fixing data in ways that can't be reused or shared across the organization.

Clarify accountability for data quality

Develop a process whereby business users can report data quality issues and then work with data stewards to research the error's source and develop a resolution. Relieve business analysts from the burden of researching data quality issues – free them to do their jobs as analysts. Identify data quality specialists, both data stewards and data quality professionals, who are responsible for resolving data quality issues. Those issues can range from root-cause analysis, metadata management and policy definition to documentation and monitoring. This approach – which acknowledges the tremendous importance of data quality – is a huge savings to most organizations.

Applaud your successes

If you've crafted your plan carefully, you'll collect baseline statistics and then measure improvements to data quality over time. Demonstrate the business value to users through your own case studies. The importance of data quality is supreme – better data translates directly into better business value. Be sure you understand how those in different parts of the enterprise measure themselves, and tie the improvements in data quality to improvements in their overall success. Then communicate how to share the value of better data across business areas.

Recognizing the multifaceted nature of questions raised by people with a variety of business interests and perspectives will position you to design a sustainable approach to data quality. As you flesh out details in the data quality plan, include technical experts to work with your business analysts. Together, they can account for the data's full spectrum of definitions and characteristics while engineering ways to fix and maintain it. The investment will pay off in the short run and for years to come.



Carol Newcomb is a Data Management Consultant for SAS with more than 25 years' experience in information management. She specializes in the design and implementation of data governance programs and data strategy for a broad range of industries. She is author of the SAS e-book *When Bad Data Happens to Good Companies* and has written numerous blogs and white papers, including *Implementing Data Governance in Complex Health Care Organizations*.