



BUSINESS INTELLIGENCE

Data Quality Risk Mitigation

Do business. **We're IT.**

COMSYS recently acquired *Praeos Technologies Inc.*, a leading Atlanta-based provider of IT consulting services specializing in the business intelligence, data warehousing and analytics sectors.

Praeos' strong practice in business intelligence and data warehousing have become the platform for these services offerings within the COMSYS Managed Solutions Group. Together, these combined business intelligence practices, now called COMSYS BI Practice make COMSYS a leading provider in these sectors.

INTRODUCTION

By its very nature, data quality is a variable and often unknown quantity. Why? The data quality issues inherent in *yesterday's* data sets do not equate to quality issues that will arise in *tomorrow's* data sets. To successfully execute on your organization's Enterprise Business Intelligence (EBI) strategy, you must acknowledge these issues and prepare your team to deal with them. This "fact" of EBI life demands a straightforward approach for success, and this white paper discusses how risk mitigating techniques addressing data quality can be applied throughout the phases of a successful methodology.

EBI INITIATION PHASE

When solving any issues, the first step is "*admitting you have problem*". The first step in data quality risk mitigation is accepting that "you don't know what you don't know." Just a few of the unknowns: diminished institutional knowledge of data over time, inadequate controls in source systems, and unknown points of manual intervention. While one risk mitigation is to quantify as many of these unknowns as possible (e.g., data quality analysis), it is critical your organization understands that "perfect knowledge" is unrealistic. It may be helpful to highlight relevant examples to illustrate this point to promote acceptance.

Identify and Educate Stakeholders on Data Quality

Creating formal stakeholders is very important to the success of the entire EBI initiative. Formal stakeholders include:

- **Cross-line Data Stewards:** responsible for in-depth knowledge of organizational data and sensitive to confidential information. Use data knowledge to identify, quantify, and help resolve the impacts of data quality issues.

- **Data Quality Committee:** acts as a steering committee for data quality issues. Typically comprised of Cross-line Data Stewards.
- **Executive Sponsorship:** ensures that the right issues are routed to the right person for resolution. Interfaces with executive leaders at a high level.
- **Internal Audit:** assists with data quality issues that require the attention of External Audit.

Data quality is an ever-changing factor; issues will arise, often times late in the project lifecycle. The formal stakeholders must understand the importance of working together in partnership to convey realistic EBI solution expectations to the informal stakeholders and executive leadership, regularly and consistently, throughout every phase of the project.

Formal stakeholders should be prepared to deal with the potential for sensitivity around data quality issues with cross-enterprise impact. For example, issues that precipitate a need to restate corporate earnings would be highly sensitive and carry a wide organizational impact.

Atmosphere of Collaboration

In addition, the formal stakeholders should understand the appropriate escalation path for each data quality issue. This includes a clear plan with defined expectations, roles and responsibilities for the data users, informal stakeholders, formal stakeholders, and Executive Sponsors.

Plan for Data Quality Issues

As the team begins defining the project objectives and risks, data quality must be addressed. The formal stakeholders need to define data quality expectations so the IT team can define the tasks to address them both during project development and post project operations.

ANALYSIS PHASE

A risk mitigation strategy begins with stakeholder education and continues through all phases of the EBI initiative lifecycle. Planning for the activities and phases of an EBI initiative takes place during the analysis phase. The team can mitigate risk around data quality by planning for data quality analysis activities or mini-phases with separate deliverables.

Estimate for the Unknown

The team should use slack and conservatism in design/build/test task estimation to address data quality unknowns and mitigate risk. Typically, EBI initiative estimates should include 25% extra time across the initiative per design/build/test task. This allows for unknowns to arise, escalate and be resolved within an appropriate timeframe.

Perform Data Quality Analysis

Often times, EBI initiatives omit Data Quality Analysis activities (e.g., defining detailed representation of how data subjects are represented in source data inventory, or documenting critical data relationships). Without these activities, appropriate measures to find and fix the data quality issues cannot be taken. By defining these activities as true data quality analysis activities with separate deliverables, more issues are found and given an appropriate timeframe for resolution or acceptable work arounds. When data quality activities are not performed, it is only a question of when the data quality issues will become self-evident. The later in the project, the greater the impact will be to schedule, cost and stakeholder confidence.

Incorporate Project-specific Stakeholders

The Data Analyst should be added to the team of formal stakeholders during the analysis phase. This person is critical to the separate deliverables needed for data quality analysis activities. They should possess a solid knowledge of an RDBMS platform where they build SQL scripts for data profiling and a working knowledge of a data profiling tool. The Data Analyst will work closely with the initial stakeholders and provide feedback to the Data Quality Committee at critical project milestones.

Don't Confuse Analysis/Design Overlap with Completion of Analysis

Stakeholders often face organizational pressure to begin Design before Data Quality Analysis has completed. As the Analysis phase progresses, additional DQA becomes less likely to drive significant data structure changes. The sizing and hardware procurement needs of a new project may create additional pressure to move into Design. However, these pressures should not be misconstrued as decision points for the completion of DQA.

While the exact completion point of DQA can be imprecise, it should only be determined by the progress of learning about the data—how critical data subjects and relationships are represented in the source systems throughout historical feeds—not on the progress of other analysis-phase activities.

DESIGN PHASE

Continue Data Quality Analysis

A risk mitigation strategy continues into the Design phase of an EBI initiative. The strategy evolves from planning into change request tasks, also known as “scope creep.” This phase leverages the partnership of informal stakeholders who may provide quick data quality issue resolutions.

Consider Change Controls

Consider a well-formulated change control process that allows new data quality issues to be addressed in this phase while keeping all formal and informal stakeholders satisfied with the forward movement of the EBI initiative.

Incorporate Informal Stakeholders

Informal stakeholders should be added to the partnership environment (e.g. source owners, SMEs, and stewards). They need to understand that the project follows an approved plan that was established in the previous phase. In addition, they should be informed about the inherent data quality risks expected during an EBI initiative in the Design, Build and Test phases.

Don't Communicate Overconfidence

Perhaps you have yet to discover a large data quality issue at this point in your project, but the possibility remains. Do not become overconfident, and do not give in to the temptation to move into build early. Remember, your DQA activities were necessarily focused on a subset of the data in history. Instead, review the task estimates and the activities/phases of the EBI initiative in further detail. There is a definite point of diminishing returns with regard to data quality tasks and activities/phases - but there is also much more to learn and prepare in this phase.

BUILD PHASE

Keep the Slack

A risk mitigation strategy continues into the Build phase, where risk mitigation focuses on maintaining sponsorship confidence while addressing outstanding data quality issues through aggressive prioritization. Re-planning Build, Test, and Deploy activities after the completion of Design should continue to incorporate slack to account for unknowns in the data.

Keep the Faith – Collaborate and Review Successes

Sometimes the new data quality issues in the Build phase make it difficult to see the light at the end of the tunnel. The separate data quality documentation that was created during the Analysis phase should not only prove just how many successes have happened throughout the initiative, but also keep morale high and the sponsorship actively engaged.

An aggressive prioritization schedule that aligns with change control procedures can overcome any remaining data quality issues that arise.

CONCLUSION

Data is constantly changing. There will always be data quality issues with any EBI initiative. But a proper risk mitigation plan can:

- set realistic expectations for the project
- provide a meaningful stakeholder team structure through each phase of the project
- formalize an escalation plan to resolve data quality issues as they arise
- maintain a sense of continuity for the lifecycle of the initiative