

Data Integration and Migration

Warning: Lack of Profiling Can Doom a Project

Scope creep is a constant threat in data integration and migration projects. Are you already at risk for running over budget and off schedule with a project? Here are some of the warning signs:

- The business users and the IT users are not communicating.
- You do not know whether critical fields are actually populated with the right information.
- There is no way to confirm if integration designs accomplish business requirements for all data conditions.
- It is unclear whether the coding conventions for products/parts are consistently the same in different data sets.

Top Five Reasons Why Integrations Fail

More than 80% of all data integration projects fail according to several surveys over the past few years. Experts agree that the biggest reason for this high failure rate is that the integration team did not understand the extent and nature of problems buried in the data.

Scoping underestimates data problems—During initial project planning and scoping, data problems are not yet identified. Addressing these unplanned data issues easily consumes as much as 70% of the time and budget in rework later on, resulting in costly delays.

Insufficient business user input—Business users understand the meaning, importance, and relevance of specific fields and values, yet participate very little after sign-off on metadata and data models until User Acceptance Testing, where more problems are identified.

Expect too much from ETL tools—ETL tools are optimized to move and transform records, but not identify data inconsistencies and errors. Likewise, they are not intended to interpret the meaning of data or perform logic-based matching for the purposes of identifying duplicates or householding individuals and corporations.

Requirements change—An integration strategy must take into account ongoing changes. Prior to Phase 1 project completion, the user community will have change requests both in terms of data and functional requirements. Rapidly accommodating data changes ensures that users receive the information they need.



Benefits of High Quality Data

Lowers integration costs
Lowers project risks
Increases user acceptance
Meets requirements

How To Avoid Integration Problems

Project managers can keep their integration projects on time and within budget by taking a few steps early on in the process:

Build a cross-functional team—A cross-organizational, cross-functional core team should include IT resources, business users, subject matter experts, business analysts, and an executive sponsor.

Risk assessment—Create accurate estimates for how long it will take to prepare the data for integration, and what must occur during integration, based on a full assessment of the actual data sets.

Profile upfront—Early profiling identifies data problems prior to design and development. This helps the team identify exactly what data can be integrated and how to best accomplish this.

Plan for change—Expect a constant stream of new requests from users for additional data—from new sources as well as current ones. Help manage change requests through a simple assessment of the data itself. It may be quite simple to include a few extra, valid fields.

Data Integration Innovations

Data integration projects have become much more involved than one-time migrations or weekly batch updates. Innovations in technology have opened up many possibilities for meeting the integration needs of complex, distributed organizations.

Integration options include replication and synchronization techniques, traditional ETL options, and other methods for sharing data in real-time—virtual integration, application integration, and service-oriented architectures that leverage standardized services.

Data integration initiatives are quickly becoming the thread tying together complex organizations and making real time information sharing possible.

Quality Data, Despite Integration Technology

Regardless of the many possible technical solutions available, the data itself remains the most significant and complex obstacle to address as part of any integration initiative. Delivering high quality data to the business remains the ultimate goal for integration projects. To this end, companies are seeking out a data quality solution that interoperates with all types of integration technologies to meet project-specific requirements as well as enterprise-level requirements.

Trends Supporting Enterprise Integration

Service Oriented Architecture (SOA)

Companies are using SOA's so that they can define and develop a process once, then leverage that work across any of the systems and business processes that require it, reducing development time and ongoing maintenance while ensuring consistency across systems and functions.

Real-time processing

Organizations need instant access to complete information in order to deliver high levels of customer satisfaction and ensure the operational efficiency necessary to remain competitive. Optimal real-time capabilities consistently deliver high quality data with high performance response rates, regardless of volumes.

Master data management (MDM)

Organizations are taking on enterprise-wide initiatives to manage a business's core information about suppliers, customers, products, inventory, and other basic information, to be used repeatedly in transactional systems. EIM (Enterprise Information Management) and CDI (Customer Data Integration) are two other popular concepts in data integration that also address data consolidations that span systems and business units.

Corporate data standards

Master data, compliance, and data governance all represent huge, enterprise-wide data initiatives. Just as Rome wasn't built in a day, companies realize that achieving these goals takes a long time and must be based on the careful coordination of

multiple, consistent integration efforts. Portable, reusable data quality rules and standards are the best way to cost-efficiently ensure corporate standards across multiple efforts over time. They allow organizations to economically and efficiently use rules developed in one project for another, thereby promoting consistent standards and levels of quality across all records.

Consolidating instances for an enterprise view

Companies are streamlining their IT systems to consolidate multiple instances of ERP systems or other enterprise applications into one system, for a single view. Consolidations also build master data.

Monitoring data for changes

Companies recognize that data integration is not a one-time effort, but a long-term plan for maintaining data integrity and responding to changes. These companies are investing in data quality solutions that also allow them to monitor incoming data in real time, at the source systems, and correct it before it gets stored and replicated in enterprise systems, as well as highlight anomalies that do exist.

Unicode

Companies are adopting Unicode-enabled technologies so that they can quickly and easily handle data from around the world.

Trillium Software Solution for Data Integration

The Trillium Software System provides a complete data quality solution for data integration, across varying integration technologies. It streamlines the process of assessing, correcting, consolidating, and monitoring data.

Data Profiling and Risk Assessment

TS Discovery provides a non-technical team a way in which the data itself can be easily assessed so realistic timelines are set, based on known issues within the data. It goes beyond traditional profiling with easy-to-use workspace and tools that bring business users into the data assessment process. IT and business users have the same view of the data.

Business user-friendly

TS Discovery's intuitive graphical user interface makes it easy for business users to look at the source data without knowing SQL. Through point and click methods, users interact with the data, isolated from negatively impacting source system performance.

Collaborative workspace

Core integration team members from both business and IT can assess the data, flag specific records or fields for further investigation, share notes, and determine comprehensive integration designs that attend to all complexities of the data. A common view simplifies communicating about what is needed, what is not and what the right formats and structure must be.

All data profiled

Automated profiling of all the data makes no assumptions about what the content and structure of the data sets are. Instead it systematically checks for anomalies, outliers, misspellings, dependencies, and possible keys for all the records. It delivers a full report with easy drill-down capabilities.

Correct and reformat data

TS Discovery corrects and standardizes name and address data. Users can reformat columns of information, populate new columns based on business rules, and write rules to correct and standardize any type of data—all directly within the profiling workspace.

Export tables

TS Discovery offers the unique ability to build new tables for export to TS Quality and other applications, including columns created by correcting and reformatting data.

Data Quality with Interpretation

TS Quality is a rules-based data quality engine that provides everything you need to build data quality processes during data migration, and on a continuous basis to maintain the quality of the data in real time as well as high-performance batch.

TS Quality integrates data across platforms, systems, and geographical boundaries. It automatically consolidates data from disparate systems as well as from different countries around the world.

Reusable quality rules

Portable rules-based processes created in TS Quality ensure consistent results across platforms and save development time and money, while promoting corporate quality standards. Rules are exportable in many formats, including XML, for easy integration into other types of applications.

Semantic interpretation

TS Quality interprets data in context, identifying errors in meaning, misfielded data, and unstructured text data and then correcting and standardizing them accordingly.

Deployment options

Universal interoperability and rapid integration are the hallmarks of TS Quality. Ready-made connectors reduce development costs for integration projects.

Monitor

Use the Trillium Software System for periodically assessing the quality of existing data, measuring the quality of newly acquired data, and repeatedly correcting and standardizing new and changed data.

Data alerts

Monitor source data to detect anomalies and identify specific events, receiving alerts when such conditions arise.

Maintain quality over time

Periodically assess existing data. Correct and standardize new data before it is stored in enterprise systems.

Benefits of Data Quality for Data Integration

Lower integration costs—With the help of extensive profiling, the UK Ministry of Defence efficiently combined records from 860 systems and millions of inventory items, saving \$36M in supply chain management processes in the first two years.

Increase user acceptance rate—Using the Trillium Software System to correct and standardize the data in its little-used data warehouse, a conglomerate increased use of their marketing and analytic systems dramatically in the first year.

True data support for SOA—An international online employment agency calls data quality functionality through a Web Service to correct and standardize data in real time for its Siebel system, list cleansing efforts, and single view creation to promote on-demand integration across systems.

Foundation for MDM/CDI—A heavy manufacturer uses the Trillium Software System to build master data for its SAP Business Warehouse. In three months, it had a harmonized view of 70 databases, including a single view of the supply chain and over 2.8 million materials and parts stored in five languages, working in parallel with an ETL process.

Improve communication—Business users who can easily view source data in a workspace they share with IT can more easily define their business needs in data terms and better understand the issues involved in the integration process.

Putting It All Together

Integration has no end in organizations. More and more data flows in continuously. The data itself changes, and business users change and expand on the information they require. Systems offer new ways of getting more information and new insights from the data. By using the Trillium Software System, companies successfully complete today's data integration while building a foundation for tomorrow's integration solutions.

Harte Hanks Trillium Software
www.trilliumsoftware.com

Corporate Headquarters
+ 1 (978) 436-8900
trlinfo@trilliumsoftware.com

European Headquarters
+44 (0)118 940 7600
trillium.uk@trilliumsoftware.com



**TRILLIUM
SOFTWARE**