

# Trillium Software System<sup>®</sup>, Enterprise Services Edition

## Product Overview

**Harte-Hanks Trillium Software**

[www.trilliumsoftware.com](http://www.trilliumsoftware.com)

**Corporate Headquarters**

+ 1 (978) 436-8900

[trilliumsoftware@trilliumsoftware.com](mailto:trilliumsoftware@trilliumsoftware.com)



**TRILLIUM  
SOFTWARE**

## The Trillium Software System®

The Trillium Software System delivers complete end-to-end data quality life cycle management with a completely integrated data quality suite. It is the only platform on the market specifically architected to profile, assess, validate, improve, and monitor global data quality across and throughout a distributed enterprise using a single user interface. Designed for collaboration and information sharing, the Trillium Software System lets businesses individually define what data quality means to their organizations and use one product suite to implement and govern data quality standards across departments, lines of business, technology initiatives, and distributed systems. Executive management, business analysts, and technologists can now align business goals with their own processes and technologies to deliver measurable improvement in data quality that translates to competitive advantage.

The Trillium Software System is comprised of several integrated modules, built from the ground up. Because Trillium Software's goal is to provide solutions based on the needs and budgets of our customers, suite components described below are also available as stand-alone products.

### TS Discovery

Collaborate across business and IT resources to assess large volumes of data within and across systems. Robust data profiling capabilities allow users to understand data domains, formats, patterns, and relationships as they exist within the data itself in preparation for data migration and data integration initiatives. In-depth data profiling ensures the data will support business requirements. Further, users can easily see whether data conforms to specific business rules and defined data standards. TSS routinely assesses data to ensure that high quality is maintained at all times and monitors production systems for anomalies.

### TS Quality

Cleanse, standardize, and match any data: name and address data; product data; asset, material, and location data; transactions; etc. Validate worldwide address data against authoritative sources and identify relationships among records. World-class global capabilities and automated, rules-based intelligence give organizations a simple but complete solution to handle massive volumes of data—out of the box. Organizations can further customize rules and adapt to meet changing business needs, responding quickly to competitive challenges and improving operational efficiencies.

### TS Enrichment

Complement, supplement, and amplify your current data by drawing on over 5000 third-party sources, e.g. NCOA. This service provides a fully automated process for appending third-party data seamlessly for storage and distribution.

### TS Insight

Gain visibility into the current state of the quality of your data to pinpoint problems quickly and efficiently. Monitor business rules and data quality metrics graphically through a customizable Data Quality Dashboard. Use scorecards and trending information, disseminated through a browser-based interface, to communicate data initiatives, results, and goals. Because TS Insight lets you see which data sources have data quality issues and which do not meet minimum corporate threshold and acceptability levels, you can forecast and allocate the right resources to improve and optimize the business processes that impact them.

# Product Architecture and Key Concepts

The Trillium Software System (TSS) is specifically architected to provide robust functionality for and seamless transition between profiling, assessing, validating, improving and monitoring global data quality across and throughout a distributed enterprise.

In this release, Trillium Software directed development efforts to augment the degree of automation across all functions. The value of automation manifests itself in two key areas:

- A powerful, automated data profiling solution that analyzes data and quickly uncovers data anomalies, inconsistencies, missing data, duplicate data, and data that does not conform to business rules.
- A robust data quality engine that lets project teams repurpose and extend out-of-the-box and customized business rules and make them available as services to the enterprise.

TSS allows organizations to integrate applications and business process throughout the enterprise into ready-made, out-of-the-box workflows and makes customization of existing processes easy and intuitive.

Thoughtful design leverages pre-packaged profiling workflows and data quality rules and processes so that enterprises can apply business rules and standards to one project and repurpose them to others at any time. By implementing this re-use strategy in an easy-to-use paradigm, TSS offers a consistent, value-based approach to managing the data quality life cycle from profiling through improvement.

Key architecture considerations:

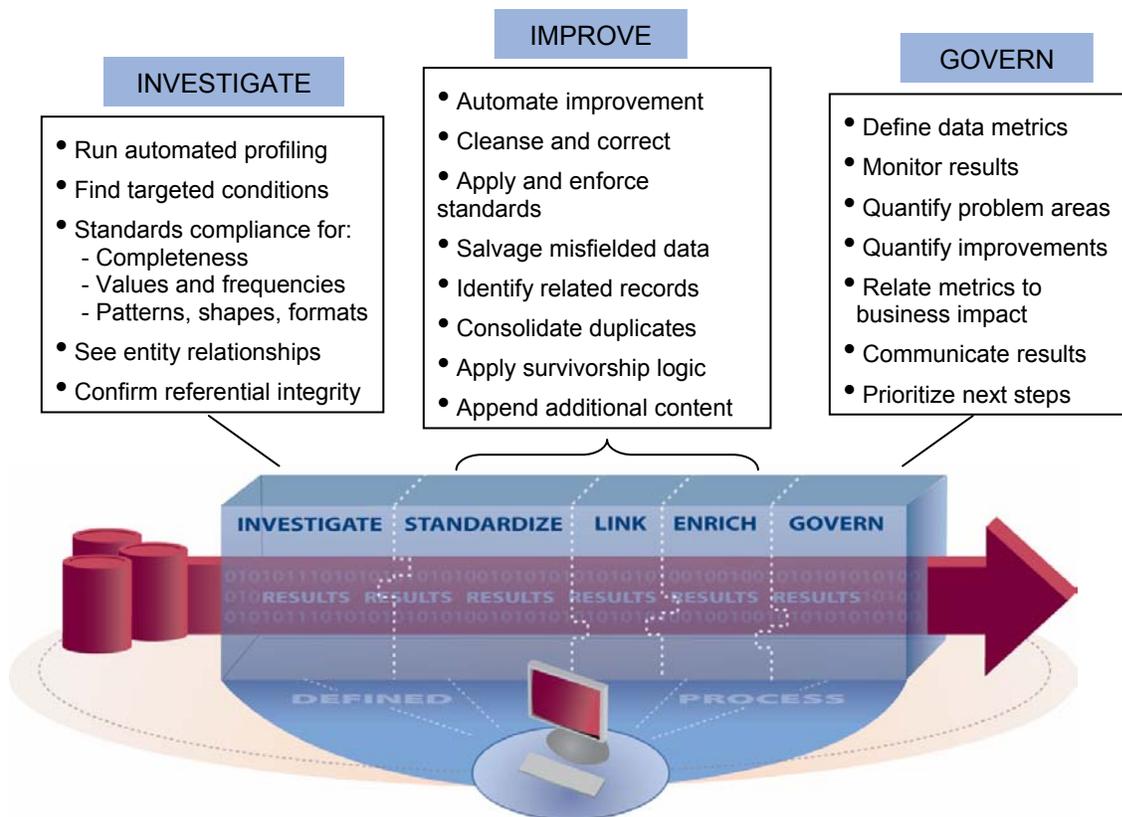
- **Ease of use:** provide efficiency and flexibility
- **Collaboration and communication:** fulfill business requirements and meet goals
- **Global readiness:** support complex needs of global organizations
- **Re-usability:** simplify development and maintenance
- **Scalability:** grow and extend usage over time, as needed
- **Interoperability:** interact seamlessly with applications throughout enterprise

Throughout this document, we discuss how we employ these key architecture considerations in the context of data quality life cycle management.

## Data Quality Life Cycle Management

A fully integrated, total data quality suite, TSS offers a single user experience for complete, global data quality life cycle management across the enterprise. Increased support for business user involvement in data quality diagnostics, analysis, and resolution ensures that efforts and results meet business expectations. TSS is purpose-built for seamless movement to and from views related to each phase of the data quality management life cycle, with particular emphasis on investigation and improvement (process design, development and test phases).

A non-technical interface lets business users, data stewards, and technologists collaborate to define, design, and execute data quality standards enforcement. Flexible and streamlined workflows let TSS adapt to your business processes and enable seamless transition from three main process areas of the data quality life cycle:



### Investigate

Invoke automated profiling capabilities to gain valuable insight about the current state of the data. Early discovery highlights issues, anomalies, and previously unknown data problems by exposing data content, context, patterns and relationships independent of the original data structure(s). Delivering a more comprehensive understanding of enterprise data across systems translates to more accurate project planning, scoping, and feasibility assessment and reduces risk in meeting daunting project deadlines and aggressive budgets.

### Improve

TSS lets data profiling output seamlessly transition into the data quality engine. To achieve economies of scale, define and design cleansing, matching, validation, and enrichment processes and repurpose them as data quality services throughout the enterprise. Deploy services in batch



or real-time using the same rule sets and standards across an unlimited number of applications and systems.

## **Govern**

Ongoing and cross-functional, data governance spans committees, business users, and technologists. Rich, powerful features and a unified interface promote collaboration and enable swift process refinement. The non-technical user interface provides visibility into the current state of data quality and related processes. The purpose-built data quality life cycle management platform lets users involve both business and project stakeholders with ease, promoting early feedback and issue resolution to ensure that business needs and expectations are met.

## **Data Quality Life Cycle Management within Enterprise Projects**

Trillium Software applies technology to data quality life cycle management in several ways. First, it streamlines the project management process by making communication and collaboration easier throughout the course of technology initiatives. Second, TSS enables seamless transition between the conceptual, data quality life cycle stages of *Analysis*, *Design*, and *Development* (see page 11) and enables the cyclical application of these processes over time (as described below). During the analysis stage, users concentrate on data investigation. During the design stage, they employ as much out-of-the-box content as they wish and define additional business rules and processes. During the development stage they can create and or modify these business rules and processes as required. The ability for any user to transition easily from one stage to another highlights Trillium Software's value-based approach to total data quality.

## **Multi-user Environment Supports Collaboration**

TSS allows enterprises to leverage the combined value of IT with valuable business-user insights regarding how data is created and used, for a complete and thorough understanding of enterprise data. The single interface and streamlined workflows facilitate active participation of the business community in the process of addressing data quality as an organization and helps align data and its quality to business priorities.

## ***Cross-functional Communication***

Because of its ability to bridge the communication gap between business and IT, TSS empowers a profound change in the way companies use their data as an asset. For the first time, organizations can manage data quality throughout its life cycle – from introduction into the business environment through all life cycle phases including cleansing, enrichment, and integration into down-stream applications, and ending in archiving and finally, deletion. It enables cross-functional teams to analyze data, set standards, monitor, communicate and collaborate with ease.

## ***Seamless Transition across Tasks***

Built-in rules optimize the discovery process so teams can collaborate about the data and implement immediately the insight that drives the best-fit process design. Teams can easily leverage insights derived from the investigation phase to drive process design. Process definition wizards facilitate the creation of new data quality processes during development and enable users to view results immediately to tune and interact with data quality processes. The Trillium Software System reflects decades of field-tested, best practices related to the investigative and improvement phases of the data quality management life cycle. Finally, during the QA/test phase, teams can review monitoring results and further tune business rules and processes to validate that results meet or exceed business expectations.

Country-specific, out-of-the-box content exists for quick-starts of global projects and users can easily tune these rules and processes and review results with key stakeholders.

## Global Readiness

Trillium Software System provides the most out-of-the-box capabilities with pre-built country-specific rules, automated data flows, and field-tested best practices. While other vendors outsource their global business rules to third parties, TSS delivers country-specific business rules built and tested within the product, from the ground up. It easily processes mixed country data (records) dynamically, identifying the true country of origin, and then subjects the data to a process that applies appropriate culture- and country-specific logic automatically. No other product is able to deliver such localized, culturally-aware, and intelligent recognition logic out of the box.

### Support for Global Needs

<b>Robust</b>	Australia Canada	United Kingdom United States	Japan*			
<b>Mature</b>	Belgium China* France Germany Hong Kong	Italy Malaysia Netherlands Portugal Singapore	South Korea* Spain Switzerland Taiwan*	<b>House Level Address Validation</b>		*Double byte
<b>Standard</b>	Argentina Austria Brazil Brunei Chile	Colombia Denmark India Ireland Jamaica	Mexico New Zealand Peru Philippines Saudi Arabia	South Africa Sweden UAE Venezuela	<b>Complete customer data standardization and matching rules</b>	
<b>Basic</b>	(All Countries)		Basic Country Data Project		<b>City-level Address Validation Formatting and Correction</b>	

TSS is used worldwide to cleanse data in different languages from different countries. Trillium Software has a structured program dedicated to enhancing and extending global coverage—country rules are constantly incorporated into deep, out-of-the-box intelligence. Although country support is enabled (turned on or off) depending on licensing purchased, Trillium Software’s complete global capabilities are incorporated into the product, allowing you to expand globally at any point in time based on business need, without process reengineering or reconfiguration.

### Unicode Support

TSS provides the broadest double-byte Unicode support available and is the only solution on the market today that handles both single byte and double byte data within the same product. Additionally, it can translate data from one code page to another, even mixed data within the same record. This offers organizations scalable growth options as they expand either the scope of their data quality efforts or the breadth of their global business through acquisition or market growth.

Processing data for double-byte countries poses significant challenges, especially given that spaces are not used to indicate a word or thought break, and character placement (and what characters neighbor a particular character) significantly influence the meaning and context of data being processed. Broad-based and comprehensive out-of-the-box intelligence is required to implement context-sensitive processing to global data.

## Data Quality Services

A powerful data quality engine lets project teams repurpose and extend out-of-the-box and customized business rules and make them available as services to the enterprise. It allows them to integrate applications and workflows throughout the enterprise into ready-made, out-of-the-box workflows and makes customization of existing processes easy and intuitive. These rules can be packaged as data quality services, put in place in a **Production Data Quality Services Environment**, and used to cleanse both operational source systems as well as any centralized repository.

### Create Reusable Data Quality Processes

Data quality processes increase significantly in value when they can be used throughout an organization as opposed to serving only as a point solution. By creating and deploying data quality services, corporate standards can be enforced as appropriate, development lifecycle becomes significantly reduced, and Centers of Excellence and Best Practices can easily be shared and propagated throughout the enterprise.

A key goal at this stage is to *automate* as much of the data cleansing and enhancement as possible. Automation of the data quality services provides three key benefits:

- ⇒ It reduces manual intervention and therefore long-term costs of system operation
- ⇒ It increases the consistency of data values
- ⇒ It provides an auditable and repeatable process, thus increasing security and compliance with regulations such as Sarbanes-Oxley, Basel II and HIPAA

Arguably, the data quality engine and associated rule-sets are the most important part of any integration initiative. If you don't have a strong capability to produce valid, complete and accurate information, then you are building an expensive new repository filled with the same unreliable data that already resides in sources systems. Fortunately, there are tools available today that provide powerful data cleansing, based on out-of-the-box, tunable rules.

The user interface presents a development environment that allows a user to develop data quality processes and flows that can be used throughout an organization. Any portion of, or complete process can be reused, in either a centralized or distributed environment. This gives organizations complete flexibility to establish and enforce corporate standards as well as develop application-specific rules.

### Central and Distributed Strategy Support

TSS is architected to maximize flexibility and customizability and provides each organization the technology options most suitable to achieve their business goals.

Rules and processes can be:

- Defined once and integrated independently into multiple applications and workflows.
- Implemented once as centralized data quality services supporting multiple applications.
- Defined and integrated independently into disparate applications.
- Deployed as a single data quality service that supports multiple applications that request rules and standards specific to each application.
- Callable by one or more applications.
- Deployed in either batch or real time, using the same rule sets and standards across an unlimited number of applications and systems.



## **Service Oriented Architectures and Master Data Management**

Many companies are just realizing the many benefits of a service oriented architecture approach to enterprise data management as well as activities surrounding Master Data Management. As the need to take advantage of these benefits becomes more critical, project teams find themselves faced with data integration and migration tasks of enormous scale.

TSS aids with the integration design supported by the data itself and facilitates risk identification, source system assessment, and ongoing monitoring of SOA and MDM projects, regardless of size and complexity. TSS also supports ongoing data governance activities as systems are integrated into enterprise initiatives and tracks the quality of data available within the new system or network of systems to ensure that investments continue to provide high quality information to users over time.

# Investigate Data as First Step

## Perform Initial Assessment Using Automated Profiling

When new data sources are introduced to TSS they initiate an extensive automated profiling process. Both summary and detailed metadata results are stored within the data quality repository. By reverse engineering data, the metadata created represents the data itself, as opposed to metadata derived from documentation, application logic, business definitions, or database constraints. This produces an unbiased view of a data set, and any anomalies, conditions, relationships, etc. that exist within it. It is truly a view into the data without preconceived notions or limitations.

TSS allows users to explore metadata statistics and results within context by presenting the actual data in a well-organized manner. The information is easily navigated and digested using drilling capabilities that allow users to drill through to native records for complete context and relativity. Anomalies are automatically highlighted in red in the interface.

Using this approach TSS absolves the need to:

- define what profiling tests to run against each field
- run SQL queries (or submit multiple query requests to support drilling)
- support or protect source systems from extensive query activity which drains resources
- isolate operational systems from non-technical resources

Automated data profiling capabilities preserve time and resources throughout the course of any integration project by executing due diligence into the state of source data and delivering deep insight during the scoping and development phases.

## View Entity Metadata

Metadata is collected for each **Entity** (a table, a file, or the results of a dataset created as part of a defined process). Entity metadata includes:

- attribute count
- row counts
- row lengths (maximums, minimums, averages, etc.),
- schema details
- documentary descriptions

Metadata about potential keys and relationships is also collected (covered on next page).

## View Attribute Metadata

Metadata is collected for each **Attribute** (a column, a field, an element). Attribute metadata includes:

- the ranges of values
- unique (distinct) values
- frequency and distribution of values
- data types (string, integer, etc.)
- null counts, space counts, blank counts
- basic statistics (maximum and minimum values, maximum and minimum lengths, etc.)

- patterns (formats, masks, shapes of the data)
- other metadata derived from the actual data

## Analyze Keys

TSS generates metadata regarding the quality of fields functioning as keys (e.g., primary keys, identifiers, etc.). **Key Analysis** measures to what degree an attribute would serve as a key, without preconceptions of whether or not the field IS a key. It indicates how many unique values are within the field, how many duplicate values are within the field, and how many records are affected by duplicate (non-unique) values. Users can investigate results, drill into results for more details, and define the status at a field level regarding whether or not the field should be a *defined* key (permanent key).

This series of tests is extremely helpful in two ways: analysts have insight into the quality of fields that are known keys and have easy access to any records that have duplicate key values. Also, where a known key field has significant quality problems, other natural keys (simple or compound) are easily identified to help define uniqueness across a dataset.

## Identify Inconsistencies

The analysis and presentation of attribute **Patterns** (formats, masks, shapes of the data) identifies potential problems within structured data fields, such as dates, postal codes, product codes, customer codes or any attribute that should conform to a particular format and structure.

For text values, identifying potential problems, inconsistencies, or near duplicates can be more challenging. TSS applies phonetic representations of data using **Soundexes** and **Metaphones** to identify related values, duplications, misspellings, inconsistent punctuation, and other anomalies in textual data, unveiling quality issues that would otherwise be overlooked.

## Drill-down into Data for Context

Summary metadata and related results are represented to support further analysis. Using basic Windows Explorer concepts, analysts drill through results and observe increasing detail in areas where they choose to focus. Drilling is supported down to the actual record so that results can be seen in context of the original source data.

TSS offers a safe repository for exploration without affecting the performance of operational systems that serve as data sources. The architecture of the product not only supports extensive drilling without performance challenges, but isolates production and operational systems from repeated and unrestricted queries that support investigation and analysis of initial results. System administrators prefer this architecture given that data investigation is being performed by members of the business community, who may not understand the impact of their queries and drilling on a live, production system.

## Make Sense of Unstructured Text Fields

Analysts can perform **Word and Phrase Analysis** on particular fields. The results of this profiling test present a view of what words or phrases exist within a specified field and how many times that particular word or phrase appears. TSS automatically imports this information into the **Word and Phrase Rule Builder** to create standards for those fields as part of an automated improvement process. These rules can then be incorporated into the data quality process and used throughout the organization to organize, structure, and standardize data sourced from unstructured text fields (descriptions, comments, etc.).

## Quantify Targeted Conditions

### Measure Compliance with Data Standards

Analysts can measure how closely attributes comply with **Data Standard Definitions (DSD)**. In addition to the out-of-the-box test run, a user can customize a DSD to specify compliance with known standards relating to the following aspects of an attribute:

- User-specified Datatype, Schema, Schema length
- Patterns (defined by acceptable or non-acceptable patterns or combination of both)
- Values, Range, Uniqueness
- Spaces, Nulls
- Sum Check

Users can further customize tests to allow for thresholds or defined percentages of acceptability.

### Quantify Compliance with Critical Conditions or Business Rules

Analysts can view results that quantify to what degree data complies with a known critical condition by creating **Business Rules**. To compare sources against the same criteria and to track rule compliance trends over time, business rules can be set to run against more than one source. Results are separated by the number of records that pass and fail the expressed logic. Users can drill-down into summary results to glean further detail.

### Understand Dependencies across Attributes

Analysts can view **Dependencies**. There are naturally occurring data dependencies, where the value of one attribute is consistently associated with the value of another: e.g., a product code should always be associated with the same product name and product type values, and a zip code should always be associated with the same city value. When it does not, a potential quality issue is likely.

TSS finds dependencies within the data, displays the attributes involved and any conflicting values, and enables users to drill down to exceptions. TSS can also apply dependency analysis to normalize source data.

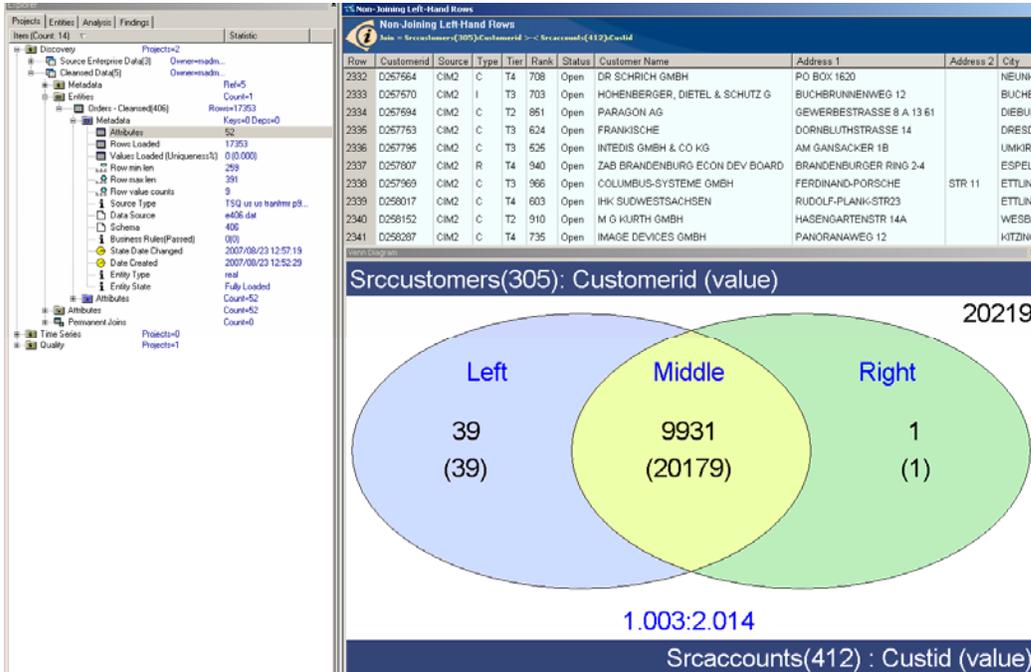
### Analyze Joins

Analysts can easily inspect how well data complies with prescribed **Joins**, either within or across datasets (files, tables, entities, and/or heterogeneous sources) and define a join(s), or request that all possible join combinations be analyzed (or a subset of all joins). Results can be analyzed by looking at the degree to which the two sets of data joined (percent), the cardinality of the join, and the outlier records (non-joining records, orphans).

TSS shows you how well your data will integrate, what problems will be encountered, and options to rectify issues, well before a first attempt at the actual integration process.

### Graphical Display of Information

TSS allows users to review the results of joins and relationships as **Venn diagrams** and **Entity Relationship Diagrams (ERDs)**. Venn diagrams show the relationship between two entities, the attributes and number of rows involved, and the number of outlier records. Users can drill down to see the actual values within original source records. ERDs show the relationships across entities, the cardinality, and are used to compare against data models, to ensure that the data as it naturally relates conforms to other data models.



## Communication across Teams

Built for team collaboration, TSS provides a number of ways to facilitate communication from one person to another within the tool itself.

## Notes

Users can create **Notes** associated with a particular entity and provide information to anyone that accesses the entity. Users can append additional information to a note and change priorities, status levels, etc. Using Notes features, a business analyst can impart important information to a technology team member or vice versa, or someone from one department can ask or answer questions for someone in another department. TSS maintains a full log of all history.

## Recodes

Business community members have expert knowledge about the business, how data is created, and how data is intended to be consumed. Often, these business users can determine what the literal values should be in records where there is a quality issue and provide recode values for others to use. The **Recodes** are immediately applied to correct a field, but can also be immediately incorporated into an automated improvement process.

## Flag records

At any time users can **Flag** records—those flags are carried with the entity. Flags can be set on records that: fall outside of a join, have erroneous values, are updated with a recode value, etc. Flags are an effective way to indicate what records meet or fail to meet a specified condition.

## Report Generation

TSS generates customized reports in a variety of formats from any view within the user-interface. Create HTML reports or export data, metadata, and findings into files and popular Windows desktop products to share information with team members not using the tool.

# Standardize and Cleanse Any Type of Data

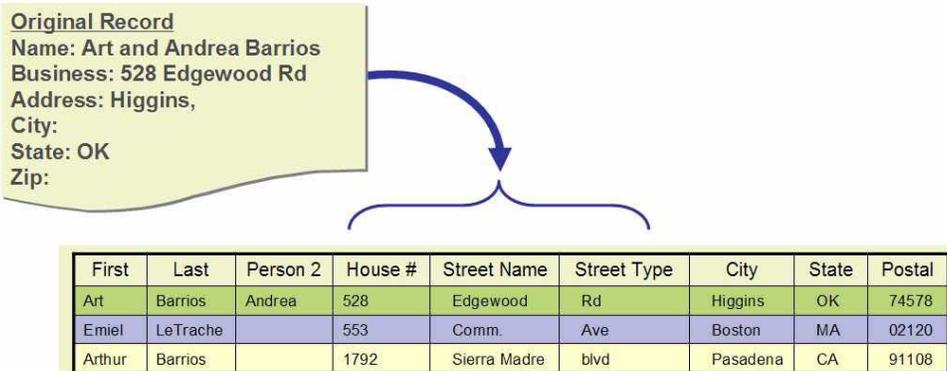
## Perform Complex Standardizations

The **Transformer** performs general cleansing on any or all incoming records by:

- Capturing valuable data buried in text fields based on literal values or data format (phone numbers, email addresses, account numbers, etc.)
- Flagging fields for special processing based on conditional logic
- Setting values, moving whole or partial data values from anywhere within a field, or performing specific actions based on conditional logic
- Translating global data from one code page to another for consistent processing

## Cleanse Name and Address Data

The **Customer Data Parser** is optimized for name and address data. It is driven by modifiable, country-specific rules tables containing definitions and interpretations for all commonly occurring words and word patterns that occur in names and addresses. It uses a combination of identifying key words (e.g., “Street” or “Ave” or “Road” are all street types) and patterns (e.g., numbers preceding a word preceding a street type is the house number), in order to intelligently process data within context.



## Cleanse Data Specific to your Business

The **Business Data Parser** is used for cleansing and standardizing complex, non-name and address data, for example product descriptions, text fields, comments, unstructured data, etc. The Business Data Parser identifies key words, understands patterns associated with key word types, and derives relevant content based on the data available, code translations, and logic and rules provided as standards. The **Word and Phrase Rule Builder** assists in creating the rules used by the Business Data Parser to identify key words and phrases, and apply re-codifications.

## User – Controlled Parsing

The parsers identify data captured in the wrong field, interpret the type of data components available across the entire record, and process each component within context, according to standards and rules set by the user. Each improvement in a record not only increases the value of the information on the record for future business use, but also increases the effectiveness of any matching and linking processes applied to the data. Parsed records are assigned codes that indicate the success or failure of the parser to interpret meaning from the record. Using these codes, users can tune TSS to interpret data and further manipulate records with specific parsing conditions. This gives the user complete control over the data and improves results.

# Validate and Enrich Global Addresses

## Automated Country-Specific Rules

The **Global Data Router** automatically identifies the country of origin for each record and forwards the record to a process that utilizes the rules engine specific to that country. Users do not need to manually separate records, maintain separate files, or process a single file multiple times in order to expose records to rules that process data according to the context in which it was first created.

This feature is especially powerful for data sources that provide data from multiple countries, but do not capture “country” as a field, such as web forms or call center screens.

## Address Repair and Validation

Country-specific **Geocoders** validate and repair address data to meet the standards of local, authoritative sources. Geocoders use country-specific postal directories (USPS in the US, Royal Mail PAF in the UK, etc.) to verify whether the address actually exists and, where appropriate, makes any necessary corrections or additions to the address, including its post code. Each record is allocated a fail-level code to indicate an address match or the reason for validation failure.

## Conditional Logic

Address data components are recompiled into a single field as a validated mailing address. However, if desired, the **Data Reconstructor** empowers a user to incorporate if-then logic to build other outputs or to rebuild addresses differently based on the level of validation from the geocoder. For instance, users may opt to use source data in lieu of a validated address when vanity cities are identified and updated according to USPS standards, in favor of customer preferences.

## Geographic Enrichment Options

Addresses can be further enriched with geographic data to serve a number of different business purposes or downstream applications, such as logistics routing, insurance zone analysis, market targeting, mapping, tax jurisdiction, etc.

### Geographic Enrichment for US

For US addresses, census and geographic data can be appended to records that have passed through a geocoder(s). In addition to enriching records with information from the USPS postal file that supports mail delivery and zoning, latitude and longitude data can be appended at several different levels of detail. Two options include assigning lat/long based on the center of the zip code, and assigning lat/long based on the actual house number “rooftop”. This data is derived from US postal and census data and maintained by TeleAtlas (formerly GDT). Additionally businesses can use Delivery Point Validation (DPV) to assert that the actual address exists.

### Geographic Enrichment for Other Countries

For Canadian records, census and geographic data can be appended to records that have passed through a geocoder(s). In addition to enriching records with information from the Canada Post file that supports mail delivery and zoning, latitude and longitude data can be appended. This data is derived from Canadian postal and census data and maintained by TeleAtlas.

## Identify Relationships, Link and De-duplicate

The **Relationship Linker** is driven by user-defined rules that identify duplicate records or relationships among records related to the same entity. It identifies matching records referring to the same business, household, individual/contact, product, etc. and identifies relationships linking a contact to a business, an individual to a household, a product to a product class, or other.

Users determine the definition of each entity. A household for example, could consist of any one person at a given address or people with the same surname at a given address. The matching rules compare data items, such as product names, product hierarchies, or organization-specific data; not just names and addresses. Rules are completely flexible, allowing the user to decide what constitutes a match, a non-match, or a suspect match. The same matching rules can be used in any deployment or implementation, guaranteeing consistency across the enterprise. Country-specific Best Practice match rules are provided for use as a starting point.

Relationships are identified for records reflecting situations such as contacts with multiple addresses, businesses with different naming conventions, product names with abbreviations, multiple contacts at the same business, multiple individuals within the same household, or multiple products in the same product class.

## Match within or across Files and Against Databases

TSS easily matches and links records within a single file or across an unlimited number of files in high performance batch processes that can be incorporated into any job schedule or workflow. Able to perform reference matching to match a file or transaction against a master database or external data source, it provides the ability to maintain unique reference numbers.

Reference matching is often performed in online environments, comparing an incoming record with an underlying database and preventing duplicate records from ever entering the system. Reference matching is frequently used by customers for matching against third-party data sources, such as geo-demographic data or suppression files.

## Single View

The **Commonizer** is used to create a single view by standardizing data across a set of matching records linked by a common key assigned during the matching and linking process. Rather than being forced to select a single record from a group of matching records, it is possible to collect the best data from the set of records and build a new “best” record. Each record might contribute a value that none of the other records contains. Consolidation and survivorship is driven by user-defined rules specifying which field values should be carried forward. Rules are defined at a field by field level, so the user has complete control over the logic driving survivorship for every piece of data. Most commonly applied survivorship and consolidation options used include:

- highest / lowest value
- longest / shortest value
- most / least commonly occurring value
- most / least recent value
- value from a specified input source

# Incorporate Data Governance into Processes

As part of developing data quality processes, define how your organization will measure and monitor the status and effectiveness of your enterprise data quality over time. Use TSS to identify critical data assets and define data quality metrics, data standards, business rules, and critical conditions as well as the process by which these will be regularly measured and reviewed.

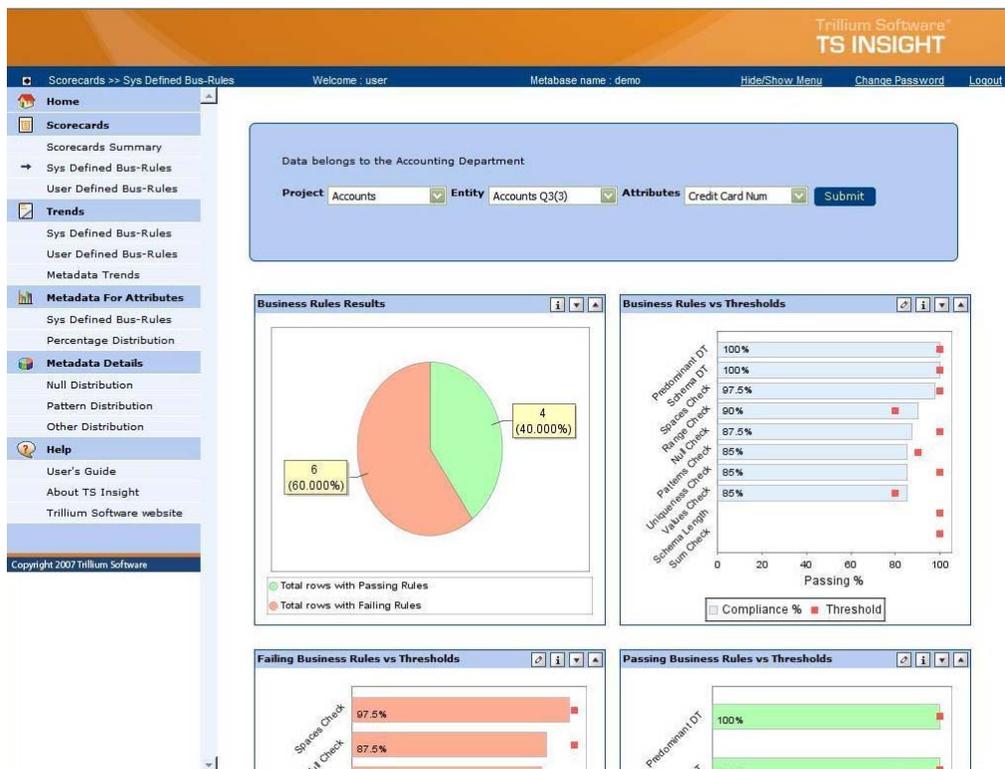
Further support data governance by developing data quality processes that enforce data standards as defined by the data governance team. Measure, monitor, and govern compliance with such standards over time using TSS. The data quality repository provides an ideal architecture for maintaining an auditable repository of quantified results and supporting data that can be used to communicate with key stakeholders and management over time to optimally manage allocation of resources and prioritization of next steps for data governance teams. Additionally, the compilation of data quality indicators and metrics and the trends of these over time can be used to determine a true ROI and the actual value of the data quality program to the business itself.

## Complement with TS Insight for a Data Quality Dashboard

TSS serves as the engine that powers TS Insight, a data quality dashboard solution designed to support ongoing monitoring and communication of data quality throughout a cross-functional or distributed team.

TS Insight displays results of TSS analysis within a data quality dashboard that provides users an at-a-glance view of how data currently measures against data standards and defined quality compliance thresholds. Data quality scorecards, graphs, and trends are easily viewed through a web browser and each user can customize their interface to focus on data they are most interested in tracking over time.

Data Governance teams use TS Insight as a way to track, monitor, and communicate with senior management about data quality throughout the enterprise.



## Integration with Other Applications

The Trillium Software System delivers unbeatable enterprise performance and scalability—it uniquely boasts consistent sub-second performance in both batch and real time environments. A number of deployment options are available through which the Trillium Software System can be implemented at an enterprise level. Both the Director and Web Services offer options supporting service-oriented architecture needs.

### Director

The **Director**, an optional component that functions as a data quality application server, can be deployed in either a centralized or a distributed environment. The Director routes requests for data cleansing and matching from multiple real-time and batch applications simultaneously to appropriate servers. It provides multi-threading, load balancing, and tracing features, and gives the user full control and monitoring capabilities through an administrative GUI interface.

### Web Services

Trillium Software functionality is available through Web Services. Interfaces optimized for IBM WebSphere®, BEA WebLogic® and Microsoft .Net® are available to facilitate seamless integration into multiple workflows, providing data quality services to applications throughout the enterprise.

### Enterprise Application Connectivity

Trillium Software partners with other software vendors to provide ready-integration with many applications. The Trillium Software System can be called in real time or as part of a batch process. Various enterprise application connectivity options give customers guaranteed interoperability and significantly reduce development and implementation costs. Trillium Software technology partners are long-standing and have proven integration and cooperative working partnerships that deliver solutions for the fast path to success.

### ETL Connectors

The Trillium Software System is frequently embedded within ETL applications as part of a job workflow. Cleansing and matching functions are called by ETL tools, ETL applications pass data to the data quality engine, and results are passed back to the ETL tool. Customized Connectors are available for the following ETL tools and provide tight integration and seamless data flow from one environment to the other:

Ab Initio Co>Operating System\*

Ascential Data Stage

Informatica PowerCenter

Microsoft Data Transformation Services

Oracle Warehouse Builder

*\*Maintained by Ab Initio*

### CRM and CDI Connectors

TS Quality can be called through customized connectors to interface with a number of different partner CRM and CDI applications, for either real-time interaction or batch processing. These connectors provide a seamless integration of TS Quality's data quality functions within a workflow specific to the third-party application, such as entering a new customer, logging a service call

request, incorporating a response to a marketing campaign or online product purchase into a unified view, etc.

Partner-specific Connectors, delivered with customizable configuration files provide flexible options for deployment within these third-party enterprise applications:

SAP – SAP certified

Siebel – Siebel certified

Oracle Customer Hub

Siperian Hub\*

*\*Maintained by Siperian*

## XML for Data Standards Sharing

TS Quality takes advantage of the industry standards available to ease application integration, including XML. Data quality standards and rules can be stored in XML in addition to the traditional text-based form. These files can be shared with any other XML-compatible technologies, facilitating the exchange of metadata within an enterprise.

## Repository API

TSS offers an API that lets users easily extract both summary and detailed metadata. Using this API, users can export this data into any third-party reporting packages they choose.

## Technical Specifications

### Minimum Requirements—Client

#### Supported Platforms - Client

Microsoft Windows 2000 Professional/Server, SP4

Microsoft Windows XP Professional, SP2

Microsoft Windows Server 2003 Standard Edition, SP1 (32 bit)

*The operating system should be running a minimum DirectX level of 8 and the client machine must have Internet Explorer v5.5 or greater.*

### Minimum Hardware Requirements - Client

#### Screen Resolution

1024x768 pixels minimum, 1280x1024 pixels or above recommended

#### Color Depth

16 bit (65536 colors) minimum

#### DirectX Level

8.0 minimum

#### Processor

Intel or AMD (Pentium III or Athlon) 1.0 GHz minimum

#### Memory (RAM)

512 MB minimum

#### TCP/IP Connection

10 Mbps minimum

#### Available Disk Space

200 MB minimum

Hardware Requirements – Server			
	Minimum	Recommended	Power Server
<b>Memory Requirements</b>	RAM: 2 GB Maximum Users: 2 Loads: Overnight Only	RAM: 4 GB Maximum Users: 5 Loads: Moderate Daytime	RAM: 8GB Maximum Users: 10 Loads: Daytime
<b>Network TCP/IP Connection Speeds</b>	100 Mbps	1 Gbps	
<b>Hard Disk Storage</b>		100 MB on multiple server disks or RAID array plus 2 to 5 times the size of the data source (dependent on data duplication and other factors)	

*These figures represent guidelines only. Please see your Account Manager to determine specific site requirements.*

Supported Server Platforms	Required patch (if any)
<b>UNIX</b>	
AIX 5L (v5.1, 5.2, and 5.3)	File system should be large file enabled. Requires April 2006 IBM C++ Runtime Environment Components for AIX
Solaris 8	SunOS Release 5.8 requires Patch ID 108434-22 or greater. Prerequisite Patch ID 109147-07 or greater.
Solaris 9	Patch ID 111711-16 or greater.
Solaris 10	No patch required
HP-UX 11.i V1 (PA RISC)	PHSS_33035 (aC++ runtime patch) PHSS_33944 (linker patch) PHSS_33901(libcl patch )
HP-UX 11.i V2 (IA64)	PHSS_35978 (aC++ runtime patch) PHSS_35979 (linker patch) PHSS_34859 (library patch) PHSS_34853 (math patch)
Linux Red Hat Enterprise Linux 2.1 and above	No patch required
Linux SUSE Version 9.2 Professional and above	No patch required
<b>Additional UNIX Server Requirements</b> The following components are required and <i>must</i> be installed: <b>Inetd (Xinetd for Linux)</b>	
<b>Microsoft Windows</b>	
Microsoft Windows 2000 Server	Service Pack 3
Microsoft Windows Server 2003 Standard Edition	Service Pack 1 (32 bit)
All required patches are available from your operating system support provider.	

## Suggested Server Hardware Configurations

Platform	Server	Speed	Processor
IBM AIX	RS/6000	900 MHz	eServerP5 and pSeries Platform Power 5
Sun Solaris	Sun Sparc	900 MHz	Ultrasparc IV
HP-UX (PA RISC)	HP PA RISC (PA 2.0)	750 MHz	PA RISC PA 8800
HP-UX (IA64/Itanium)	HP Itanium 2 (IA64)	1.4 GHz	Intel Itanium 2
Linux and Windows	Any	3 GHz	Intel Xeon MP

## Data Requirements

### Flat File Data Sources

TS Discovery can import (or link) to data from any of the following flat file data sources:

#### TS Quality files

*Duplicate row, checksum calculation, sparse row counting, or row length statistics are not supported for Trillium connections.*

#### Delimited files

- With ASCII, extended ASCII, or Hexadecimal delimiters
- With or without ANSI SQL DDL

#### Flat, fixed length files described by COBOL copybooks.

- Data must match Copybook layout
- Various character encoding including, but not limited to ASCII, EBCDIC, and Unicode
- Big or little endian byte orders
- One or two byte data alignment
- Common COBOL copybook features such as:
  - REDEFINES clause;
  - OCCURS clause;
  - FILLER clause;
  - COBOL data types: 9, X, A, B;
  - numeric storage formats: COMP-3, Packed-DECIMAL, COMP, COMPUTATIONAL, or BINARY;
  - comments beginning with \*;
  - descriptors such as USAGE IS, DISPLAY, SEPARATE, LEADING, and TRAILING;
  - unnamed fields; and
  - unsigned COMP-3 fields.

*Copybook features such as Multiple Record Types and OCCURS DEPENDING must be manually removed from the Copybook and the data file prior to loading into the Trillium Software System.*

## Relational Data Sources

TS Discovery can import (or link) to data from Relational Database Management Systems (RDBMS):

- via direct connection to Oracle and DB2;
- via ODBC compliant RDBMS connection; or  
*Duplicate row checksum calculation, sparse row counting, or row length statistics are not supported for ODBC connections.*
- via RDBMS extraction into a delimited file with a corresponding ANSI DDL.  
*NATIONAL data type for RDBMS loads is not supported.*

## Support for Relational Data Sources

Supported Database Clients for Direct Connections	Version	Driver
DB2	7.1, 8.1 (Fixpack 5 or greater)	IBM
Oracle	8i*, 9i and 10G†	Oracle

Tested ODBC Connections	Version	Driver
Allbase	7	HP MB Foster drivers
DB2	7.1, 8.1 (Fixpack 5 or greater)	IBM
Informix	9.3/Iconnect 2.8 32-bit client	IBM driver
Microsoft Access	Access 2002 with SP3	Microsoft driver
Microsoft SQL Server	SQL Server 2000 with SP3	Microsoft driver
Oracle	8i, 9i and 10G	Oracle
Sybase	12.5.x	Sybase driver
Teradata	v3.05.00.00 against a version 2 release 6.2 database	NCR Corp driver
<i>ODBC Connections to any other database or database version require verification from Trillium Software Customer Support.</i>		

\* Direct connection via an Oracle v8i client is not supported for the HP Itanium or Linux platform.

† Direct connection via an Oracle v10G client is not supported for HP-UX on a PA-RISC platform.