

The Value of Metadata-Driven Automation for the Modern Enterprise

Automated, meta-data driven code generation for faster time-to-value in data preparation, deployment and governance

The Automation Framework: Why It Matters

Data governance is more important to the enterprise than ever before. It ensures everyone in the organisation can discover and analyse high-quality data to quickly deliver business value. It assists in successfully meeting increasingly strict compliance requirements, such as those in the General Data Protection Regulation (GDPR). And it provides a clear gauge on business performance.

A mature and sustainable data governance initiative must include data integration. This often requires reconciling two groups of individuals within the organisation: 1) those who care about governance and the meaningful use of data and 2) those who care about getting and transforming the data from source to target for actionable insights. Both ends of the data value chain are covered when governance is coupled programmatically with IT's integration practices.

The tools and processes for this should automatically generate “pre-ETL” source-to-target mapping to minimise human errors that can occur while manually compiling and interpreting a multitude of Excel-based data mappings that exist across the organisation. That approach introduces errors that ultimately impact data quality. Additionally, the efficiencies gained through automation, including minimising rework, can help cut system development lifecycle costs in half.

An automation framework sustains data governance and accelerates project delivery through comprehensive and automated code generation for data integration, transformation and movement.

Data Governance and the System Development Lifecycle

Boosting data governance maturity starts with a central metadata repository (data dictionary) for version-controlling metadata imported from a broad array of file and database types to inform data mappings. It can be used to automatically generate governed design mappings and code in the design phase of the system development lifecycle.

The right toolset — one that supports a unifying and underlying metadata model — will be a design and code-generation platform that introduces efficiency, visibility and governance principles while reducing the opportunity for human error. Automatically generating ETL/ELT jobs for leading ETL tools based on best design practices accommodates those principles; it functions according to approved corporate and industry standards.

Automatically importing mappings from developers' Excel sheets, flat files, Access and ETL tools into a comprehensive mappings inventory, complete with automatically generated and meaningful documentation of the mappings, is a powerful way to support governance while providing real insight into data movement — for lineage and impact analysis — without interrupting system developers' normal work methods.

GDPR compliance, for example, requires a business to discover source-to-target mappings with all accompanying transactions, such as what business rules in the repository are applied to it, to comply with audits. When data movement has been tracked and version-controlled, it's possible to conduct data archeology — that is, reverse-engineering code from existing XML within the ETL layer — to uncover what has happened in the past and incorporating it into a mapping manager for fast and accurate recovery. This is one example of how to meet data governance demands with more agility and accuracy at high speed.

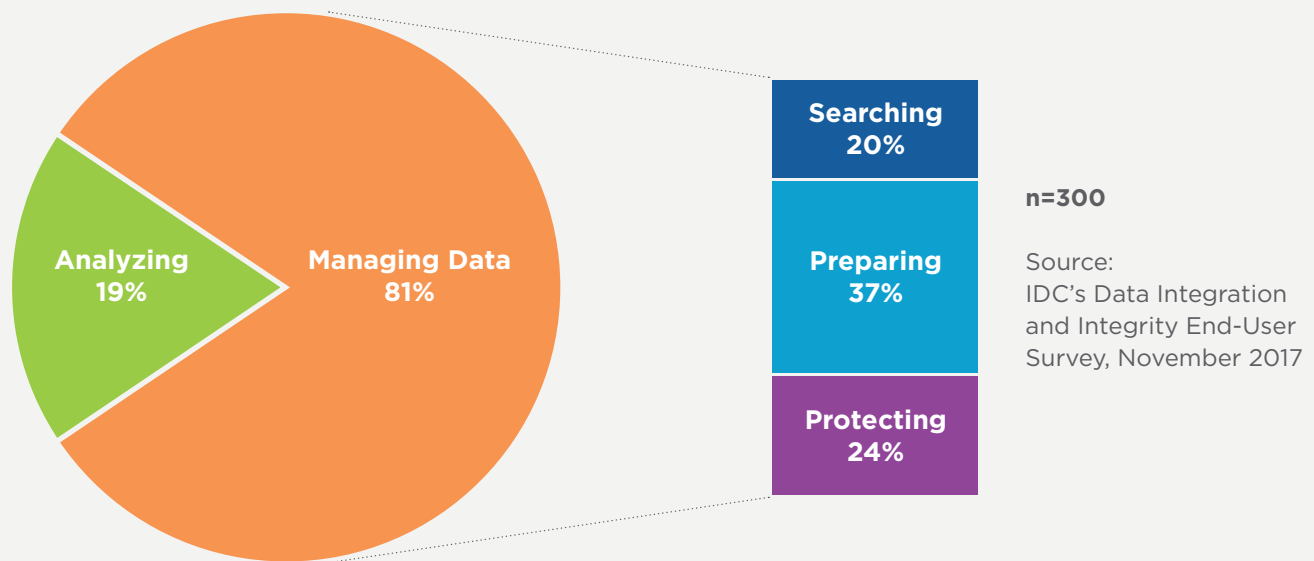
The Value of Code Automation Templates (CATs)

CATs open the door to simplicity, consistency and reusability of high-quality code that can be created and applied to almost any process where data integration matters. CATs are meant both to help improve governance across the software development life cycle (SDLC) stack and also to deliver instant business value. When CATs serve as blueprints to automate common and routine tasks in any development project, they open development to even junior-level staff to move the business forward.

Deploying reusable automation templates for data integration results in time and cost savings for multiple processes. For example, large migration projects can be more organised, version-controlled and faster. The organisation can quickly build and populate a Data Vault, or other structure, for long-term data storage using hubs, links and satellites for flexibility. And if a business goes through a merger or acquisition, CATs can be used to compare repositories from both companies for more complete analysis and development of the integration strategy.

PERCENTAGE OF TIME SPENT ON DATA ACTIVITIES (WEEKLY)

Q How many hours per week on average do you spend on each of the following data-related activities?



Automated code generation that delivers high-quality, reusable code reverses the ratio of time spent on data management versus data analysis to improve business outcomes.

Metadata-Driven Automation: How to Do It

The **erwin Automation Framework** is a metadata-driven universal code generator that works hand in hand with **erwin Mapping Manager (MM)** for pre-ETL enterprise data mapping; the governance of metadata; the governance and versioning of source-to-target mappings through the lifecycle; the governance of data lineage, impact analysis and business rules repositories; and automated code generation templates that save time and money in performing routine data integration tasks.

Its critical capabilities are:

- **The erwin Automation Framework's metadata-driven automation templates for universal code create simplicity, reliability, consistency and customisation for the integrated development environment.**

CATs can be created – for virtually any process and any tech platform – using the SDK scripting language or the solution's published libraries to completely automate common, manual data integration tasks. CATs are designed and developed by senior automation experts to ensure they are compliant with industry or corporate standards as well as with the business's best practice and design standards.

The 100-percent metadata-driven approach is critical to creating reliable and consistent CATs. It is possible to scan, pull in and configure metadata sources and targets using standard or custom adapters and connectors for databases, ERP, cloud environments, files, data modeling, BI reports and Big Data to document data catalogs, data mappings, ETL (XML code) and even SQL procedures of any type.

- **The erwin Automation Framework's high-quality coding provides blueprints anyone in the organisation can use.**

Stage DDL from source metadata for the target DBMS; profile and test SQL for test automation of data integration projects; generate source-to-target mappings and ETL jobs for leading ETL tools, among other capabilities. It also can populate and maintain Big Data sets by generating PIG, Scoop, MapReduce, Spark, Python scripts and more.

BEING ABLE TO RELY ON AUTOMATED AND REPEATABLE PROCESSES CAN RESULT IN UP TO 50 PERCENT IN DESIGN SAVINGS, UP TO 70 PERCENT CONVERSION SAVINGS AND UP TO 70 PERCENT ACCELERATION IN TOTAL PROJECT DELIVERY.

- **It incorporates data governance into the system development process.**

An organisation can achieve a more comprehensive and sustainable data governance initiative than it ever could with a homegrown solution. The framework's ability to automatically create, version, manage and document source-to-target mappings greatly matters both to data governance maturity and a shorter-time-to-value. This eliminates duplication that occurs when project teams are siloed, as well as prevents the loss of knowledge capital due to employee attrition.

Another value capability is coordination between data governance and SDLC within the **erwin EDGE platform**, including automated metadata harvesting and cataloging from a wide array of sources for real-time metadata synchronisation with core data governance capabilities and artifacts.

- **The erwin Automation Framework focuses on the value of data lineage and impact analysis for governance and risk assessment.**

Automated reverse-engineering of ETL code into natural language enables a more intuitive lineage view for data governance. With end-to-end lineage, it is possible to view data movement from source to stage, stage to EDW, and on to a federation of marts and reporting structures, providing a comprehensive and detailed view of data in motion. The process includes leveraging existing mapping documentation and auto-documented mappings to quickly render graphical source-to-target lineage views including transformation logic that can be shared across the enterprise.

Similarly, impact analysis — which involves data mapping and lineage across tables, columns, systems, business rules, projects, mappings and ETL processes — provides insight into potential data risks and enables fast and thorough remediation when needed. Impact analysis across the organisation while meeting regulatory compliance with industry regulators requires detailed data mapping and lineage.

- **The use cases for the erwin Automation Framework support a wide spectrum of business needs.**

Intelligent automation delivers enhanced capability, increased efficiency and effective collaboration to every stakeholder in the data value chain: data stewards, architects, scientists, analysts; business intelligence developers, IT professionals and business consumers.

The solution makes it easier for them to handle jobs such as data warehousing by leveraging source-to-target mapping and ETL code generation and job standardisation. It's easier to map, move and test data for regular maintenance of existing structures, movement from legacy systems to new systems during a merger or acquisition, or a modernisation effort.

- **In addition to erwin's mapping manager, it seamlessly integrates with erwin's enterprise modeling and data governance solutions.** The **erwin Automation Framework** is a valuable addition to the **erwin EDGE** portfolio that transforms enterprise data into accurate and actionable insights by connecting all the pieces of the data management and data governance lifecycle.

As with all erwin solutions, it embraces any data from anywhere (Any²) with automation for relational, unstructured, on-premise and cloud-based data assets and data movement specifications harvested and coupled with CATs.

Benefits of the erwin Automation Framework

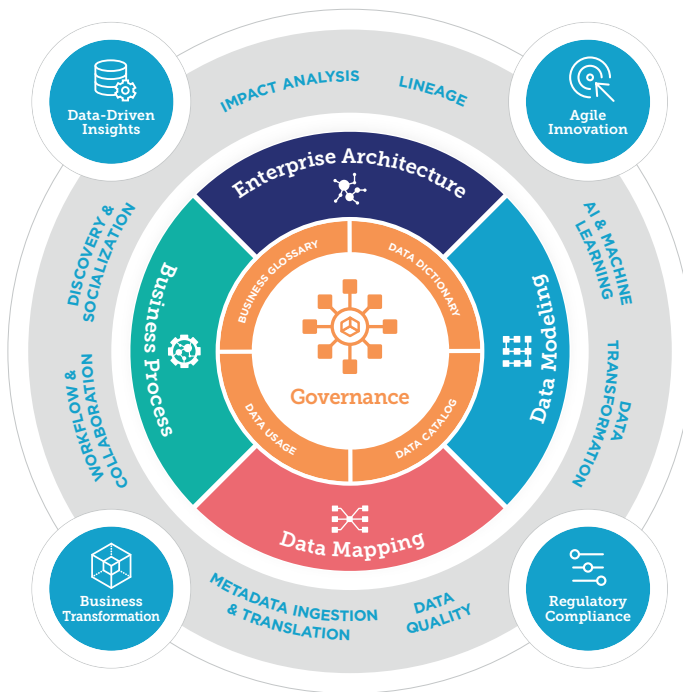
- Centralised and standardised code management with all automation templates stored in a governed repository
- Better quality code and minimised rework
- Business-driven data movement and transformation specifications
- Superior data movement job designs based on best practices
- Greater agility and faster time-to-value in data preparation, deployment and governance
- Cross-platform support of scripting languages and data movement technologies

A Key Component of the erwin Data Governance Platform

In most companies, an incredible amount of data flows from multiple sources in a variety of formats and is constantly being moved and federated across a changing system landscape. Often these enterprises are heavily regulated, so they need a well-defined data integration model that will help avoid data discrepancies and remove barriers to enterprise business intelligence and other meaningful use. IT teams need the ability to smoothly generate hundreds of mappings and ETL jobs. They need their data mappings to fall under governance and audit controls, with instant access to dynamic impact analysis and lineage.

With the **erwin Automation Framework**, data professionals can meet these needs at a fraction of the cost of the traditional manual way. They can have fast access to metadata with out-of-the-box scanners that bring more and richer information into the repository, so it's easier to build accurate mappings quickly. Centralised mapping and metadata accelerates the automation process while keeping everything version controlled, easily maintained and fully auditable. Comprehensive, reusable integration code is automatically generated in compliance with standards without demanding manual effort.

The solution accelerates and documents data movement and transformation projects with automated code generation — while reducing delivery timeframes and errors. This ability is a key component of the **erwin EDGE platform** in delivering an “enterprise data governance experience” for collaboration between IT and the business to discover, understand and unlock the value of data both at rest and in motion.



The erwin EDGE Platform: Comprehensive Data Management and Governance

Mature and sustainable data governance requires collaboration from both IT and business stakeholders, backed by a technology platform that accelerates the time to data intelligence. The erwin EDGE integrates data preparation, enterprise modeling and data governance.

For more information about the erwin Automation Framework or to request a quote, please contact info@sandhill.co.uk



erwin provides the most comprehensive data management and governance solutions to automate and accelerate the transformation of data into accurate and actionable insights. The erwin EDGE platform combines data governance, enterprise architecture, business process, data modeling and data mapping. Thanks to the broadest set of connectors for data preparation, modeling and governance, government agencies, financial institutions, healthcare companies and other enterprises around the world can use data to fuel their compliance, innovation and transformation initiatives.

Connect with us
at **sandhill.co.uk**



© 2018 erwin, Inc. All rights reserved. All trademarks, trade names, service marks, and logos referenced herein belong to their respective companies.