research

CHECKLIST REPORT 2020

# Accelerating Time to Value with Enterprise Data Preparation

## Eight steps for meeting data needs of self-service business users, data scientists, and enterprise IT governance

By David Stodder

Sponsored by:

Informatica

tdwi TRANSFORMING DATA WITH INTELLIGENCE™

MAY 2020

TDWI CHECKLIST REPORT

# Accelerating Time to Value with Enterprise Data Preparation

## Eight steps for meeting data needs of self-service business users, data scientists, and enterprise IT governance

By David Stodder

TABLE OF CONTENTS




tdwi

**Transforming Data With Intelligence™**

555 S. Renton Village Place, Ste. 700
Renton, WA 98057-3295

T 425.277.9126
F 425.687.2842
E info@tdwi.org

tdwi.org

## FOREWORD

Organizations are excited about data's potential, but gaining its full value is increasingly challenging. Data continues to grow rapidly in volume and diversity; meanwhile, users in every corner of the enterprise are demanding data for strategic and operational decisions, customer interactions, developing predictive insights, and collaborating on business processes. As more users analyze and share data, concerns about data privacy, security, and governance are rising.

These trends are putting pressure on data preparation—the sequence of activities that take data from raw ingestion through profiling, collection, integration, cleansing, transformation, enrichment, and governance to make it ready for users.

Preparing quality data at the scalability, performance, and throughput levels necessary today is not easy; without the right technologies and practices, data preparation is often slow, inconsistent, and incomplete, compromising the data's value. To keep workloads for dashboards, analytics, artificial intelligence, data science, and business applications well-provisioned with data, organizations need to address weaknesses in data preparation.

Left on their own, many users will resort to one-off data preparation using spreadsheets, desktop databases, or custom programs that prove inadequate for the speed, quantity, and variety of data. Recent self-service data preparation technologies are an improvement over manual work, but these tools often leave users with data silos full of inconsistent and insufficiently governed and secured data.

Organizations must establish an enterprise data preparation strategy that balances the flexibility self-service users want with the consistency, quality, and governance that modern, centralized enterprise data preparation offers.

Data scientists, analysts, and business users typically want to enrich analytics with multiple sources, but increasing the number of data sources increases data prep's complexity. Fortunately, modern enterprise data preparation solutions can apply artificial intelligence (AI) techniques such as machine learning (ML) plus automation to scale data integration and preparation.

Organizations also need modern enterprise solutions to cleanse big data so users can uncover errors, inconsistencies, and anomalies. Fronted by easy-to-use graphical interfaces, today's solutions hide complexity from users and apply AI/ML to accelerate data discovery, cleansing, and other preparation. Some solutions offer AI-driven recommendations that help users shape data preparation specifically to their needs.

Good data preparation increases data trust and security, but without an enterprise strategy, disconnected data preparation processes can make trust and security hard to establish. Enterprise data preparation can enable organizations to bring standardization and efficiency to data preparation processes that increase data accuracy, quality, validity, completeness, and conformance. Data engineers and managers need to ensure data quality both on premises and for data in the cloud.

An enterprise data catalog integrated with enterprise data preparation can increase both data trust and accessibility. A data catalog enables users to more easily find trusted data across the enterprise. Catalogs also promote reuse of data preparation routines, which can make data pipeline development faster and more standardized.

## FOREWORD CONTINUED

Finally, for data governance and regulatory compliance, organizations need data catalogs and enterprise data preparation to protect data privacy (including by masking sensitive information) and monitor data use and lineage.

This TDWI Checklist discusses these issues in more depth through eight considerations aimed at enabling you to accelerate time to value with modern enterprise data preparation. Data-informed decision making is a competitive advantage; gaining faster value from data is critical to business objectives.

# 1 INCREASE TRUST BY IMPROVING DATA QUALITY THROUGH ENTERPRISE DATA PREPARATION

Trust in the data is essential if executives, managers, and frontline workers are to make data-informed decisions and generate business value, but establishing trust is getting harder with the explosion in data volume and diversity.

Central to data trust is data quality; two-thirds of organizations surveyed by TDWI research (67 percent) cite poor data quality and completeness as their biggest challenge (see Figure 1).[1] Organizations need data quality practices and technologies that use AI and automation to scale effectively as the data universe expands.

Analytics, AI, dashboards, and data-driven business applications require a foundation of good data quality so that as data use grows, errors and inconsistencies do not spread downstream into business activities. An insurance company, for example, may choose to integrate claims data from multiple sources and analyze patterns and data relationships by including relevant semi- or unstructured contextual data from a data lake. If data sets have missing fields or incorrect data not discovered before users build predictive models and generate results, business actions based on the insights will be flawed. Once burned by bad data, users lose confidence; the flawed decisions damage not only the firm's business processes but also its reputation.

Departmental data quality solutions may only fix problems in one or a limited number of sources or for restricted quantities of data. These technologies and practices may be sufficient for specific projects and offer users some advantages through self-service capabilities.

However, departmental solutions often expose organizations to inconsistencies across data silos and provide little management visibility into whether users are sharing trusted, quality data. Organizations should evaluate solutions that address data quality, consistency, and completeness at enterprise scale to reduce the

**What are your organization's biggest challenges in enabling data assets to be used effectively for analytics, AI, and other data consumption? (Please select all that apply.)**

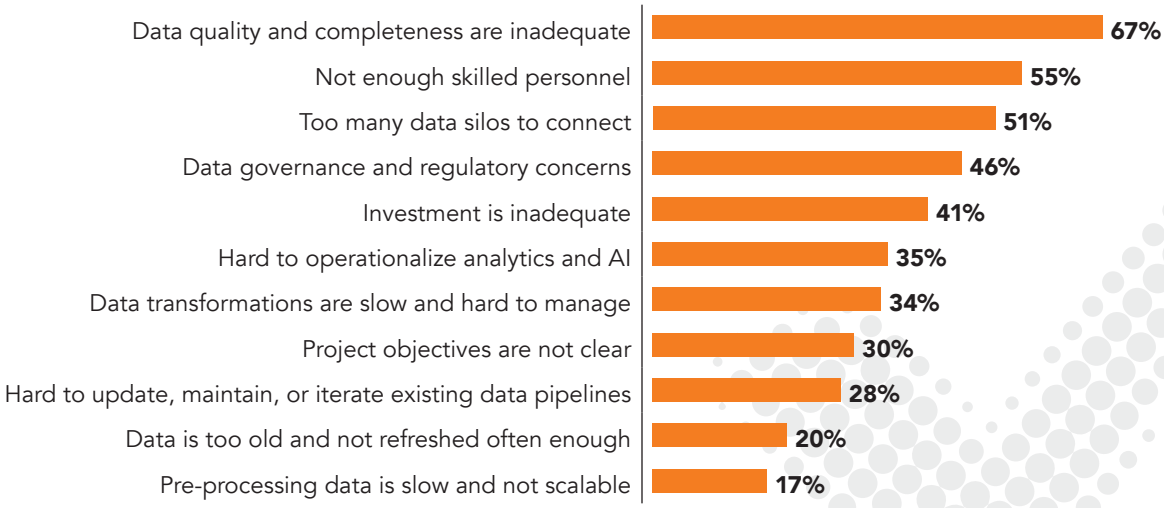


**FIGURE 1.** Based on answers from 138 respondents. Source: *TDWI Best Practices Report: Faster Insights from Faster Data*, online at tdwi.org/bpreports.

[1] For a further discussion of these results, see *TDWI Best Practices Report: Faster Insights from Faster Data*, page 18, available online at tdwi.org/bpreports.

## INCREASE TRUST BY IMPROVING DATA QUALITY THROUGH ENTERPRISE DATA PREPARATION CONTINUED

chance of bad data spreading into shared analytics, dashboards, and business applications.

To solve data quality issues at an enterprise level, organizations should:

- **STANDARDIZE DATA PREPARATION.** Organizations will reduce inconsistencies across departments if they formalize rules for data validation and correction at an enterprise level. Standardization can extend to documenting requirements and using enterprise solutions that monitor and measure progress toward meeting requirements.
- **APPLY AI AND AUTOMATION TO DATA QUALITY.** A modern enterprise data preparation solution should augment manual work by automating data quality processes. Use AI to increase speed and effectiveness in spotting errors and discovering missing data, anomalies, and other inconsistencies (such as in customer names and addresses). AI and automation are key to enabling enterprise IT to scale up data quality processes and handle increased demands as organizations democratize data and analytics.
- **IMPROVE MANAGEMENT OF AND ACCESS TO ENTERPRISE METADATA.** Data catalogs that standardize data definitions and other critical knowledge about data sources enable organizations to move faster to improve data quality and ensure that data cleansing steps are complete and effective. This topic will be discussed further in the next section.

## 2 ESTABLISH AN ENTERPRISE DATA CATALOG TO MAKE IT FASTER AND EASIER TO DISCOVER DATA

Metadata repositories such as data catalogs enable organizations to collect knowledge about how the data is defined, its location, lineage information about its origin and use, and how the data is related to other data. This knowledge is essential as data volumes become larger, more diverse, and more distributed. TDWI research sees growing interest in data catalog development and deployment; organizations that have developed and deployed a data catalog show higher levels of user satisfaction with data.

Driving interest in data catalogs is the desire to help users discover the data they need amidst increasingly voluminous sources; 79 percent of organizations surveyed regard this as the most important goal of developing a data catalog. Additionally, more than half of organizations (59 percent) want to use a data catalog to coordinate data meaning across sources.[2]

Enterprise data catalogs are particularly useful for this purpose because they can provide a centralized resource for resolving differences in how data is defined and related to higher-level entities such as customers, suppliers, or products. This can help organizations address any lack of data consistency and quality among departments.

An enterprise data catalog integrated with enterprise data prep can make it easier to share data definitions during preparation processes and use them to locate related and trusted data more rapidly wherever it is stored. Despite these advantages, not all organizations have a centralized data catalog because its development and maintenance has traditionally involved significant manual effort.

Modern solutions can apply AI/ML to augment human efforts to build enterprise data catalogs; almost a third of organizations (30 percent) surveyed by TDWI regard AI's role in building a data catalog or similar metadata repository as critical.[3] Organizations should evaluate solutions that use AI/ML to drive faster development of enterprise data catalogs and make it easier for business groups and IT to collaborate on defining new data and maintaining the catalog.

An enterprise data catalog can bring greater efficiency to data pipeline development. Using AI/ML and automation capabilities, a modern enterprise data catalog can help organizations curate data for pipelines by exposing which data sets are available; this cuts down on the time it traditionally takes for users to find trusted, relevant, and available data for pipelines.

Some catalog solutions enable users to browse information about data sets, classify them, and examine data lineage about the data's creation. To promote consistency, governance, and reuse, organizations should evaluate solutions that provide visibility into who has used particular data sets for pipelines—as well as for data views, analytics models, and other data-driven services.

[2] Ibid., 26.
[3] Ibid., 31.

# 3 USE ENTERPRISE DATA PREPARATION AND THE DATA CATALOG TO INCREASE USER AGILITY

Nearly all users want to be less dependent on IT. TDWI research shows that increasing users' self-reliance with BI, search, data exploration, and analytics is a top organizational priority, with 89 percent of those surveyed calling it important. The desire for self-service extends to data preparation; users want to spend less time waiting for IT's help in creating data pipelines and moving data through preparation and data quality processes.

However, self-service can have downsides. First, the majority of users do not have advanced tools for preparing data. To do it themselves, most users extract data into spreadsheets, desktop databases, or primitive open source tools, creating many of the consistency and quality problems discussed earlier. These practices cannot keep up with the speed of business given rising data volume and complexity.

Self-service users are forced to spend the majority of their time on data prep rather than on analytics. TDWI research finds that about half of organizations surveyed (49 percent) say that their users spend at least 61 percent of their time on data preparation, leaving too little time to focus on performing analytics and data interaction to develop business-critical data insights (see Figure 2).

Second, working on their own, self-service users may not access the enterprise data catalog, if there is one, to find data beyond the subsets and extracts with which they are working. Greater enterprise data catalog use could ensure that they access governed and trusted data sets and find additional relevant data.

Third, these users may not be aware of the time they could save by reusing data preparation routines vetted by IT and made available through integrated enterprise data prep and catalog systems.

Organizations should ensure that users are fully aware of the capabilities in enterprise data preparation and catalogs. They should train data stewards and power users to mentor less data-savvy users in how to use enterprise data preparation and the catalog resources to find, access, and use data faster and more efficiently. Data stewards can help users work with enterprise-level systems

**Thinking of your organization's most recent BI and analytics projects, what percentage of the total time was spent preparing the data compared to the time spent performing analysis and data interaction?**

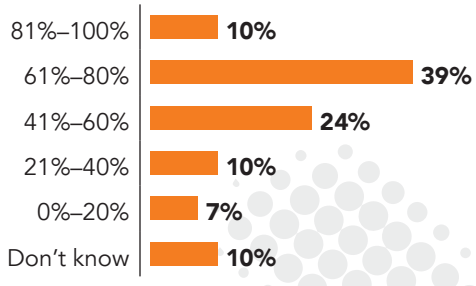


**FIGURE 2.** Based on answers from 130 respondents. Source: *TDWI Best Practices Report: Faster Insights from Faster Data*, online at tdwi.org/bpreports.

## USE ENTERPRISE DATA PREPARATION AND THE DATA CATALOG TO INCREASE USER AGILITY CONTINUED

to standardize data profiling, transformation, and calculations as the users interact with different data sets.

Organizations should implement modern enterprise data prep and catalog systems that can increase users' agility and save them time. They should focus on how enterprise data preparation and catalogs can reduce routine, repetitive work—for example, supplying data for standard dashboards and reports, affording users time to address unexpected, ad hoc business questions through analytics.

As noted, AI/ML capabilities in modern solutions can increase scale and drive automation for faster data discovery, cleansing, and further preparation; intelligent solutions can learn from usage patterns and recommend relevant data sets. Through integrated use of an enterprise data catalog, organizations can properly and automatically register governed and trusted data sets for reuse and sharing.

Self-service will continue to be a dominant trend because users demand agility. Organizations can address this need through modern enterprise systems. Enterprise data preparation and cataloging can let users stop doing all the work by hand so they can spend more time answering business questions than dealing with data prep and access intricacies.

# 4 MODERNIZE DATA PREPARATION TO IMPROVE ANALYTICS AND DATA SCIENCE DEVELOPMENT

AI/ML-enhanced enterprise data preparation and cataloging can improve productivity and efficiency for data scientists and other advanced analytics creators who often have to work manually or use primitive open source tools to find and prepare data. Even more than standard business users, data scientists and advanced analysts spend much of their time on data discovery and preparation, crowding out hours that they could be spending on analytics and AI/ML development.

Enterprise data preparation and cataloging can help these users; they typically work with large and varied volumes of data from structured, semistructured, and unstructured sources, often stored in a data lake, or (increasingly) in the cloud. They need technologies that can support highly interactive, fine-grained, and multivariate data analysis to develop models and discover patterns, correlations, and data relationships.

In some cases, data scientists and analysts want pipelines that simply locate raw data and move it rapidly into a data lake, including through data streaming. In other cases, they need more extensive data profiling, quality, transformation, and enrichment steps.

Banks, for example, are interested in using AI/ML to determine the creditworthiness of customers applying for loans and the value of those loans to the business. They want to look beyond FICO scores and income data to examine more variables—not only data provided by loan applicants but also related contextual data, especially if the applicant has little or no credit history. A broader view provided by analysis of contextual big data can improve the bank's judgment about an applicant's creditworthiness.

In this scenario, enterprise data prep can replace crude tools and manual spreadsheet work with smarter and more automated tools and more consistent preparation processes. Enterprise data preparation systems integrated with governance and security can apply data masking where appropriate to de-identify data, control access, and adhere to data privacy regulations. In these ways, enterprise data preparation can help banks reduce latency and errors in loan approvals, creating a competitive business advantage.

TDWI finds that for most organizations, the lack of skilled personnel is a leading barrier to the expansion of data science and advanced analytics. Enterprise data prep can help organizations overcome this barrier by enabling users to develop pipelines and preparation routines that they can standardize and reuse. This will enhance the productivity of not only data scientists but also "citizen data scientists," i.e., other users who want to perform advanced analytics and apply insights to business decisions.

Enterprise data preparation can thus reduce pressure on organizations to hire or contract outside data scientists in order to accomplish key projects. Organizations should evaluate solutions that offer easier-to-use capabilities that enable a wider range of users to address data preparation requirements for analytics and data science.

# 5 INCREASE THE VALUE OF CLOUD DATA LAKES WITH ENTERPRISE DATA PREPARATION AND CATALOGING

Data lakes have become more common as organizations collect different types of valuable data. Among organizations surveyed by TDWI research, 44 percent are using a data lake to make data available sooner for analytics and AI/ML workloads that explore high-volume, raw, and diverse data.[4] Apache Hadoop technologies once dominated data lakes, but today many organizations have shifted to the cloud to take advantage of flexibly priced object storage from providers such as Amazon, Google, and Microsoft.

However, data lakes are in danger of becoming enormous, impenetrable data swamps unless organizations have the right technologies to learn what the data means and extract value. Instead of enabling faster data availability, cloud data lakes could make finding and analyzing data slower and more difficult.

Cloud-based enterprise data prep technologies are a good fit for refining the contents of cloud data lakes and curating the data so users can interact with relevant and trusted data sooner. Data preparation technologies are key because data coming from multiple sources (sometimes in the hundreds) typically does not conform to a predefined schema, model, or set of data definitions before it is loaded into the lake, as would be the case for a data warehouse. Therefore, to prepare the data, organizations need to prep the data either before or after data ingestion, if not both.

Organizations may need to profile and validate some or all of the data as it is ingested so they gain immediate knowledge about the data and its sources. Rather than incur delays and inconsistencies that come with manual custom coding, data scientists and engineers can use data preparation tools capable of scaling to handle the volume and speed of data lake ingestion. Organizations may then want to use data prep tools to streamline additional data cleansing, transformation, and enrichment after loading to meet specific data pipeline requirements for analytics, AI/ML, and other applications.

Fundamental to gaining value from a data lake is effectively using an enterprise data catalog to manage metadata information. Metadata is crucial to giving data scientists a faster path to finding data, but it is also important for understanding what diverse data means, how different data sets are related, and how the data is relevant to analytics and AI/ML projects.

Ideally, organizations should tag data as it is loaded into the data lake and collect information that describes the sources, their content, who created the data, and information about its structure. This will be crucial in guiding how data scientists, analysts, and engineers interact with the data to determine whether to refine it further through enterprise data preparation processes.

The combination of cloud-based enterprise data preparation and the data catalog can enable organizations to position the cloud data lake within the fabric of their data architecture rather than as something separate from other systems such as the cloud and/or on-premises data warehouse. Organizations can then apply a consistent set of tools and a data catalog across all systems rather than using siloed tools for each one.

[4] Ibid., 27.

# 6 ENHANCE OPERATIONALIZATION BY INTEGRATING DATA PREPARATION WITH DATAOPS

With many organizations focused on using analytics and AI/ML to drive digital transformation of business processes, pressure is growing to reduce friction that impedes operationalization of prepared and governed data sets. As data grows in volume, speed, and complexity, operationalization is becoming even more challenging.

Projects such as digital transformation involve many stakeholders, from data scientists, analysts, and data engineers to business process owners, application developers, and IT managers. It can be helpful if organizations use methods and frameworks that offer repeatable (but not overly rigid) development guidelines for stakeholder teams.

Agile methods have become popular for data warehouse, BI, and analytics projects. These build on a foundation of software engineering principles for enhancing broader, less hierarchical team collaboration among stakeholders. A key agile goal is to deliver incremental value through "sprints" as projects develop so users can test components; then, given feedback, teams engage in continuous cycles of improvement toward higher quality deliverables.

Many organizations expand their agile experiences and implement DevOps methods, which bring together software development and IT operations to collaborate on improving development efficiency. DevOps enables automation on a larger scale so that organizations produce better-engineered software faster.

Today, organizations are building on agile and DevOps to implement DataOps methods. DataOps blends aspects of agile and DevOps to give organizations a framework for collaboration that reduces latency in constructing large numbers of high-quality data pipelines. Like agile and DevOps, DataOps supports the notion of continuous improvement and delivery; stakeholders can determine whether the pipelines fit their needs and propose improvements. To make DataOps work, organizations need scalable, more automated, and integrated enterprise data preparation that revolves around strong metadata management to reduce delays in locating data and operationalizing trusted data sets.

In sum, enterprise data prep and data catalogs together can help organizations succeed with DataOps in three important ways:

- **CONTINUOUS INTEGRATION AND COLLABORATION.** DataOps frameworks can guide stakeholders' collaboration throughout data life cycles. Enterprise data preparation can support this generally iterative process on a larger scale than departmental solutions. Stakeholders can use an AI/ML-enhanced enterprise data catalog to shorten the path to find relevant data, including as data is streamed in real time into cloud data lakes.
- **CONTINUOUS DELIVERY USING AN ENTERPRISE DATA CATALOG.** Enterprise data preparation and catalog systems combined with DataOps help organizations produce data pipelines that deliver governed data more efficiently. Organizations can use an enterprise data catalog to map defined business terms to data sets, making it easier to discover data relationships, patterns, and correlations.

  The catalog helps reduce disputes about the provenance and quality of the data, enabling data scientists to accelerate predictive model design and testing. Organizations seeking to respond to customer churn, for example, could

## ENHANCE OPERATIONALIZATION BY INTEGRATING DATA PREPARATION WITH DATAOPS CONTINUED

operationalize data and models faster so they can develop data insights in time to positively impact customer loyalty and engagement.

- **CONTINUOUS DEPLOYMENT OF DATA SETS FOR PIPELINES.** DataOps frameworks make it easier for organizations to envision how they will ultimately operationalize and deploy data sets. Modern enterprise data preparation and catalogs that use AI/ML to guide automation support iterative data pipeline development, making it easier to adjust and improve pipelines over time as they are deployed.

Organizations should evaluate how enterprise data preparation and data catalogs can underpin stakeholder collaboration fostered through DataOps to bring order to data pipeline development, which is often chaotic, inconsistent, and slow. Together, the technologies and framework will enable organizations to replace manual data prep with more automation and scale pipeline development through better reuse, standardization, and speed.

# 7 GAIN A HOLISTIC VIEW TO STREAMLINE END-TO-END DATA PREPARATION

In most organizations, data preparation consists of a hodgepodge of uncoordinated tools for profiling, cleansing, transformation, enrichment, and governance. As discussed in the previous section, using a DataOps framework with enterprise data prep and cataloging can help organizations improve collaboration between stakeholders, which is critical to coordinating different data preparation steps.

However, whether organizations implement a formal DataOps framework or not, they should explore how they could use aspects of the framework plus modern technologies to gain an end-to-end view of existing data preparation and pipeline development, including dependencies between processes. TDWI research finds that lack of visibility into how processes such as transformations depend on each other can lead to interruptions, bottlenecks, performance problems, and redundancy. Organizations can use this visibility to spot data preparation processes that are no longer needed, for example, ETL routines serving unused data pipelines, analytics models, or dashboards that should be discontinued.

An end-to-end, holistic view can reveal where modern enterprise data prep using AI and automation could replace pockets of unnecessary manual work. TDWI research finds, for example, that only 9 percent of organizations surveyed consider themselves very successful in automating repetitive tasks (38 percent are somewhat successful). Only 4 percent indicated that their organizations are very successful in creating end-to-end process efficiency from data collection to delivery (22 percent are somewhat successful, 41 percent call themselves "average," and 33 percent say they are unsuccessful).[5]

An important element of DataOps is clearly identifying the desired "end" of an end-to-end process. This includes understanding the personas who will ultimately use the data and analytics models and where they may experience problems. To this point, TDWI research finds that data scientists are far less satisfied with their ability to access data and information than traditional BI and data warehouse users such as business and data analysts. This suggests that organizations should focus on how they can streamline data preparation for data scientists.

Data scientists' workloads and data interaction requirements differ from those of traditional BI and data warehouse users. A holistic, end-to-end view will help organizations see common, recurring, but perhaps less-well-understood problems that data scientists are having. Organizations can use DataOps plus modern enterprise data preparation and cataloging to improve data scientists' experiences, standardize data prep among data science projects, be aware of dependencies, and increase reuse.

In examining personas, organizations should not overlook users and developers of embedded analytics and dashboards. Applications and cloud-based services featuring embedded analytics, dashboards, or AI/ML developed to monetize an organization's data will typically include external business partners and customers as users. Thus, it is critical that the collection of data preparation routines that serve these applications and services be efficient, well-integrated, and dependable. Organizations should map the sequence of data preparation processes and data pipelines used to support embedded applications and services so they can quickly remedy slow data loading, poor data quality, and inconsistent or incomplete data transformations.

[5] Ibid., 11.

# 8 IMPROVE GOVERNANCE WITH ENTERPRISE DATA PREPARATION AND THE DATA CATALOG

Governance rules and policies spell out how an organization protects sensitive data assets, meets regulatory requirements for data privacy, and takes steps to improve data trust. Yet, even as regulations such as the European Union's General Data Protection Regulation (GDPR), the California Consumer Privacy Act (CCPA), and others in the U.S. and around the world make governance a higher priority, the data landscape is becoming more challenging to govern. Two of the most difficult issues are the growth of self-service data preparation and management, which often spawns "shadow IT" data silos, and data lakes that ingest large quantities of big data, not all of which may be well governed.

Enterprise data preparation and data catalogs can play a critical role in establishing governance for each of these situations. Together, the systems can improve documentation and visibility into data lineage, which is necessary for data governance and compliance with data privacy regulations. By tracking data lineage through end-to-end data prep visibility and recording this lineage in an enterprise data catalog, organizations can learn the origin of the data, who and what applications are using and sharing it, and how it has been transformed and enriched in data pipelines, data lakes, or data warehouses. Organizations can establish governance as data is ingested into a data lake and used in analytics, AI/ML development, and dashboards.

Modern enterprise data catalogs implement AI/ML to improve mapping of data definitions and business terms to data sets, which makes it easier to find and govern all data related to a customer, for example. Organizations can then protect this data, whether it exists in a cloud data lake, a data warehouse, or business application database. Organizations can also use catalogs to speed response to compliance audits or meet other regulatory reporting requirements.

Addressing problems with self-service data preparation silos requires both technology solutions and organizational leadership. Governance committees populated by business and IT stakeholders can facilitate communication about priorities and how to overcome data ownership issues that may be preventing oversight. Data stewardship can play a role in governance by facilitating sharing of best practices and assignment of governance accountability to stakeholders for certain data domains. Some organizations will have data stewards mentor users of dashboards and analytics models to ensure that data use meets governance and quality standards. Data stewards should be represented on governance committees.

To govern data, IT will need to control data access privileges to enterprise data sources. This includes monitoring what data users can import, extract, and download. Organizations therefore need enterprise data preparation and data catalog systems that enable IT to manage access privileges. IT should integrate governance with tools for data masking and de-identification. If these tools are integrated with enterprise data preparation systems, IT can ensure that throughout preparation cycles the organization is protecting sensitive information through masking or by blocking user access. Finally, governance oversight should ensure that users delete unused data within a set time period so that it is not accidentally exposed.

## A FINAL WORD

Faster time to value with data depends on improving data preparation—an area where organizations commonly experience delays and inefficiencies. This TDWI Checklist has noted eight key issues.

We have highlighted the value of AI/ML-infused enterprise data preparation and enterprise data catalogs for relieving diverse users of common burdens in data preparation and pipeline development. At the same time, modern enterprise data preparation and data catalogs (plus frameworks such as agile methods and DataOps) can support flexibility to meet specific user requirements. This will enable organizations to find the right balance between user needs for self-reliance and enterprise requirements for standardization and governance.

## ABOUT OUR SPONSOR

Informatica

Informatica is a proven enterprise cloud data management leader that accelerates data-driven digital transformation. Informatica enables companies to fuel innovation, become more agile, and realize new growth opportunities, resulting in intelligent market disruptions. Over the last 25 years, Informatica has helped more than 9,500 customers unleash the power of data. For more information, call +1 650-385-5000 (1-800-653-3871 in the U.S.) or visit https://www.informatica.com/.

Connect with Informatica on LinkedIn, Twitter, and Facebook.

## ABOUT TDWI RESEARCH

TDWI Research provides research and advice for BI professionals worldwide. TDWI Research focuses exclusively on analytics and data management issues and teams up with industry practitioners to deliver both broad and deep understanding of the business and technical issues surrounding the deployment of business intelligence and data management solutions. TDWI Research offers reports, commentary, and inquiry services via a worldwide membership program and provides custom research, benchmarking, and strategic planning services to user and vendor organizations.

## ABOUT THE AUTHOR

**David Stodder** is senior director of TDWI Research for business intelligence. He focuses on providing research-based insights and best practices for organizations implementing BI, analytics, data discovery, data visualization, performance management, and related technologies and methods and has been a thought leader in the field for over two decades. Previously, he headed up his own independent firm and served as vice president and research director with Ventana Research. He was the founding chief editor of *Intelligent Enterprise* where he also served as editorial director for nine years. You can reach him at dstodder@tdwi.org, @dbstodder on Twitter, and on LinkedIn at linkedin.com/in/davidstodder.

## ABOUT TDWI CHECKLIST REPORTS

TDWI Checklist Reports provide an overview of success factors for a specific project in business intelligence, data warehousing, analytics, or a related data management discipline. Companies may use this overview to get organized before beginning a project or to identify goals and areas of improvement for current projects.