



CHECKLIST 2022

# Deriving Maximum Value from the Enterprise Data Marketplace

By James Kobiellus



# Deriving Maximum Value from the Enterprise Data Marketplace

By James Kobiulus

**T**rustworthy data is a foundation of modern business. With better data and analytics at their disposal, organizations can drive better decisions, more agile processes, and smarter applications throughout their operations.

Enterprise data marketplaces provide key self-service channels for accelerating the application of the best data, analytics, machine learning, and other digital assets to business scenarios. These marketplaces enable users to search for, find, and procure trustworthy data sets and such data-intensive assets as structured reports, analytics dashboards, and pretrained machine learning models. They allow employees to easily shop for and request access to the data they need. Typically, such a marketplace provides an experience that simplifies data searches so users can more successfully find data and understand its applicability to their analytical and other business needs.

Five steps for deriving maximum value from the enterprise data marketplace:

- 1 Consolidate enterprise data and analytics assets in a cloud service
- 2 Inventory and understand enterprise data and analytics assets and their usage scenarios
- 3 Catalog enterprise data and analytics assets
- 4 Apply strong governance to enterprise data and analytics assets shared through the marketplace
- 5 Enhance the consumability of marketplace-accessible data and analytics assets

Strong governance is an operational necessity within enterprise data marketplaces. It enables organizations to manage their information assets so they remain correct, consistent, compliant, unbiased, and secure for all intended uses. These same controls help ensure that only the best, most relevant data and analytics can be sourced through cloud-based marketplaces. In this way, data marketplaces can drive faster business decisions by enabling more effective customer engagement, deeper operational efficiencies, faster time to market, and more compelling competitive differentiation.

TDWI research indicates that enterprises are quite aware of the importance of data governance (see Figure 1), of the need to modernize how their organizations manage this function, and of the enabling approaches for making the transition to cloud-native data governance architectures. Though only a little over a third (37 percent) of respondents said that modernizing data governance was a leading priority for their data governance programs, the vast majority (94 percent) considered data governance an opportunity overall due to the fact that it enables compliance while providing internal standards for improving data and its management.<sup>1</sup>

This TDWI Checklist focuses on how enterprises can build an online, self-service data marketplace to ensure that their users are provided with trusted data assets and are able to derive the maximum value from them. To help enterprise data professionals build a high-powered data marketplace, the report provides guidance on several key best practices. The relevant stakeholders to whom these practices are addressed include data managers, data owners, data curators, data stewards, and data product managers.

## 1 Consolidate enterprise data and analytics assets in a cloud service

Data's power grows as more of it is unified into a central resource available to many users, applications, and enterprise decision-making and process automation scenarios.

### Which of the following statements best represents your general assessment of data governance?

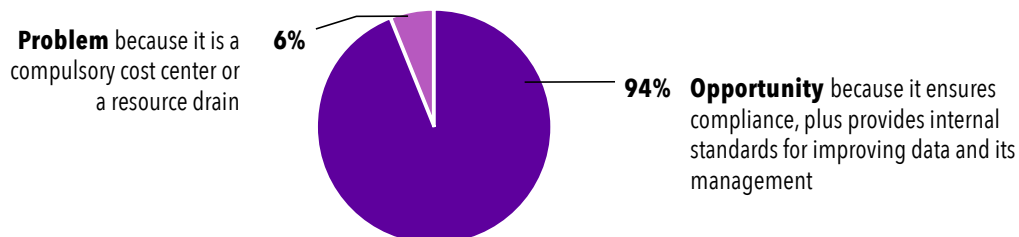


Figure 1. Based on 141 respondents.

<sup>1</sup> See the 2021 TDWI Best Practices Report: Modern Data Governance, online at [tdwi.org/bpreports](https://tdwi.org/bpreports).

The critical first step is for enterprise data managers to move enterprise data and analytics assets into a cloud-native service. As enterprises incorporate data into everything they do, more are unifying their data and analytics assets into an elastic, scalable, high-performance, and fully managed cloud-based data lakehouse. This approach—which converges the data warehouse and the data lake—enables unified access, search, governance, and administration of all assets within an open architecture.

When a data marketplace is deployed on top of this service, it can deliver a unified view of enterprise data, analytics, and other assets that are available for access and use. It can also provide a simplified experience for acquiring, analyzing, manipulating, and otherwise consuming it all. To facilitate this simplification, enterprise data professionals should consider at this stage whether refactoring, normalizing, or modernizing data structures should be undertaken as part of the migration.

Ideally, the cloud service should incorporate:

- **Data storage.** It must unify storage of structured, semistructured, and unstructured data used in BI, AI/ML, and other analytics applications. The service, which may be a centralized or distributed architecture, should support schema-on-read operations on data that is stored in its natural format, usually as object blobs or files. It should also support schema-on-write operations on data that is stored in relational, columnar, key-value, graph, and other structured formats.
- **DataOps pipeline.** The cloud service should integrate with a unified pipeline that processes ETL, data ingestion, data cleansing, and other DataOps workloads. The pipeline capabilities unify data from different sources and then format it all into a common model for storage and processing within the lakehouse infrastructure.
- **Statistical modeling.** The service should incorporate a library of machine learning, deep learning, natural language processing, and other statistical models and algorithms. These perform statistical processing on data that is stored, processed, and managed within the lakehouse infrastructure.
- **Data protection.** The service should integrate with tools for security, backup/recovery, and life cycle management of all data that is stored, managed, and processed in the repository. These should support enforcement of policies on the various operations applied to data, including creation, updating, deletion, storage, retention, classification, tagging, distribution, and archiving.
- **Online transaction processing.** The service should support ACID (atomic, consistent, isolated, durable) transactions on stored data, as well as granular operations on transactional CRUD (create, read, update, delete) tables. It should support robust transactional controls such as snapshot isolation, data versioning, rollback, and schema enforcement.
- **Mixed-workload management.** The service should support allocation, monitoring, and control of a wide range of batch, streaming, and low-latency application workloads. These should be able to execute dynamically across servers, storage devices, and other infrastructure components.

Unification of these disparate capabilities depends on a marketplace architecture that is agnostic to the underlying hardware, software, and cloud service providers.

## 2 Inventory and understand enterprise data and analytics assets and their usage scenarios

To determine which assets to make available through the data marketplace, enterprise data owners must oversee the profiling of both those assets and the usage scenarios to be served.

In addition to preparing an inventory of all data and analytics assets to be offered, enterprises must involve the assets' owners in planning the necessary governance, stewardship, and policy management processes to be instituted in the marketplace.

However, before all that, be sure to clarify whether it's truly your data to provide. If the data under your stewardship is not your intellectual property, you can't legally productize it. To the extent that it isn't, you'll need to negotiate with the owners for the rights. Clearly, you'll need to compensate them under a negotiated price and terms. Don't take any of this for granted. Whoever they are (your existing customers, suppliers, and/or partners), they'll be the foundation of your data monetization value chain.

Also, make sure that you understand who would value the data, why, and how much. Profile the use cases and likely users of the data before you consider providing it through the marketplace. You should also look at what data is used the most, which data is recent and complete, and what data employees would use if it was available and trustworthy. If you don't profile your data assets and likely consumers beforehand, there can be no solid business case for providing your data to other employees through the internal marketplace.

## 3 Catalog enterprise data and analytics assets

In order to facilitate searching, querying, and accessing data assets in the marketplace, enterprise data curators must oversee tagging, classification, and indexing of these assets. The data assets—as well as the tags, glossaries, policies, and other descriptive information associated with them—must be published through a data catalog that is a central infrastructural foundation of the marketplace.

As the foundation of a cloud data marketplace, data catalogs enable data tagging, data classification, and data contextualization with standardized business glossaries. When it provides a platform for comprehensive data curation, the data catalog can help transform data into a strategic business asset.

To derive maximum value from their cloud data assets, modern organizations are using data catalogs to enable several curatorial capabilities:

- Applying curation to new cloud data sources.** Users are bringing a dizzying range of new data types into their cloud data environments. Leveraging self-service access to the data catalog and other platforms that are needed to discover and import this data, users are tapping into such data sources as mobile devices, edge applications, intelligent robots, and other cloud-connected endpoints. As new types of data come into cloud platforms, the data catalog enables better visibility into and understanding of the data so curators can better classify and manage it to meet compliance and other requirements.

- **Enforcing consistent data curation across complex business clouds.** Organizations are confronting the need to implement data governance workflows that orchestrate smoothly across disparate cloud-connected source systems, processing pipelines, stewardship tasks, and production applications. As the complexities of enterprise multicloud and cloud-to-edge environments grow, data professionals are reorganizing their processes to ensure consistent application of data governance and curation policies across the entire enterprise.
- **Integrating curation with cloud-native data platforms.** In the pre-cloud era, many enterprises established data curation and governance infrastructure that runs on legacy, on-premises systems. However, they may now be scrambling to integrate their data cleansing, metadata management, and data lineage analysis capabilities so they integrate tightly with the cloud-native computing environments to which they've migrated most of their core business data and applications. If enterprises fail to do this, they run the risk of introducing inaccuracies, inconsistencies, and noncompliant changes across data assets that are scattered across silos associated with various data stores.

Data catalogs are a pivotal platform in the curatorial process. More enterprises are embracing them to harness insights that would otherwise stay dormant and overlooked. Modern data catalogs support relevant metadata discovery (business, operational, and technical) by scanning each new data set for sensitive data, automating the classification and curation of the data, mapping out lineage between data assets, and tagging data for security and curation purposes.

Data catalogs have become ubiquitous in enterprise data environments, with intelligent metadata,

recommendation engines, and automated task-specific guidance as essential features. These capabilities help organizations manage their growing information assets across more complex hybrid clouds.

In self-service environments, ML-augmented data catalogs enable business users to discover, refine, explore, and apply complex data sets more rapidly and intelligently to diverse applications. When incorporated into a data catalog, ML can help automate data clustering, tagging, and domain/entity recognition. The ML-augmented catalog can intelligently scan data assets from across the enterprise, automatically classify them, and add tags and business context to the metadata.

Data catalogs also generally have built-in, easy-to-use interfaces that enable users to easily search for cataloged data, gain deep visibility into it, and request access to it. This enables self-service discovery of the data best-suited for the task at hand, such as training a machine learning model or curating customer data sets.

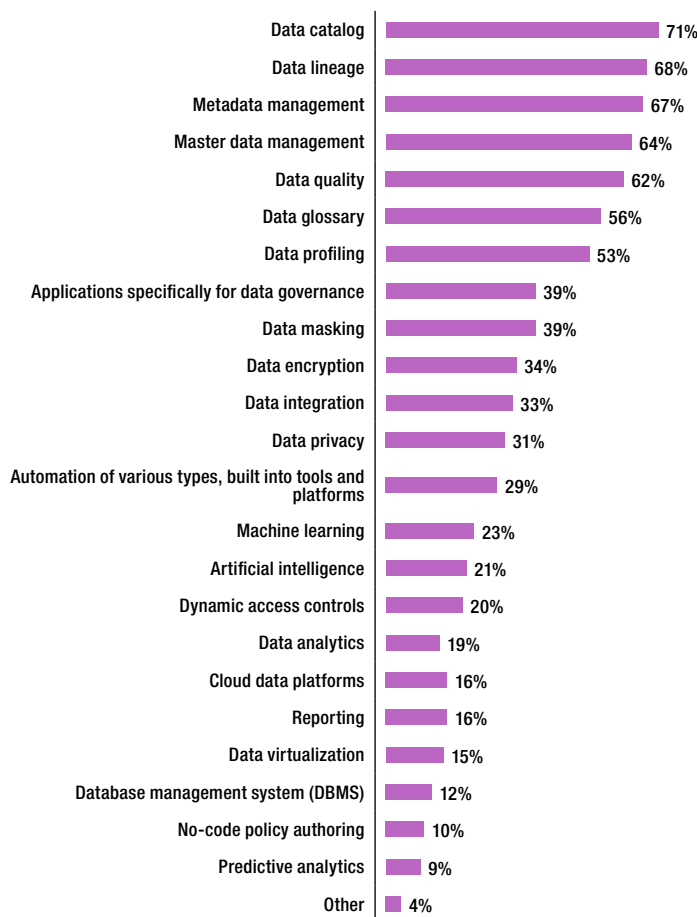
## 4 Apply strong governance to enterprise data and analytics assets shared through the marketplace

To guarantee that assets provided through the marketplace are high quality, relevant, and comprehensible to their intended users, enterprise data stewards must ensure that the assets are managed in accordance with all applicable quality, compliance, security, and other policies.

Enterprises must ensure that the data, machine learning models, and other assets exchanged through the online marketplace are fit for their intended consumers. This requires applying governance consistently and appropriately to all data available through the marketplace. When enforced in internal data marketplaces and throughout enterprise value chains, strong governance and other enterprise controls ensure that data and other assets are trustworthy and relevant. Trustworthy data depends on the work of stewards to ensure that data is kept correct, consistent, complete, compliant, secure, and comprehensible. Relevance depends on curators, who sift through data to tag, contextualize, and recommend it for various business purposes.

To achieve these outcomes, organizations are adopting modern cloud-native platforms, data catalogs, and other tools that automate, accelerate, and augment data governance and curation functions (see Figure 2 for examples). Available solutions enable many types of users to discover and quantify the condition of data sets. Generally, tools allow data stewards and curators to design metrics that represent acceptable parameters for data quality and relevance. Most offer built-in management dashboards so technical and business managers can monitor data’s compliance levels according to thresholds recorded in governance policies.

**Which types of software tools, platforms, and capabilities can automate or assist with data governance? Select one to twelve answers.**



**Figure 2.** Based on 125 respondents. From 2021 TDWI Best Practices Report: Modern Data Governance, online at [tdwi.org/bpreports](https://tdwi.org/bpreports).

These capabilities can automate, accelerate, and augment the processes that enable robust data and model governance. Strong governance must be applied consistently and appropriately to all new assets exchanged through the data marketplace. It generally requires that enterprises ensure that their investments in the enabling infrastructure—including applications for data cataloging, data cleansing, metadata management, business glossary management, and data lineage—integrate fully with the cloud-native computing environments and data platforms that support the marketplace.

Under any data governance program, IT-centric processes—such as data ingestion, profiling, and cleansing—must be aligned with business-level processes such as data tagging, stewardship, and curation. The latter are the true stakeholders of data governance. As data owners and users, they know from experience how data quality, usability, compliance, and trust drive high-priority business outcomes.

## 5 Enhance the consumability of marketplace-accessible data and analytics assets

Data’s value will be squandered if users can’t access it or make sense of what they find through the enterprise data marketplace. Consumability is the key to achieving full value from marketplace-accessible data and analytics assets.

One key to achieving full consumability is building marketplace processes to ensure the data is comprehensible to its intended users. As the complexity of data grows, ontologies become a more important tool for this. Ontologies—and the related notions of glossaries and taxonomies—are principally oriented toward data’s analytical uses within and across disparate data-store implementations. Framed in Resource Description Format and other formats, ontologies are artifacts of analysis that are geared to semantic query and knowledge discovery. They provide views of the concepts, relations, and rules for a particular area of business information, irrespective of how that information may be stored as data.

Another key to consumable data is to increase its searchability. A cloud data catalog must be searchable. To this end, the catalog is often implemented in conjunction with a common business glossary, ontology, and metadata layer. Taken together, these infrastructural layers ensure the discoverability, consistency, and clarity of data’s meaning, no matter what the source, the intermediary processing pipeline, or current location or landing zone it occupies within a marketplace architecture.

In order to ensure that data assets are fully consumable—hence, actionable—enterprise data product managers must also ensure that the data marketplace provides intuitive, self-service access to it all. Users should be empowered with an easy, self-service process for requesting access to data and quickly receiving it. There should also be workflow capabilities that enterprises can use to automate the processes and collaboration necessary to manage and govern the data available through the marketplace.



## Concluding thoughts

Data's value comes from being easily discoverable, searchable, trusted, and consumable. Data marketplaces are important channels for boosting these capabilities so data, machine learning models, and analytics assets can be fully used to grow revenues and otherwise boost the enterprise's bottom line.

Ensuring that only the most trustworthy, relevant, actionable, and consumable data and analytics are accessible through the marketplace requires strong data governance. As discussed in this report, the principal steps for delivering these capabilities are as follows:

- **Move data analytics assets to a cloud service.** Moving important data and analytics assets to the cloud makes it easier to scale to meet needs, allows more people to access the data, and helps provide more agility to the business. When deployed into a cloud data marketplace, the service can deliver a unified view of enterprise data, analytics, and other assets that are available for acquisition. It can also provide a simplified experience for acquiring, analyzing, manipulating, and otherwise consuming it all.
- **Understand enterprise data and analytics assets and consumption use cases.** In addition to bringing data and analytics assets together in the cloud to prepare for a marketplace, the enterprise must engage the owners of these data sets in planning the necessary governance, stewardship, and policy management processes that will form the bedrock of the marketplace's ongoing operations.
- **Catalog enterprise data and analytics assets.** Data catalogs have become ubiquitous in enterprise data environments, with intelligent metadata, recommendation engines, and automated task-specific guidance as essential features.
- **Apply strong governance to enterprise data and analytics assets that are shared through the marketplace.** Enterprises should integrate their investments in the enabling governance infrastructure—including applications for data cataloging, data cleansing, metadata management, business glossary management, and data lineage analysis—with the cloud-native computing environments that support the marketplace.
- **Enhance the consumability of marketplace-accessible data and analytics assets.** As an aid to data's consumability, common business glossary, ontology, and metadata layers are often implemented in conjunction with a data catalog. Taken together, these infrastructural layers ensure the discoverability, consistency, and clarity of data's meaning, no matter what the source, the intermediary processing pipeline, or current location or landing zone it occupies within an online marketplace architecture.

If enterprises establish cloud-based data marketplaces with strong governance controls built in, they'll be taking a key step in maximizing the value from their data and analytics assets. By driving the best-fit data and more predictive models into every decision and process, the enterprise data marketplace can become a pivotal platform for achieving more effective customer engagement, deeper operational efficiencies, faster time-to-market, and other strategic outcomes.

## About our sponsor



[collibra.com](http://collibra.com)

Since 2008, Collibra has been uniting organizations by delivering trusted data for every use, for every user, and across every source. Our Data Intelligence Cloud brings flexible governance, continuous quality, and built-in privacy to all types of data. The *Forbes* Global 2000 relies on Collibra to create the critical alignment that accelerates workflows and delivers better results faster. We have a diverse global footprint, with offices in the U.S., Belgium, Australia, the Czech Republic, France, Poland, and the U.K.

To learn more, visit [collibra.com](http://collibra.com), follow [@Collibra](https://twitter.com/Collibra) on Twitter, or follow us on [LinkedIn](https://www.linkedin.com/company/collibra).

## About the author



**James Kobiellus** is senior director of research for data management at TDWI. He is a veteran industry analyst, consultant, author, speaker, and blogger in analytics and data management. He focuses on advanced analytics, artificial intelligence, and cloud computing. Kobiellus has held positions at Futurum Research, SiliconANGLE Wikibon, Forrester Research, Current Analysis, and the Burton Group and also served as senior program director, product marketing for big data analytics, for IBM, where he was both a subject matter expert and a strategist on thought leadership and content marketing programs targeted at the data science community. You can reach him by email ([jkobiellus@tdwi.org](mailto:jkobiellus@tdwi.org)) on Twitter ([@jameskobiellus](https://twitter.com/jameskobiellus)) and on LinkedIn (<https://www.linkedin.com/in/jameskobiellus/>).

## About TDWI Research

TDWI Research provides industry-leading research and advice for data and analytics professionals worldwide. TDWI Research focuses on modern data management, analytics, and data science approaches and teams up with industry thought leaders and practitioners to deliver both broad and deep understanding of business and technical challenges surrounding the deployment and use of data and analytics. TDWI Research offers in-depth research reports, commentary, assessment, inquiry services, and topical conferences as well as strategic planning services to user and vendor organizations.

## 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.



A Division of 1105 Media  
6300 Canoga Avenue, Suite 1150  
Woodland Hills, CA 91367

[E info@tdwi.org](mailto:info@tdwi.org)

[tdwi.org](http://tdwi.org)

© 2022 by TDWI, a division of 1105 Media, Inc. All rights reserved. Reproductions in whole or part are prohibited except by written permission.

Email requests or feedback to [info@tdwi.org](mailto:info@tdwi.org).

Product and company names mentioned herein may be trademarks and/or registered trademarks of their respective companies. Inclusion of a vendor, product, or service in TDWI research does not constitute an endorsement by TDWI or its management. Sponsorship of a publication should not be construed as an endorsement of the sponsor organization or validation of its claims.