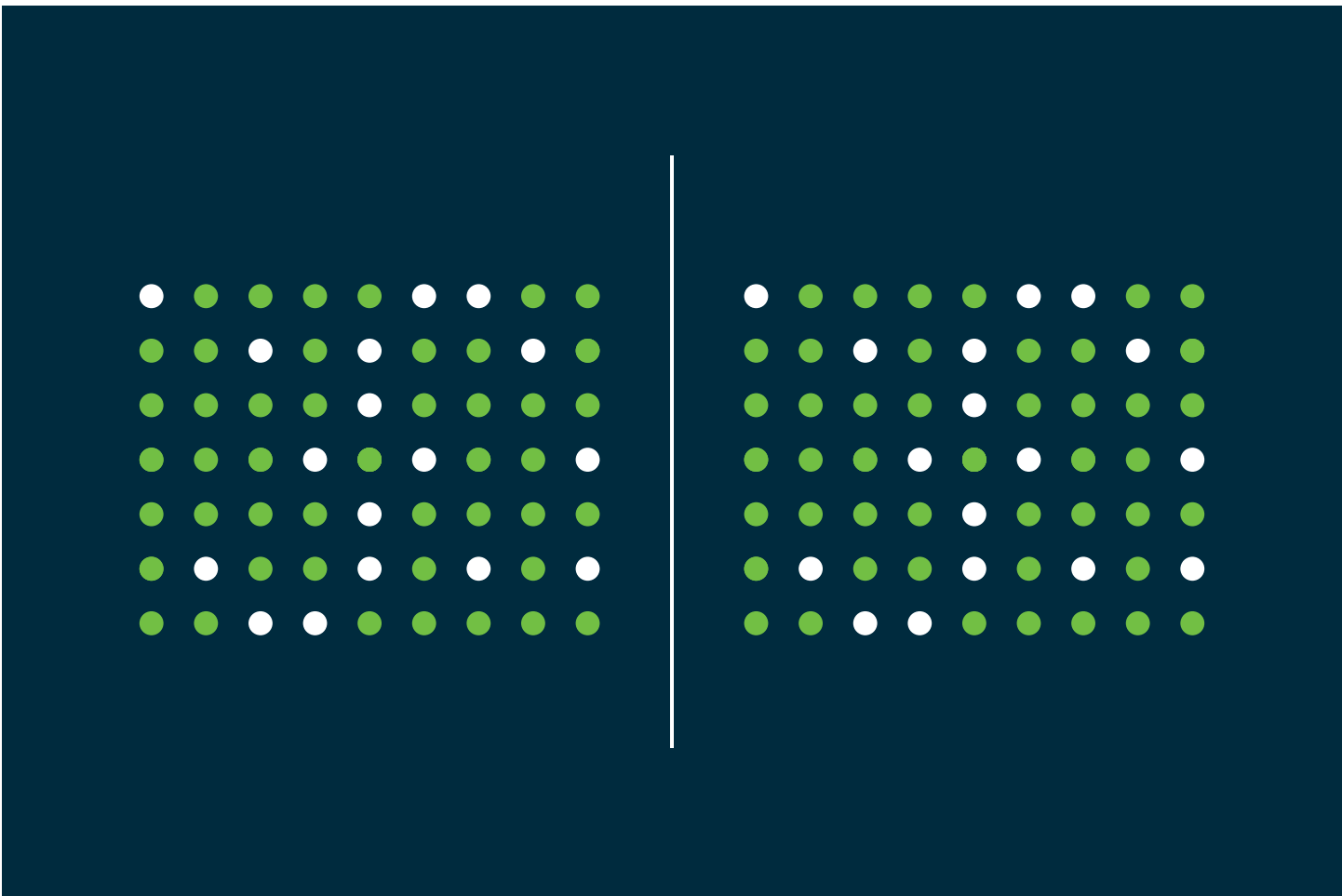# Managing reference data intelligently

Executive summary

Reference data is the foundational context for organizational data, affecting reporting accuracy and data governance. However, reference data management is becoming increasingly complex with exploding data volumes. Organizations often struggle to maintain and take advantage of reference data because of inconsistent usage across the enterprise, a lack of visibility and transparency, and a lack of governance policies to remediate issues, which can result in significant operational risks and ramifications.

Organizations must incorporate reference data management into their data management strategy and when investing in a technology, they must seek a data governance solution with native reference data. Organizations need a solution that can manage complex relationships across the enterprise and enable governance, lineage, stewardship and operational efficiency.

This whitepaper defines the concept of reference data management and illustrates the challenges associated with reference data, the required capabilities for a valuable solution and the benefits of a data governance solution with native reference data management.

What is reference data management?

Most enterprise applications, like ERP and CRM solutions, use code-value sets for categorizing business entities such as products, customers or vendors. For example, an enterprise's CRM system may specify that all customer phone numbers must be ten digits. These defined sets rarely change and data users consistently share them in lookup tables, drop down lists or prefilled forms. They represent reference data.

While reference data may be small in quantity, its quality has an extensive impact. Widely used in transactions and transformations, reference data quality issues can directly affect revenue, operational costs, reporting and governance.

In other words, reference data is the data used to categorize other data. Typical examples are country codes, measurement units, currencies or financial hierarchies. Reference data is often standardized by external bodies for industry-specific use, such as financial product categories or ZIP codes. Organizations may also have their own internal set of reference data that partners, vendors and customers share. Examples include product codes, customer codes or gender codes.

While reference data may be small in quantity, its quality has an extensive impact. Widely used in transactions and transformations, reference data quality issues can directly affect revenue, operational costs, reporting and governance. As a result, efficient reference data management will help an organization maintain the high quality reference data needed to ensure the creation of positive business outcomes.

Why use reference data management?

Reference data affects every part of the organization. It is foundational data that provides context to master transactional and reporting data. When organizations manage and govern reference data appropriately, they can leverage reference data for operational efficiency and accurate reporting.

Data governance and IT teams most often use reference data management to address some of the following use cases:

- **Maintaining code set consistency across enterprise applications.**
  With reference data management, an organization can build a code set inventory for associating with critical data elements and promote shared understanding

- **Improving data quality for efficient transactions and reporting.**
  Inconsistent reference data can cause transactional errors, which may require manual corrections. With effective reference data management, organizations can improve data quality for downstream applications and avoid conflict during transactions and reporting, thus ensuring more accurate data and higher productivity.

- **Optimizing data catalog usage with native reference data management.**
  Organizations need to centralize their data in a catalog with integrated reference data management to optimize the governance and compliance initiatives across the organization

Fresenius Medical Care

FRESENIUS
MEDICAL CARE

Industry
Healthcare

Headquarters
Massachusetts,
USA

**Number of
employees**
128,300

Annual revenue
$17.3 billion

Watch the story

Problem
Fresenius struggled with reporting because it generates incredible volumes of data from more than 347,000 dialysis patients at 4,000 clinics, 45 production sites, and 120,000 employees across the globe

Solution
Data stewards used Collibra to assign characteristics, relations, and groupings and pushed the governed medical information into their data warehouse. The data warehouse performs change data capture and distributes reference data to tables for analytical end user consumption

Result
Analysts can make trusted business decisions because they can examine reports and dashboards that are built on governed and high-quality data

However, while data governance and IT teams may serve as the primary users for reference data management, it is critical to emphasize that departments across the enterprise benefit from reference data management. When communicating to internal partners, data and IT teams must highlight the importance of reference data for the business user. Reference data benefits the business users' day to day and helps ensure:

- **Agreed upon metrics and hierarchies.** Shared understanding across the organization helps build common metrics and hierarchies that can be easily leveraged for efficient operations

- **Clear data controls.** Managing reference data access control helps establish ownership and accountability. It goes a long way in improving the data governance that is essential for trust in data

- **Higher trust in data quality.** Consistent reference data usage can help build a single trusted view of data across the organization

- **Faster delivery of insights from data.** Streamlining and automating reference data management seamlessly provisions reference data to all stakeholders. With access to quality reference data, business users can quickly derive insights from data, powering their business decisions

Furthermore, reference data management is crucial for the entire enterprise because it enhances the quality and usability of data across the organization.

Cigna

Industry
Healthcare

Headquarters
Illinois, USA

Number of employees
74,000

Annual revenue
$153.6 billion

View the webinar

Problem
Stakeholders across the organization didn't trust customer data because of a lack of context around the data and the inability to understand the data

Solution
Cigna's data governance team leveraged Collibra's reference data management to get stakeholders speaking in the right language. Business and technical lineage allowed them to get the context they needed around data. Workflow management, stewardship and the centralized platform enabled cross functional stakeholders to collaborate with each other

Result
Cigna established a culture of data transparency, where individuals can speak in a common language and collaborate around data efficiently

Challenges and consequences

Different enterprise applications use various code values for the same entities or the same code values for different entities. This mismatch creates several challenges for managing reference data effectively.

Insufficient governance

When reference data is fragmented across the organization, teams tend to rely on tribal knowledge. A lack of shared understanding and formalizing common information also leads to a lack of formal accountability. As processes get more ad hoc and offer less visibility, organizations find it difficult to automate these processes, leading to more errors.

Moreover, organizations typically handle data governance activities manually, resulting in slow and untrustworthy change management and fragmented and inconsistent reference data across the enterprise. Without strong data governance and visibility into data usage, organizations struggle to implement reference data management standards. If not addressed, mismatching code sets proliferate and create operational risks and inefficiencies.

Organizations must invest in a data governance solution with native reference data management and additional lineage, stewardship and workflow capabilities to resolve inconsistencies in the data.

Inaccurate reporting and analytics

Inconsistent code values result in inaccurate and untrustworthy reporting and analytics. For example, business analysts examine data and make recommendations for critical decisions using reports based on regions, business units or territories, all of which represent reference data. If each source uses different code value sets, manual intervention becomes necessary to ensure the accuracy of data aggregation and business analytics.

Inefficient operations

Manual reconciliation of inconsistent reference data introduces unnecessary delays in operations. Data stewards must manually address invalid or corrupted reference data and changes across multiple tools over time. Without an authoritative source for key code sets, applications are difficult to scale and deploy on a broader level.

A seemingly minor problem of inconsistent reference data can rapidly snowball into expensive manual work and rework, inaccurate reporting, poor governance and compliance risks. For compliant and competent operations, data teams must prioritize reference data management.

Required capabilities for a valuable reference data management solution

Reference data management should be a central part of an organization's data management and governance strategies. An efficient reference data management solution must manage complex relationships across the enterprise. Organizations must invest in a data governance solution with native reference data management and additional lineage, stewardship and workflow capabilities to resolve inconsistencies in the data.

Data governance

Successful management of reference data needs to be tightly integrated with data governance. Features supporting a complete audit trail and full visibility into processes, ownership and stewardship roles, and a shared understanding of reference data, will empower you to leverage all your data assets.

Data lineage

Data stewards need to map reference data arriving from various sources to a shared code set and link the data to business terms to maintain its current business context as the environment evolves. They also must synchronize reference data with definitions from standardization bodies. It is essential to record the entire data lifecycle, including every transformation at the physical and logical levels.

Workflows

To provide complete visibility into the review processes, the solution must provide collaborative workflows. Clearly defined and automated processes help organizations quickly resolve inconsistencies in reference data management and enable easy access to and use of data.

Stewardship

An efficient reference data management solution should also record reference data across an organizational structure, including communities and domains, that mirrors the business organization. Automating stewardship tasks to manage roles and responsibilities across this structure helps keep pace as the business evolves.

Policy management

For reliable reference data management and good data governance, data stewards need to be able to create, apply and enable adherence to policies for business users. Managing the associated processes and administering granular controls are critical factors for ensuring compliance at the enterprise level.

An effective data governance solution delivers benefits such as lineage, effective stewardship and policy management, which all facilitate more efficient reference data management.

Honeywell

Honeywell

Industry
Manufacturing

Headquarters
North Carolina, USA

Number of employees
113,000

Annual revenue
$36.7 billion

View the webinar

Problem
Honeywell International's data was dispersed across thousands of systems and teams, making it difficult for data consumers to find and access data. Reference data and data governance practices varied from system to system as well, so consumers struggled to trust the data

Solution
Honeywell used Collibra for reference data management, making Collibra its single source of truth. The team accelerated reference data governance with clearly defined responsibilities, ownership and workflows in Collibra

Result
The enhanced reference data and centralized platform enabled data consumers to quickly access data and trust its quality at an enterprise level

Collibra for reference data management

Collibra Data Governance enables teams to efficiently manage reference data and create a common understanding around data. With well-defined, organized and classified data, teams can confidently reconcile data across operational systems for more accurate reporting and analysis.

Get started quickly

Collibra Data Governance facilitates reference data management by making it easy to:

- Ingest reference data code sets into Collibra from a variety of sources, including databases

- Define code patterns to govern reference data at the core

- Build and manage structures over the code sets of reference data

- Map code values between code sets, combining filters and diagrams

- Version code values, code sets and code mappings

Best-in-class capabilities

Collibra allows organizations to manage the entire data lifecycle, enabling them to manage reference data with ease. Collibra's capabilities include:

- **Data governance.** One unified platform for data governance allows teams across the organization to establish a common understanding around data and keep pace with change

- **Business and technical lineage.** Automated end-to-end lineage visualization shows how code sets are built, aggregated, sourced and used, helping build trust in data

- **Collaborative workflows.** Out-of-the-box and customizable workflows and processes facilitate collaboration, close knowledge gaps, and resolve inconsistencies in reference data to ensure productive use of data

- **Data Stewardship.** Automated stewardship tasks help organizations stay current on information assets across departments and subject matter experts, data owners, and data users can collaborate across the organization through role-based dashboards and interactive views

- **Policy Manager.** Simplified processes for creating, reviewing and updating policies so that they are adopted and applied throughout the enterprise

- **Active metadata graph.** Intelligently connecting data elements to business terms, data quality, processes, policies and reports creates dynamic knowledge graphs to leverage your data most effectively

Achieve business objectives

With Collibra Data Governance, organizations can use reference data management to establish:

- **Agreed upon metrics and hierarchies**

- **Clear data controls**

- **Higher trust in data quality**

- **Faster delivery of insights from data**

Furthermore, they can leverage the power of the Collibra Platform to:

- **Ensure governance and compliance.** Leverage the comprehensive platform with automated data lineage to simplify governance and compliance

- **Empower data citizens.** With a shared understanding of reference data, empower data citizens for faster access to trusted data and more reliable business analytics

- **Deliver business agility.** Centrally manage your reference data, enable enterprise-wide access, and leverage cloud scalability to quickly respond to business growth

- **Improve operational efficiency.** Consistent use of reference data across applications and deliver more accurate reports, reducing rework or manual reconciliation

Conclusion

Reference data is the foundational context for organizational data, affecting reporting accuracy and data governance. Traditionally handled by data stewards, reference data management is becoming increasingly complex with exploding data volumes. Maintaining reference data quality is critical as it gets shared across and outside the enterprise. Challenges in shared understanding, inconsistent enterprise-wide usage, visibility, reconciliation and accurate reporting lead to operational risks.

Effective reference data management must be a part of your data management strategy and native to your data governance solution, managing complex relationships across the enterprise and delivering assured lineage, effective stewardship, and operational efficiency.

If you are interested in learning more, please visit our website and request a demo at collibra.com/request-a-demo

---

Collibra

For additional questions, contact us at:

United States
+1 646 893 3042

United Kingdom
+44 203 695 6965

All other locations
+32 2 894 79 60

By email
info@collibra.com