

How to Manage Reference Data

Your guide to planning and implementing a reference data management strategy to improve IT efficiency

About Informatica

Digital transformation changes expectations: better service, faster delivery, with less cost. Businesses must transform to stay relevant and data holds the answers.

As the world's leader in Enterprise Cloud Data Management, we're prepared to help you intelligently lead—in any sector, category or niche. Informatica provides you with the foresight to become more agile, realize new growth opportunities or create new inventions. With 100% focus on everything data, we offer the versatility needed to succeed.

We invite you to explore all that Informatica has to offer—and unleash the power of data to drive your next intelligent disruption.

Table of Contents

Executive Summary.....	4
What is Reference Data?	4
Challenges in Managing Reference Data	5
Best Practices for Implementing a Reference Data Project.....	6
Capabilities to Look for in a Reference Data Management Solution.....	8
How Reference Data Management Benefits Organizations.....	9
Reference Data Management in Action.....	9
Conclusion	10

Executive Summary

IT teams are constantly looking for ways to improve efficiency. To that end, many are streamlining data management by using a master data management (MDM) strategy to provide a single, consistent view of business data across the enterprise. MDM removes disparate, duplicate, and conflicting information to simplify reporting and business intelligence and enable better business decisions.

Yet a key area of data management complexity overlooked by many is the management of reference data. While the lookup tables that many organizations use to give users consistent values to choose when entering data such as dates, countries, gender, or industry into a database or application may appear to be simple, managing this information across a large enterprise is deceptively complicated.

This white paper gives you an overview of what reference data is, the challenges organizations face in managing it, best practices and capabilities necessary in a solution, and how the right tool can remove operational overhead so that organizations can boost productivity.

What is Reference Data?

If you were to broadly characterize all the key business data in your enterprise, you'd find four types:

1. **Transactional data** that comes from applications and operational datastores
2. **Master data** that provides a consistent definition of business entities across multiple IT systems in the enterprise
3. **Metadata** that describes the structure of an entity
4. **Reference data**

Known by many different names—lookup values, lists, code data tables, domain values, drop-down values, or enumeration values—reference data describes the code and value pairs built into code tables that provide a list of allowed values for a given field in an application or database. Most enterprise applications use reference data to classify and categorize product, customer, and supplier information. Examples include state codes, country codes, gender codes, and marital status codes. Many international and industry-specific standards define reference data standards; for example, the North American Industry Classification System (NAICS) provides codes that classify business establishments.

While the amount of reference data in most organizations is relatively small, this type of data can be found everywhere—in data warehouses, extract-transform-load (ETL) mappings, spreadsheets, relational databases, and more.

Challenges in Managing Reference Data

Among the biggest obstacles that organizations face in working with reference data is that different applications may rely on different variations of code sets to define the same thing. For example, their CRM and ERP solutions might use different country code types. As a result, integrating data across applications requires organizations to translate between the different code table representations to categorize data in a consistent manner. Today, many organizations reconcile this data as well as track any changes to it using ad hoc and error-prone manual methods. Others are unaware they need consistent reference data because they're accustomed to tolerating inconsistencies.

Errors in the reference data can have major business impacts. Reports meant for internal or external audiences that pull data from multiple applications and departments can be inaccurate. When heavily regulated industries, such as financial services or healthcare, provide inaccurate reports to regulators they can be subject to heavy fines and reputational damage. Poor quality data can also cause major issues in downstream applications.

As they synchronize values and manage changes across the enterprise, organizations face the following challenges:

Managing Multiple Codes and Their Mapping

Reference data coding systems can be quite complex. Take the case of country codes. Several systems represent countries and dependent areas, and some of these schemes offer multiple coding options. The best-known method, ISO 3166-1, offers three sets of codes, as illustrated here:

Alpha-2	Alpha-3	Numeric
US	USA	840

Industry codes are another case in point. NAICS replaced the Standard Industrial Classification (SIC) codes in 1997. Yet some regulatory reports continue to demand SIC codes. In addition, the European community uses the Nomenclature Statistique des Activités économiques dans la Communauté Européenne (NACE) classification for the same purpose but employs different codes. If you are a global organization doing business in different countries, you must map these codes correctly to be compliant in your reporting.

Data Integration

One of the primary reasons for data integration project failures is poor reference data mapping. Most organizations use different codes to represent reference data in different enterprise applications because application owners choose the codes and descriptions that best suit them. An online bill payment service may be represented in the billing system as OBS, while the customer information file may call it BILLPAY. To integrate this data, organizations must map OBS in the billing system, and BILLPAY in the customer information file to one single service type in the MDM system in order to create a single view of customers and their contracts. This challenge is compounded by the fact that data integration projects often deal with thousands of code value pairs.

Inconsistent Enterprise-Wide Representation

Most organizations do not represent reference data consistently at an enterprise level. As mentioned previously, reference data is often application specific, which leads to siloed data management. Moreover, whether the codes are internal or external, they tend to change over time. When codes change, the inability to maintain reference data centrally drives significant effort and cost for the enterprise. Mergers and acquisitions, rapid growth in data volume, complexity of reference data, and lack of governance and absence of enterprise-wide single view further increase operational risk.

Best Practices for Implementing a Reference Data Solution

Organizations need a robust reference data management strategy to avoid major data quality issues that affect reporting, regulatory compliance, transactions, and ultimately their ability to understand their data. Best Practices for planning and implementing a reference data strategy require three phases: Consolidation, Management, and Distribution. Each phase consists of the following tasks.

Consolidation

The first step is to create a golden copy of reference data by consolidating all of the organization's reference data into a central repository. Organizations can use the repository to manage reference data across the enterprise to achieve standardization, quality, and operational efficiency. Tasks include:

- Discovery. Create a list of all reference data sets necessary for smooth functioning of the business. Base this list on what is important for the enterprise as a whole rather than the demands of individual applications.
- Categorize external and internal data sets, including relevant industry standard codes and the codes the organization manages internally.
- Create a list of all internal applications and systems, and then discover, select, and understand external providers that create reference data.
- Profile data sets to gauge quality. Given the importance of data quality, this step is crucial (and often discounted).
- Analyze data and create rules around data sets for:
 - Required attributes, such as allowed length, data types, sub types, effective and expiry dates, and so on.
 - Cleansing and transforming data to maintain quality (No UNKNOWN, DUMMY, -1, -911 codes).
 - Find efficient ways to extract data so it can be loaded to a reference data hub while honoring business rules.
 - Design and standardize extract process in order to easily integrate new internal and external reference data sources.
 - Design the solution to add and update mapping and hierarchical representation of data sets.

Management

Once they have a golden version of their reference data, organizations must manage it effectively on an ongoing basis. Data stewards, business analysts, and subject matter experts should take advantage of their extensive knowledge of the organization's business process to:

- Ensure that reference data is clean, consistent, accurate, up-to-date, and sharable. For example, they should keep external reference data current by ensuring that changes are detected and assimilated as quickly as possible and that changes in operational environments that lead to rapid change in the reference data in application systems don't create problems at the enterprise level.
- Build rules to identify duplicate codes and reference data sets.
- Detect overlapping codes across different source systems and create a single version of reference data sets.
- Manage the mapping and relationships between reference data sets. Often referred to as transcoding, this step helps map a single representation of codes into multiple application-specific representations.
- Maintain hierarchical representation of reference data sets.
- Continuously audit reference data sets to capture what system or user changed the data, what was changed and when.
- Build security around reference data with attribute-level permissions for view/add/update/delete operations.

Distribution:

The "golden copy" of any master data is meant to support all downstream systems. Organizations must therefore seamlessly integrate their master reference data with downstream systems.

The following tasks will help organizations distribute reference data effectively:

- Create an extensible service layer to expose reference data in the form of web services.
- Create an infrastructure that can deliver reference data to requesting applications in sub seconds.
- Create a flexible mechanism to export and transform reference data for use by downstream applications.
- Support incompatible consuming applications (if any) by providing an ad-hoc reference data dump.
- Support the export of delta changes (incremental or changes since last export).

Capabilities to Look for in a Reference Data Management Solution

Organizations looking to implement a reference data management strategy will need the right tools to simplify reference data management, consolidate reference data across the enterprise, standardize the data, and create a single source of trusted information. Here's what to look for in a solution.

End-to-End Data Management

End-to-end data management is essential. The solution should make it easy to onboard data from a wide range of internal and external data sources; provide ongoing data management to ensure data quality, standardization, data enrichment and matching; and distribute data through seamless integration with all downstream systems.

Business User Self-Service

Business users should be able to easily use reference data in the most meaningful way with an intuitive, business-user-focused UI. Power users and business users should be empowered to define reference data models, manage the data within them, as well as govern ownership from within the same UI without the need for IT services.

Workflows and Task Management

Workflow capabilities should be available to help users define logical flow with multiple steps for the creation, modification, and deletion of reference data. Each step should be assigned to users or roles that will have permission to review, approve, or reject the tasks. Task management should enable users to see actions assigned to them and act on them. Workflows should also support exceptions such as timeout, reassign, and escalations.

Hierarchy Management

The solution should allow users to define complex, multi-level hierarchies while specifying validation rules that determine the desired number and types of attributes at each level. Users should be able to easily understand whether something has been added, removed, edited, or moved to a different parent in the hierarchy and where, as well as see which nodes are impacted, access the changes, and decide whether the changes should be approved or rejected.

Crosswalk Mapping

Crosswalk mapping enables organizations to easily see how one code scheme translates to a different scheme. The solution should make it easy to not only manage the standardized reference data sets but also their translation mappings and crosswalk relationships with application-specific versions of the same data set.

Imports and Exports

You need to be able to import external or internal reference data from a spreadsheet or flat file into a central repository. Such import must include capabilities for extraction, filtering, transformation, and enrichment—preferably driven by metadata. You should also be able to export data to a spreadsheet or flat file in order to seamlessly distribute data to downstream systems.

Cloud Scale and Agility

Organizations are increasingly looking to take advantage of cloud capabilities, such as enterprise scalability, rapid time to value, and automatic upgrades and fixes for the latest enhancements, features, and functionality. The solution should therefore enable organizations to centrally manage complex reference data information in a cloud environment.

How Reference Data Management Benefits Organizations

Organizations that employ the right reference data management solution can empower business users, simplify governance and compliance, remove operational overhead, and accelerate the time-to-value for their data initiatives.

Empower Business Users

A user-friendly, self-service reference data management tool that provides a 360-degree version of the truth empowers business users to manage the complex lifecycle of reference data sets.

Simplify Governance and Compliance

A holistic approach to reference data management ensures reference data is clean and results in accurate data interchange, publishing and entitlements, and external reporting for purposes that include regulatory compliance.

Remove Operational Overhead

Instead of manually reconciling data from different sources each time, employees and executives can now automate the consumption of data, improving employee productivity.

Accelerate Time to Value from Data Initiatives

A cloud-based data reference management solution provides faster time to value compared to other solutions available in the market, because users can provision a reference data management system without worrying about hardware and software procurement and setup.

Reference Data Management in Action

Financial services and healthcare are two industries where proper reference data management is essential to delivering high-quality services and ensuring regulatory compliance.

Financial Services

Increased global regulatory pressure is making financial institutions realize the value of putting a data strategy in place. Some of the common reference data challenges financial institutions face include an exponential increase in asset classes, new securities, the need to manage multiple Security Masters, the prevalence of different identifiers in use by front and mid-offices, including Committee of Uniform Securities Identification Procedures (CUSIP), International Securities Identification Number (ISIN), Stock Exchange Daily Official List (SEDOL), and internal identifiers.

Reference data management enables financial services companies to centralize reference data codes. Large banks with operations in multiple countries can manage individual national code sets and reconcile the differences across countries. By managing reference data, financial services companies can facilitate the seamless flow of clean, consolidated, and accurate data throughout the enterprise and streamline the value chain, manage risk efficiently, improve customer loyalty, and support sound corporate governance.

Healthcare

Healthcare organizations manage a variety of industry-specific codes while providing services to patients, doctors, hospitals, providers, and insurance companies. To facilitate a common understanding and communication across these entities, healthcare organizations need to use industry-standard codes such as ICD-9, ICD-10, Snomed, LOINC, RxNorm, DRG, Chart of Accounts, as well as provider specialty, diagnosis descriptions, and medical terms.

Reference data standardization enables healthcare organizations to deliver high-quality services to patients and ensure correct billing and payment processing. Improved reference data management enables hospitals to improve claims processing to enhance the patient experience, capture complete clinical intent to ensure optimal reimbursement, improve reporting for research and clinical trials, and remove operational overhead to cut costs and save time.

Conclusion

Managing reference data across the enterprise can be harder than it seems. Differences in coding schemes across the enterprise make it difficult to integrate reference data—and is one of the primary reasons why data integration projects fail. Errors in reference data can cause errors in reports pulled from data across the enterprise. A comprehensive, cloud-based data reference management [solution](#) that simplifies best practice processes of consolidation, management, and distribution can streamline and improve the reference data consolidation process. As a result, organizations can empower business users to manage the complex lifecycle of reference data sets, simplify governance and compliance by ensuring that reports have clean data, remove the operational overhead previously required from manual reconciliation, and accelerate time to value from data initiatives.

