

Reference Data Management

IBM Redbooks Solution Guide

Reference data refers to data that is used to categorize other data within enterprise applications and databases. Reference data includes the lookup table and code table data that is found in virtually every enterprise application, such as country codes, currency codes, and industry codes.

Reference data is distinct from transactional data and master data. *Transactional data* is the data that is produced by transactions within applications; *master data* is the data that represents the key business entities that participate within transactions. Reference data is also distinct from *metadata*, which describes the structure of an entity. Transactional data, master data, and reference data, when combined, comprise the key business data within an enterprise.

Most enterprise applications contain reference data, built into code tables, to classify and categorize product information, customer information, and transaction data. Reference data changes relatively infrequently, but it does change over time, and given its ubiquity, synchronizing reference data values and managing changes across the enterprise is a major challenge (Figure 1).

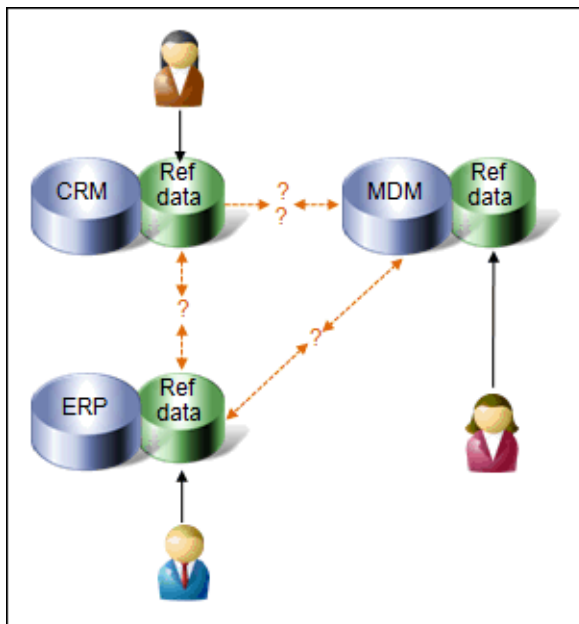


Figure 1. Reference data is found everywhere

Did you know?

Reference data has been part of enterprise applications from the beginning of the modern computing era. However, despite this fact and the fact that it constitutes a fundamental class of enterprise data, there is relatively little focus on reference data and its importance as an enterprise data asset.

Ad hoc management of reference data without a formal governance policy can create significant operational risk. For many enterprises, reference data is a major contributor to enterprise data quality problems and has a high support cost. The demands of complying with national and international industry regulations are causing companies to rethink reference data management, and compelling enterprises to manage and control their reference data by using sound data governance principles. IBM® InfoSphere® Master Data Management Reference Data Management Hub (InfoSphere MDM Ref DM Hub) is an ideal solution for reference data management.

Business value

Today, many companies have no centralized enterprise governance over reference data; critical reference data is managed using spreadsheets and manual, ad hoc methods. The difficulty of managing change across the complex web of reference data variations is not systematically addressed; errors in reference data mappings and inconsistencies are accepted and tolerated as an everyday reality. Reference data variations and inconsistencies can be a major source of data quality issues within the enterprise and cause business losses through system downtime, incorrect transactions, and incorrect reports.

InfoSphere MDM Ref DM Hub provides a robust solution for centralized management, stewardship, and distribution of enterprise reference data. It supports defining and managing reference data as an enterprise standard. It also supports maintaining mappings between the various application-specific representations of reference data that are used within the enterprise. The InfoSphere MDM Ref DM Hub supports formal governance of reference data, putting management of the reference data in the hands of the business users, reducing the burden on IT, and improving the overall quality of data used across the organization.

By centrally managing and distributing reference data within the enterprise with InfoSphere MDM Ref DM Hub, organizations can realize the following benefits:

- Reduce the risk of incorrect values
 - Automates publishing to consuming systems.
 - Provides controls and audit of changes made.
 - Provides automation for manually intensive processes.
- Reduce IT costs
 - Supports removal of redundant systems used for storing reference data.
 - Reduces integration cost for point-to-point application management for reference data.
 - Reduces cost to modify reference data fields.
- Improve business intelligence (BI) reporting and regulatory compliance
 - Automates mapping and trans-coding for reference data being provided to the data warehouse.
 - Improves warehouse data quality for BI reporting and reporting for regulatory compliance.
- Agility in modifying value quickly
 - Automates process to streamline update of consuming systems
 - Reduces authoring time and effort for making changes to reference data
 - Reduces release cycles for data integration project initiatives

Solution overview

The IBM InfoSphere Master Data Management Reference Data Management Hub was released as a separately chargeable component under the IBM Master Data Management Product ID (PID) in July 2012. The hub was developed as a stand-alone reference data domain on the InfoSphere MDM Custom Domain Hub Platform, which itself is the foundation for the InfoSphere MDM Advanced Edition. The InfoSphere MDM Ref DM Hub implements its own specialized domain model specifically for reference data, that is, reference data is supported as a first-class domain entity. The InfoSphere MDM Ref DM Hub includes a dedicated stewardship interface that is designed for managing reference data. The web-based user interface (UI) runs in the browser and no special code is required on the client. The UI is designed for business users, with intuitive and familiar navigation and controls. A flexible data model supports dynamic modeling of reference data properties through the UI, ensuring a quick implementation and minimizing the need for IT involvement on an ongoing basis.

InfoSphere MDM Ref DM Hub is designed as a ready-to-run application. It can be quick to install, easy to use and understand, and delivers real value for immediate use without requiring extensive customization. The key functions include the following items:

- Role-based user interface with security and access control, including integration with LDAP
- Management of reference data sets and values
- Management of mappings and relationships between reference data sets
- Importing and exporting of reference data in CSV and XML format through both batch and user interface
- Versioning support for reference data sets and mappings
- Change process controlled through configurable lifecycle management
- Hierarchy management

InfoSphere MDM Ref DM Hub is built on the proven InfoSphere MDM platform and delivers a master data management approach to managing enterprise reference data. It helps to reduce business risk, improve enterprise data quality, and enhance operational efficiency. InfoSphere MDM Ref DM Hub is based on a three-tiered component architecture, comprising a client and a server application interacting with a back-end database that hosts the application-specific data and required metadata. Figure 2 depicts a high-level component architecture of InfoSphere MDM Ref DM Hub.

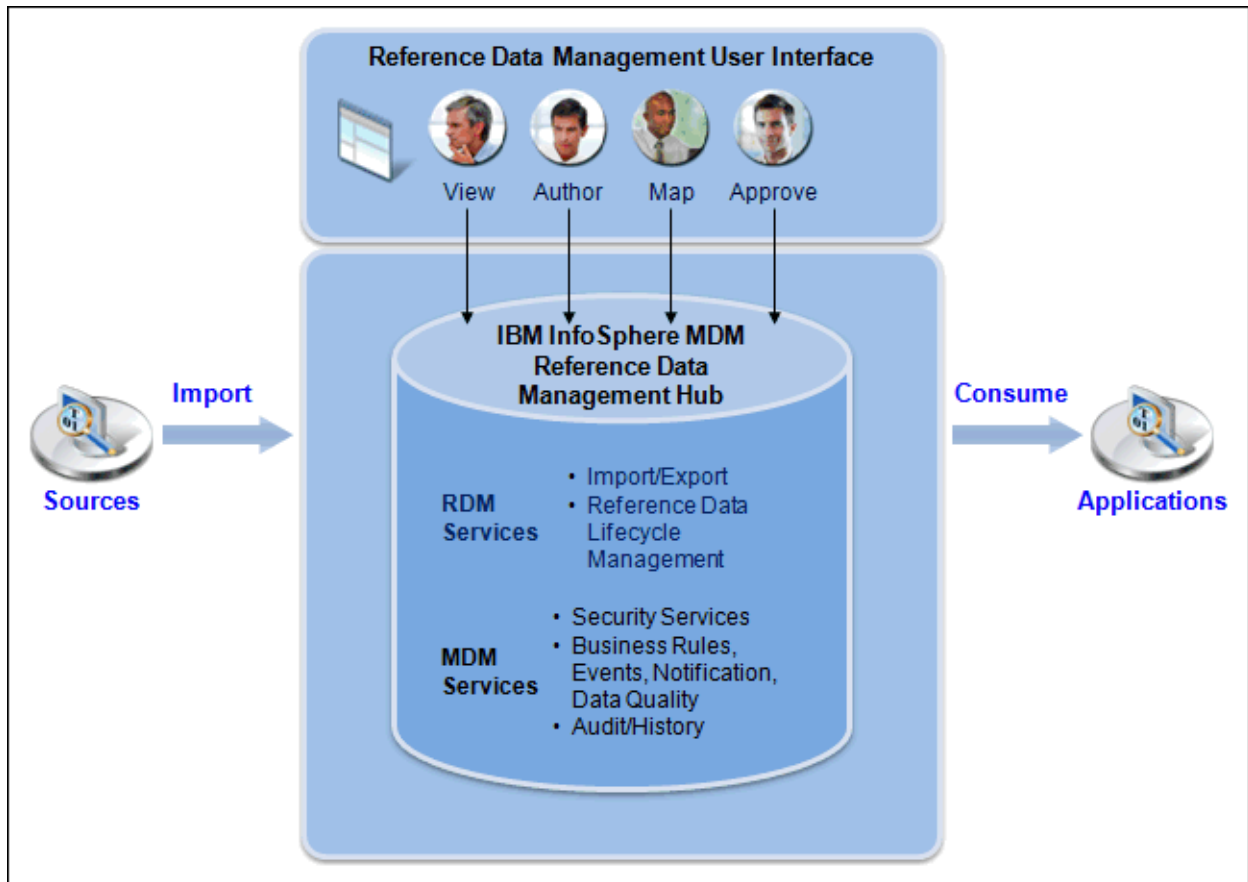


Figure 2. InfoSphere MDM Ref DM Hub logical architecture

The InfoSphere MDM Ref DM Hub user interface is a web application UI that supports collaborative authoring of reference data. Reference Data Stewards use the RDM web UI for the importing, managing, and publishing of reference data sets. The role-based UI allows a stewardship team to view, author, map, and approve reference data sets within a central repository. With this approach, reference data sets can be created and managed in a controlled manner. User actions on the web UI trigger requests, which are handled by appropriate service controllers present in the Representational State Transfer (REST) layer. The REST layer services invoke the server-side transactions to manage create, read, update, and delete (CRUD) procedures on RDM database.

The server-side is implemented on the proven InfoSphere Master Data Management Custom Domain Hub engine (the same engine that powers InfoSphere MDM Server and InfoSphere MDM Advanced Edition).

The reference data domain model elevates reference data to be a first class domain entity within MDM. By implementing the InfoSphere MDM Ref DM Hub as a new domain on the InfoSphere MDM platform, the InfoSphere MDM Ref DM Hub benefits from a wide range of base services and ready-to-use frameworks that InfoSphere MDM provides, such as business rules, event notification, data quality, and audit history. In addition, several reference data management specific services are implemented to achieve key functionality, such as import and export, reference data set lifecycle management, transcoding, distribution, and versioning.

The client and server enterprise archives reside in an IBM WebSphere® Application Server instance. The currently supported databases are IBM DB2® and Oracle.

Solution architecture

The InfoSphere MDM Ref DM Hub serves as an integration, management, and distribution point in the enterprise for reference data sets, maps between reference data sets, and hierarchies over reference data.

Figure 3 shows an overall view of where RDM fits into an enterprise reference architecture.

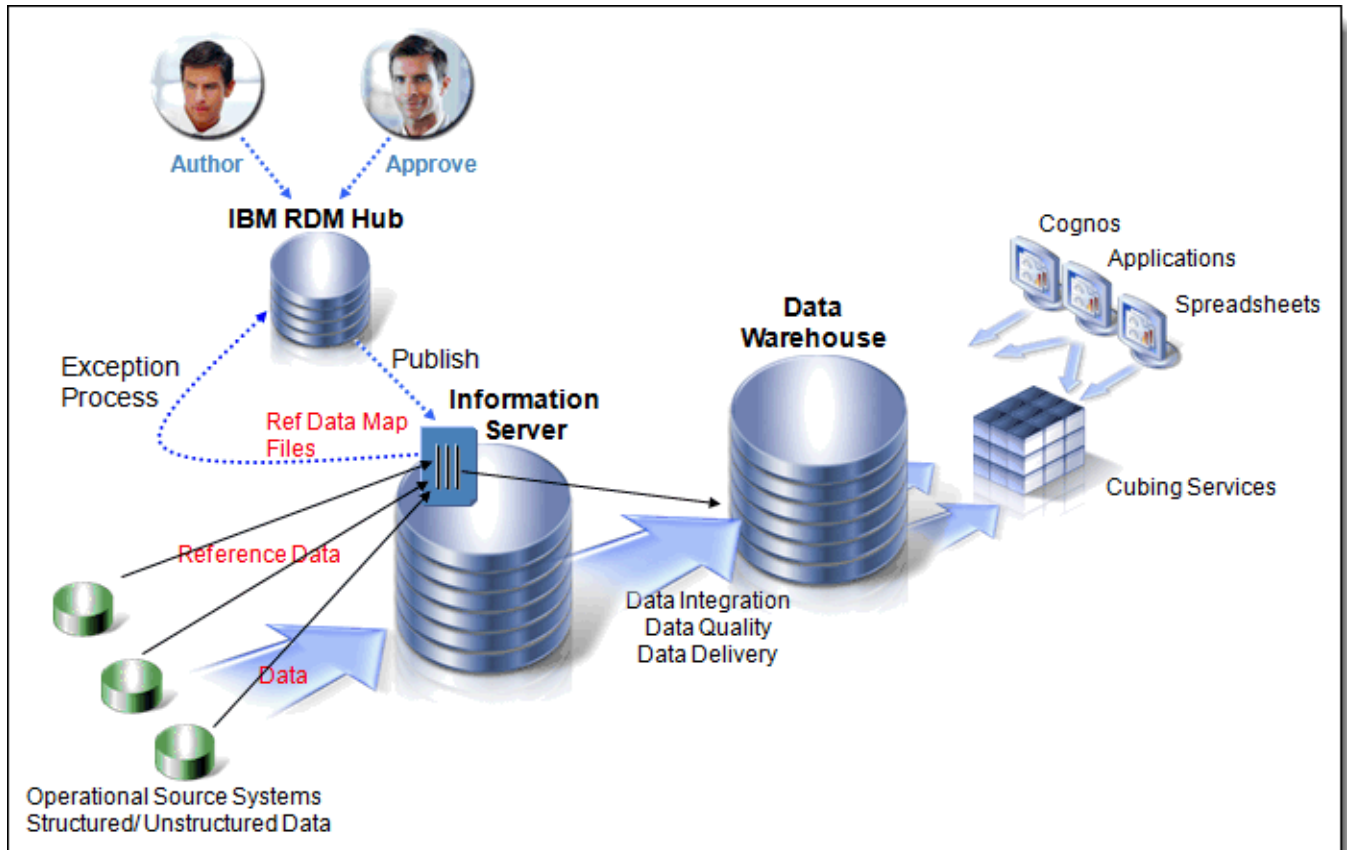


Figure 3. Reference Data Management Hub in an enterprise architecture

Reference data sets and hierarchies that InfoSphere MDM Ref DM Hub provides are consumed by enterprise information systems (such as InfoSphere MDM, SAP, data warehouses, business intelligence systems, and so on) to ensure that business objects are accurately and consistently described across the enterprise. Reference data maps are used by data integration layers (such as IBM InfoSphere Information Server, or an enterprise service bus) to map reference data values between source systems and target systems.

Stewardship of reference data consists of the following tasks:

- Importing or authoring reference data from source systems
- Managing changes to the reference data in an orderly fashion
- Distributing the reference data to downstream systems

All of the InfoSphere MDM Ref DM Hub objects can be accessed through the web services. Additionally, a REST layer on top of the InfoSphere MDM Ref DM Hub web services is used by the InfoSphere MDM Ref DM Hub user interface. The InfoSphere MDM Ref DM Hub UI is the standard way to manage the InfoSphere MDM Ref DM Hub and to govern reference data.

There are three core reference data domain objects: sets, maps, and hierarchies (Figure 4). Each object supports the standard CRUD operations. Each object also supports the notion of a validity period (the time when an object becomes active, and the time when an object is no longer valid). Sets and maps also support extensibility and lifecycle.

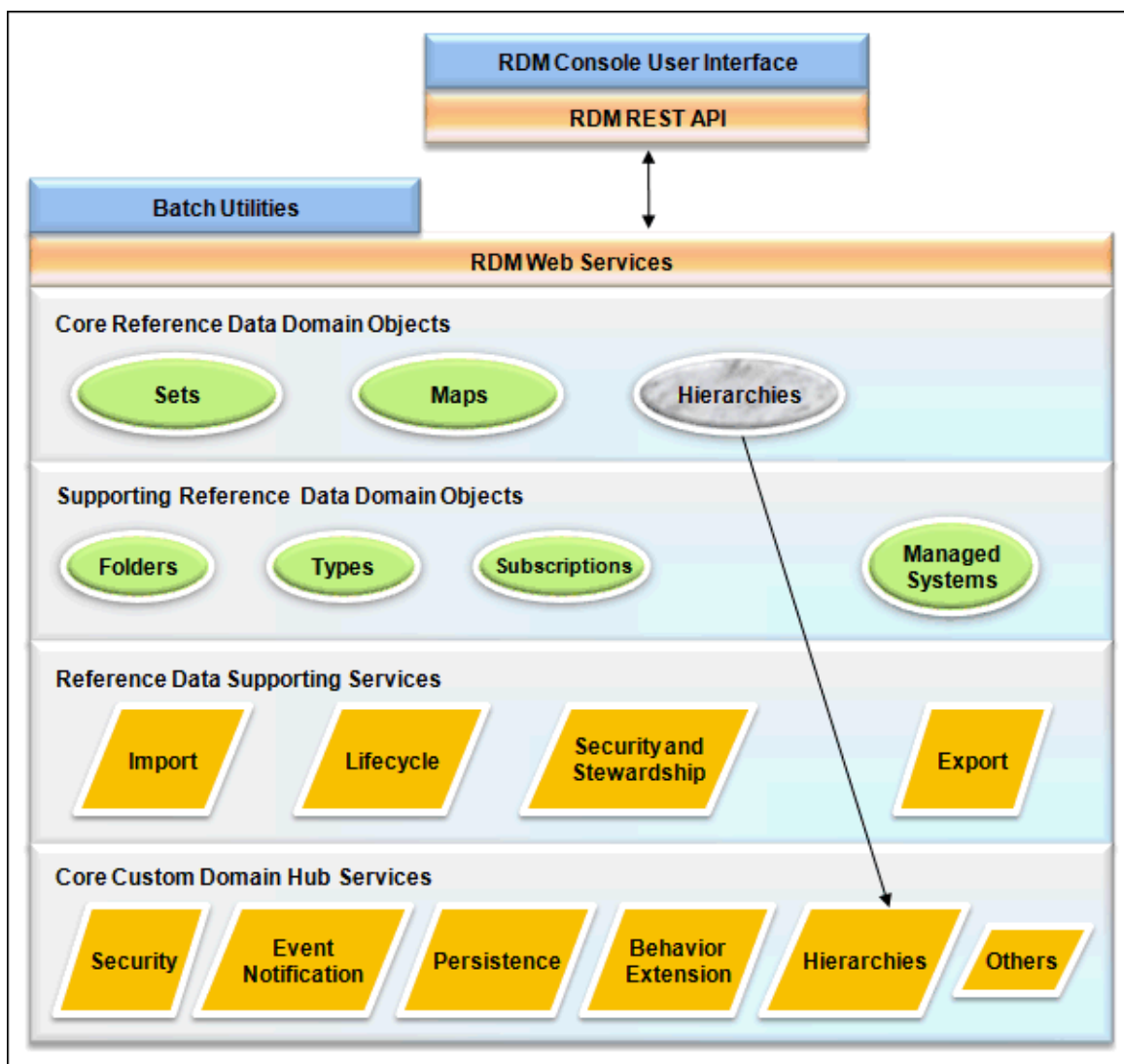


Figure 4. Reference data domain objects

In addition to the core reference data domain objects, there are some supporting objects utilized in the reference data domain. These objects range from providing underlying support for core objects (types), to providing objects that link to organizational containers (folders), and finally, to a set of objects that are linked to the core objects (subscriptions, managed systems) as part of the reference data ecosystem.

Figure 5 illustrates an InfoSphere MDM Ref DM Hub data model with various reference data objects.

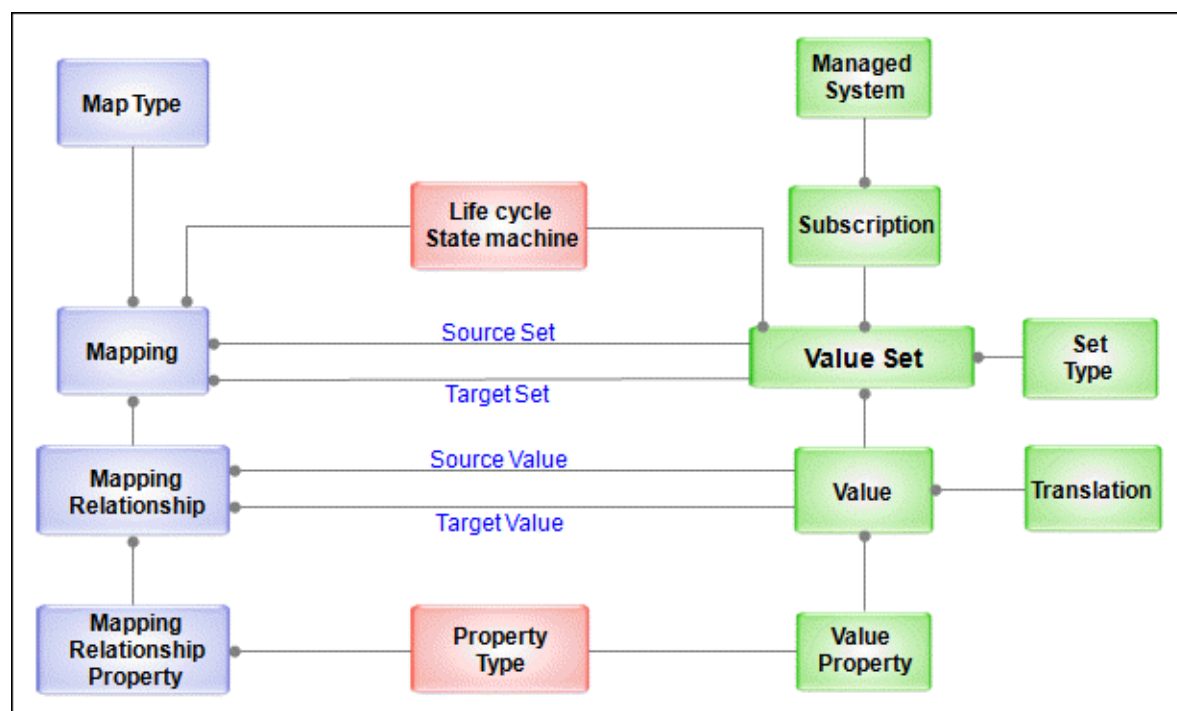


Figure 5. InfoSphere MDM Ref DM Hub data model

Usage scenarios

Banks, insurers, and other financial services organizations are under pressure to comply with increasingly rigorous standards for data accuracy, security, and provenance. At the same time, their executives require more accurate business intelligence (BI) and reporting to drive key decisions. InfoSphere MDM Reference Data Management Hub is an ideal solution for these needs.

Examples of drivers for managing reference data include:

- **Banking.** Banks face the challenge of distributing reference data changes to downstream applications and then assessing what applications are impacted by a reference data change. The subscription model of InfoSphere Reference Data Management Hub makes it possible to assess what applications are impacted by a particular reference data set change.

In European countries, you can choose a version of the NACE industry classification codes unique to the country as the national standard. Banks with operations in multiple countries must manage individual national NACE code sets and reconcile the differences across countries. The InfoSphere MDM Ref DM Hub solution allows different versions of the NACE codes to be managed from a central point, thereby simplifying the creation of mappings among different versions and supporting transcoding values across the data sets.

- **Healthcare.** Healthcare is a highly integrated industry with many different organizations working together to provide services to patients. Doctors, hospitals, plan providers, and insurance companies must all work together and communicate effectively in order to provide a high quality of service to patients and ensure correct billing and payment processes. Code Classifications such as ICD-9 and ICD-10 identify specific diagnosis and clinical procedures on claims, encounter forms, and other electronic transactions. InfoSphere MDM Ref DM Hub supports central management of healthcare code sets and the mappings between them so that changes to these code sets can be managed over

time. In addition, proactively managing the mappings between codes within applications and transactions, and the reference data codes used within the data warehouse, can result in greatly enhanced BI and statistical reporting, vital in order to compete effectively in the highly competitive healthcare space.

- *Cross-industry*: Loading data into a data warehouse or an MDM hub typically requires reconciling reference data from multiple sources. InfoSphere MDM Ref DM Hub supports the loading and publishing of reference data map files for batch and load jobs.

Integration

One of the critical functions of InfoSphere MDM Ref DM Hub is to interact with the reference data found in other enterprise systems. InfoSphere MDM Ref DM Hub can obtain reference data from key enterprise information systems and the managed reference data objects are then used in conjunction with those enterprise information systems.

The following are a few enterprise system integration examples of InfoSphere MDM Ref DM Hub:

- **Master Data Management**
The InfoSphere Master Data Management (MDM) Custom Domain Hub hosts the reference data domain, but the core server itself has some other interaction patterns with MDM. In particular, InfoSphere MDM Ref DM Hub interacts with MDM for the following key functions:
 - Management of MDM code tables
 - Management of maps used to import data from different sources into MDM
 - Management of maps used to export data from MDM to target systems
 - Runtime transcoding of reference values used in transactions from bespoke systems that create, update or retrieve data from MDM
- **MDM and SAP**
IBM InfoSphere Master Data Management Server is the repository that centralizes and manages the organization's critical master data entities such as customer, product, supplier, and more. It provides reliable and flexible delivery of the master data for SAP applications. The InfoSphere MDM Ref DM Hub is used to import the MDM-specific values and the SAP-specific values, define mappings, and export the mappings to create the transcoding tables that are read within the Enterprise Service Bus mediation flow.
- **Taxonomy Management for Enterprise Content Management pattern**
In Enterprise Content Management (ECM) systems, taxonomies are used to classify content to make it easier to find in content delivery systems. InfoSphere MDM Ref DM Hub can be used as a component in an enterprise taxonomy management solution for content management and delivery.
- **Data warehouse**
Data warehouses are used as the enterprise repository for business intelligence data. Reference data plays three key roles in a data warehouse environment:
 - Providing a consistent canonical definition of reference codes to be used in all tables in the warehouse. Having the reference data defined in InfoSphere MDM Ref DM Hub, and then pushed into the warehouse, enables consistency and governance around reference data dimensions.
 - Delivering a well managed set of hierarchies for reporting and analysis. Many reports are interactive and hierarchical in nature. You can use InfoSphere MDM Ref DM Hub to ensure the consistency, integrity, and governance of that hierarchy before the hierarchy is published to the warehouse.
 - Ensuring that reference data from source systems is mapped to the common canonical reference data used in the warehouse as part of the ETL process. InfoSphere MDM Ref DM Hub is ideal for

managing the sets associated with the different source system and the warehouse, and for empowering business data stewards to create the maps and version them for the mappings between the sources and the warehouse.

Supported platforms

IBM InfoSphere MDM Reference Data Management Hub supports IBM AIX®, Sun Solaris, and Linux Red Hat operating systems. The database systems supported are DB2 Enterprise Server Edition Version 9.7, Version 10.1, and Oracle Database 11g Enterprise Edition.

For complete hardware and software requirement information, refer to the InfoSphere MDM Reference Data Management Hub Installation Guide and the readme document.

The hardware and software requirements for InfoSphere MDM Reference Data Management Hub might be updated. To obtain the most current information for supported hardware, visit http://www.ibm.com/software/data/infosphere/mdm_server/requirements.html

Ordering information

IBM InfoSphere MDM Reference Data Management Hub V10 is only available via IBM Passport Advantage®. It is not available as shrinkwrap. This product can only be sold directly by IBM or by authorized IBM Business Partners for Software Value Plus.

For more information about IBM Software Value Plus, visit: http://www.ibm.com/partnerworld/page/svp_authorized_portfolio

To locate IBM Business Partners for Software Value Plus in your geography for a specific Software Value Plus portfolio, contact your IBM representative.

For information about Passport Advantage, visit: <http://www.ibm.com/software/passportadvantage>

Related information

For more information, see the following documents:

- IBM Offering Information page (to search on announcement letters, sales manuals, or both): http://www.ibm.com/common/ssi/index.wss?request_locale=en
On this page, enter InfoSphere MDM Reference Data Management Hub, select the information type, and click **Search**. On the next page, narrow your results by geography and language.
- IBM InfoSphere MDM Reference Data Management Hub Sales Manual: <http://ibm.co/12DX0sU>
- IBM InfoSphere MDM Reference Data Hub Information Center: <http://bit.ly/1072u0X>
- IBM InfoSphere MDM Reference Data Management Hub Product page: <http://www-01.ibm.com/software/data/infosphere/mdm/reference-data-management-hub/>
- IBM Redbooks publication: *A Practical Guide to Manage Reference Data with InfoSphere MDM Reference Data Management Hub*, SG24-8084: <http://www.redbooks.ibm.com/Redbooks.nsf/RedpieceAbstracts/sg248084.html>

Notices

This information was developed for products and services offered in the U.S.A.

IBM may not offer the products, services, or features discussed in this document in other countries. Consult your local IBM representative for information on the products and services currently available in your area. Any reference to an IBM product, program, or service is not intended to state or imply that only that IBM product, program, or service may be used. Any functionally equivalent product, program, or service that does not infringe any IBM intellectual property right may be used instead. However, it is the user's responsibility to evaluate and verify the operation of any non-IBM product, program, or service. IBM may have patents or pending patent applications covering subject matter described in this document. The furnishing of this document does not give you any license to these patents. You can send license inquiries, in writing, to:

IBM Director of Licensing, IBM Corporation, North Castle Drive, Armonk, NY 10504-1785 U.S.A.

The following paragraph does not apply to the United Kingdom or any other country where such provisions are inconsistent with local law: INTERNATIONAL BUSINESS MACHINES CORPORATION PROVIDES THIS PUBLICATION "AS IS" WITHOUT WARRANTY OF ANY KIND, EITHER EXPRESS OR IMPLIED, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF NON-INFRINGEMENT, MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE. Some states do not allow disclaimer of express or implied warranties in certain transactions, therefore, this statement may not apply to you. This information could include technical inaccuracies or typographical errors. Changes are periodically made to the information herein; these changes will be incorporated in new editions of the publication. IBM may make improvements and/or changes in the product(s) and/or the program(s) described in this publication at any time without notice.

Any references in this information to non-IBM Web sites are provided for convenience only and do not in any manner serve as an endorsement of those Web sites. The materials at those Web sites are not part of the materials for this IBM product and use of those Web sites is at your own risk. IBM may use or distribute any of the information you supply in any way it believes appropriate without incurring any obligation to you. Information concerning non-IBM products was obtained from the suppliers of those products, their published announcements or other publicly available sources. IBM has not tested those products and cannot confirm the accuracy of performance, compatibility or any other claims related to non-IBM products. Questions on the capabilities of non-IBM products should be addressed to the suppliers of those products. This information contains examples of data and reports used in daily business operations. To illustrate them as completely as possible, the examples include the names of individuals, companies, brands, and products. All of these names are fictitious and any similarity to the names and addresses used by an actual business enterprise is entirely coincidental.

Any performance data contained herein was determined in a controlled environment. Therefore, the results obtained in other operating environments may vary significantly. Some measurements may have been made on development-level systems and there is no guarantee that these measurements will be the same on generally available systems. Furthermore, some measurement may have been estimated through extrapolation. Actual results may vary. Users of this document should verify the applicable data for their specific environment.

COPYRIGHT LICENSE:

This information contains sample application programs in source language, which illustrate programming techniques on various operating platforms. You may copy, modify, and distribute these sample programs in any form without payment to IBM, for the purposes of developing, using, marketing or distributing application programs conforming to the application programming interface for the operating platform for which the sample programs are written. These examples have not been thoroughly tested under all conditions. IBM, therefore, cannot guarantee or imply reliability, serviceability, or function of these programs.

© Copyright International Business Machines Corporation 2013. All rights reserved.

Note to U.S. Government Users Restricted Rights -- Use, duplication or disclosure restricted by GSA ADP Schedule Contract with IBM Corp.

This document was created or updated on May 16, 2013.

Send us your comments in one of the following ways:

- Use the online **Contact us** review form found at:
ibm.com/redbooks
- Send your comments in an e-mail to:
redbook@us.ibm.com
- Mail your comments to:
IBM Corporation, International Technical Support Organization
Dept. HYTD Mail Station P099
2455 South Road
Poughkeepsie, NY 12601-5400 U.S.A.

This document is available online at <http://www.ibm.com/redbooks/abstracts/tips1016.html> .

Trademarks

IBM, the IBM logo, and ibm.com are trademarks or registered trademarks of International Business Machines Corporation in the United States, other countries, or both. These and other IBM trademarked terms are marked on their first occurrence in this information with the appropriate symbol (® or ™), indicating US registered or common law trademarks owned by IBM at the time this information was published. Such trademarks may also be registered or common law trademarks in other countries. A current list of IBM trademarks is available on the Web at <http://www.ibm.com/legal/copytrade.shtml>

The following terms are trademarks of the International Business Machines Corporation in the United States, other countries, or both:

AIX®
DB2®
IBM®
InfoSphere®
Passport Advantage®
Redbooks®
Redbooks (logo)®
WebSphere®

The following terms are trademarks of other companies:

Linux is a trademark of Linus Torvalds in the United States, other countries, or both.

Other company, product, or service names may be trademarks or service marks of others.