**Understanding of Informatica RDM Accelerator and Its Implementation**

**Prepared By:**

Saikat Sinha

**Contents**

[1 Introduction: 3](#_Toc470718196)

[2 Content Coverage: 3](#_Toc470718197)

[3 Understanding of Reference Data: 3](#_Toc470718198)

[4 Importance of Reference Data Management and It’s Key Features: 4](#_Toc470718199)

[5 RDM Implementation using INFA RDM Accelerator: 4](#_Toc470718200)

[6 Data Model: 5](#_Toc470718201)

[7 Source Data Design: 5](#_Toc470718202)

[8 Data Insertion in the Accelerator: 7](#_Toc470718203)

[9 Publishing Data: 12](#_Toc470718204)

[10 Advantages of RDM Accelerator: 13](#_Toc470718205)

# Introduction:

This document details a Reference Data Management (RDM) design using Informatica RDM Accelerator which is based on INFA MDM 10.1.

The RDM provides the processes and technologies for recognizing, harmonizing and sharing coded, relatively static data sets for “reference” by multiple constituencies (people, systems, and other reference data domains). There are numerous agencies that can all benefit from RDM. Due to desire of the ability to manage and map relationships between different reference data sets which exist in an enterprise, it is important to consider an RDM solution as an Enterprise-wide solution.

This document describes the General Structure of a Reference Data Management, the advantages of RDM, potential considerations to be understood before implementing, and services that can be realized through successful Implementation.

# Content Coverage:

* Understanding of Reference Data
* Implementation of RDM Accelerator through Country Codes example
* Advantages of using RDM Accelerator

# Understanding of Reference Data:

Reference data is a special subset of master data that is used for classification throughout the entire organization: postal codes, cost centers, financial hierarchies or countries. Whether the data is externally mandated or internally authored it is unambiguous and non-negotiable.

Reference data is recognized to play an important role in data governance initiatives. Whether it can be referred to as lookup values, reference codes, code tables, or reference hierarchies, reference data is an important part of the business when it is needed to govern common code types, which is why organizations all over the world are looking for a way to manage reference data.

# Importance of Reference Data Management and It’s Key Features:

 Each application in an enterprise has its own representation of code sets defining the same thing. During integration of master data (or any data) across applications, it is necessary to translate between the different code table representations in order to categorize data in a consistent way.

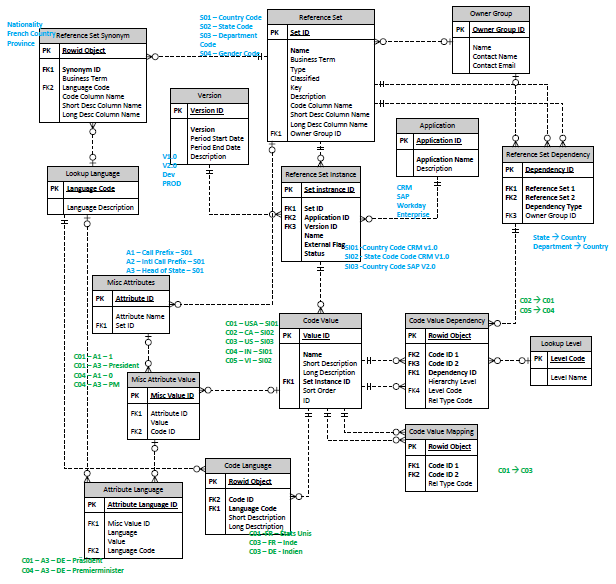
Mapping between the different representations and keeping track of changes across all the different code table variations on an ongoing basis can be a major challenge. Many enterprises struggle with this challenge and often use error-prone manual processes to record and manage changes to reference data sets. Errors in the reference data can have a major business impact. Quality issues in reference data have ripple effect and can cause major issues in the downstream applications.

* 1. **Features of Reference Data Management:**
* Robust interface controlled by role based security to support collaborated authoring of reference data.
* Versioning and auditing capability
* Ability to publish reference data
* Efficient load/extract functionality
* Good error tolerant search capability

# RDM Implementation using INFA RDM Accelerator:

Informatica Reference Data Management Accelerator is a business ready platform often hosted in a cloud platform by which organization can easily leverage their common reference data. There is no need to invest much on development activities for the typical MDM, rather it is platform which follows the concept of map and load the data in the specified data model.

# Data Model:



# Source Data Design:

Source data needs to be mapped in the accelerator data model in such a way that the logical meaning of the data remains intact. The name of source data columns may not be used as tables and columns are already defined in this model.

Among the 16 tables there are three base tables which can hold the source data

* **Reference Set** - The table stores information about the types of reference sets that you use.
* **Reference Set Instance** - An instance of a set or a dependent set that is managed by an external application. An application uses a reference data set instance. The table stores the information about an instance of a reference set.
* **Code Value** - The table stores all the actual reference data values that the applications use.

However, any other records of source which are mainly supported records can be kept in Misc Attribute Value table. There are few more tables such as Version, Application which are used to maintain the versioning of the record and keep the name of source application respectively. Insertion of the data in these tables are one time activity and can be done through IDD.

Below is an example of Country Code as source data which is mapped to the accelerator.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Reference Set** | **Reference Set Instance** | **Relation between Ref Set Ins and Code Val** | **Code value** | **Misc Attributes Values** |
| **Name:**  Country Code |  | (1,n)  [For a single instance] |  | **Attribute ID** |
| **Set ID**:Country Code | **Name**:<Country Alpha 3 Code> | Country Code Valid From Date |
| **Application ID**:<ISO> | **Short description**: <Country Name> | Country Code Valid To Date |
| **Version ID**: <3.0> | **CodeValue Identifie**r:<Country Numeric Code> | Country Code Comment |
| **Buisness Term**: Country | **NAME**:<ISO Country Code 3.0> |  | Country Code Effective Date |
| **External Flag**: <Y> | Country Code Created Date |
| **Status**:<Active> |  |
|  |
|

# Data Insertion in the Accelerator:

To insert source data in the accelerator model following sequence needs to be followed:

**Reference Set**

**[**Mandatory attribute: ***Name*]**

**Misc Attributes** (if any)

**Reference Set Instance**

**[**Mandatory attribute: ***Set ID (****FK with Reference Set****), Application ID, Version ID, Name, External Flag, Status*]**

**Code Value**

**[**Mandatory attribute: ***Name, Set Instance ID(****FK with Reference Set Instance****)*]**

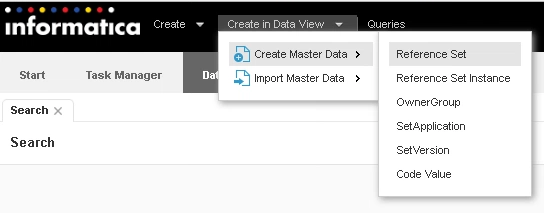
**Misc Attribute Value** (if any)

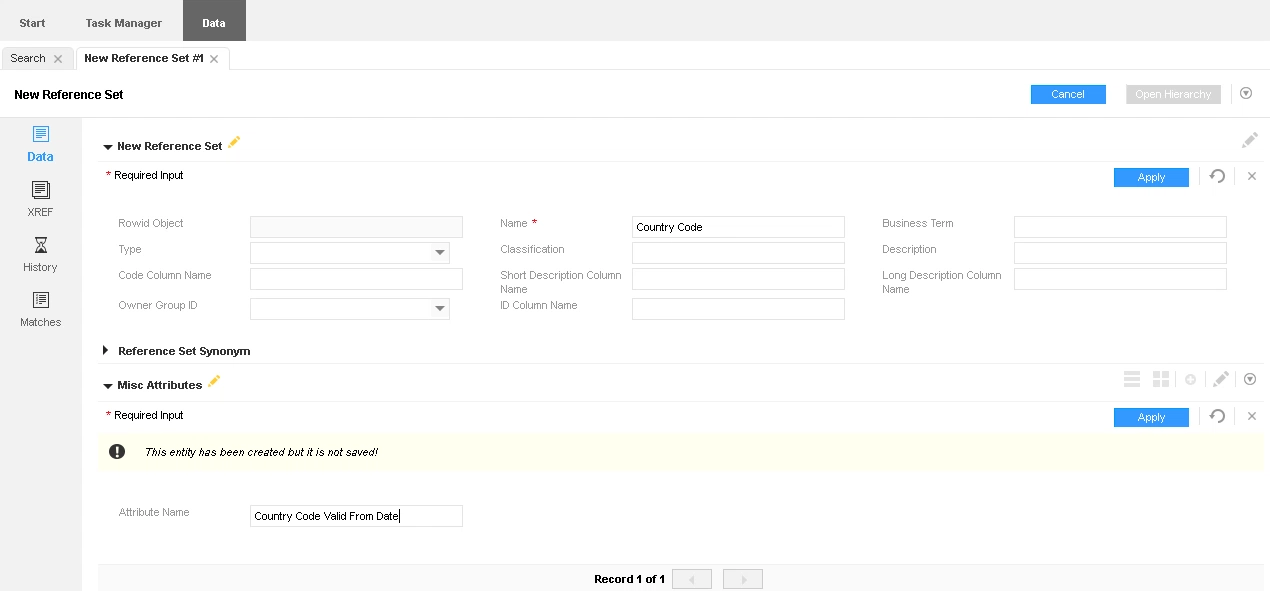
**[**Mandatory attribute: ***Attribute ID (****FK with* Misc Attributes***)*]**

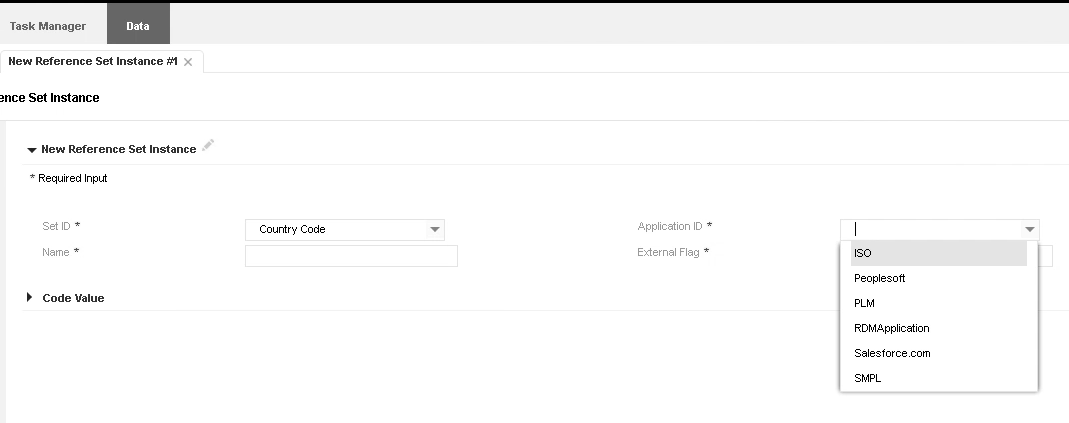
As mentioned above, data needs to be entered for Version and Application by IDD as a onetime activity before inserting data in the Reference Set Instance.

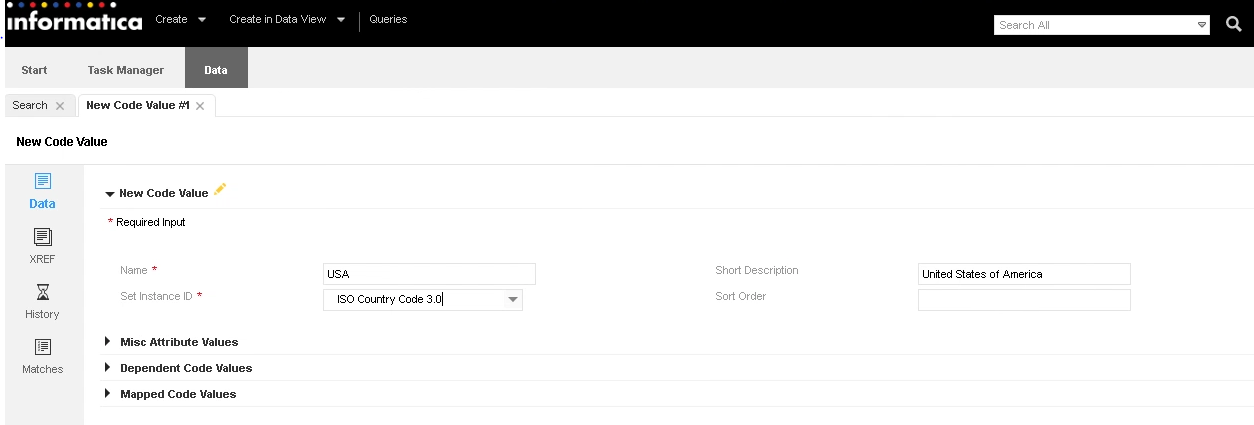
There are three ways by which data can be loaded to the accelerator.

* **Manual Insertion** – Users can insert record by Informatica Data Director (IDD) in the respective tables.



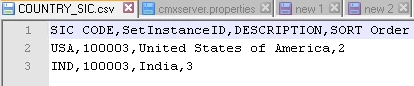


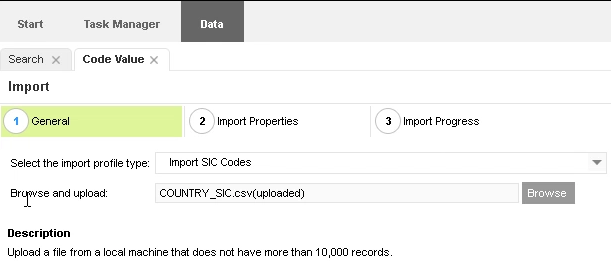


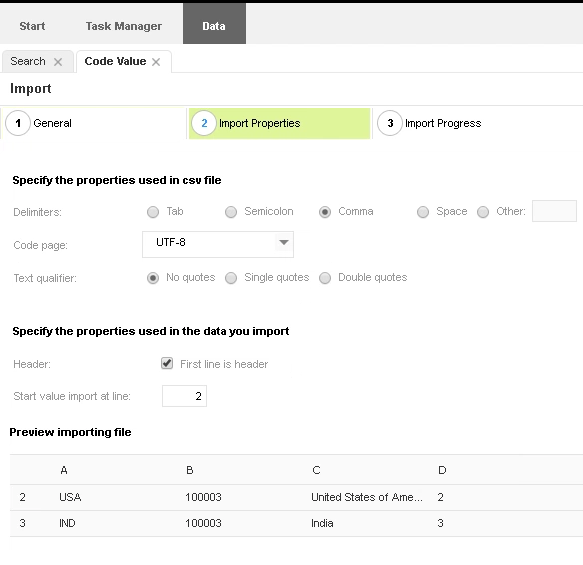


* **Uploading Data in Bulk** – Users can upload csv files only for code value table for any bulk insertion by IDD.

There are four import profile types by which csv files can be imported. Please follow the steps for importing the file to load data in code value by ‘Import SIC Codes’ type.



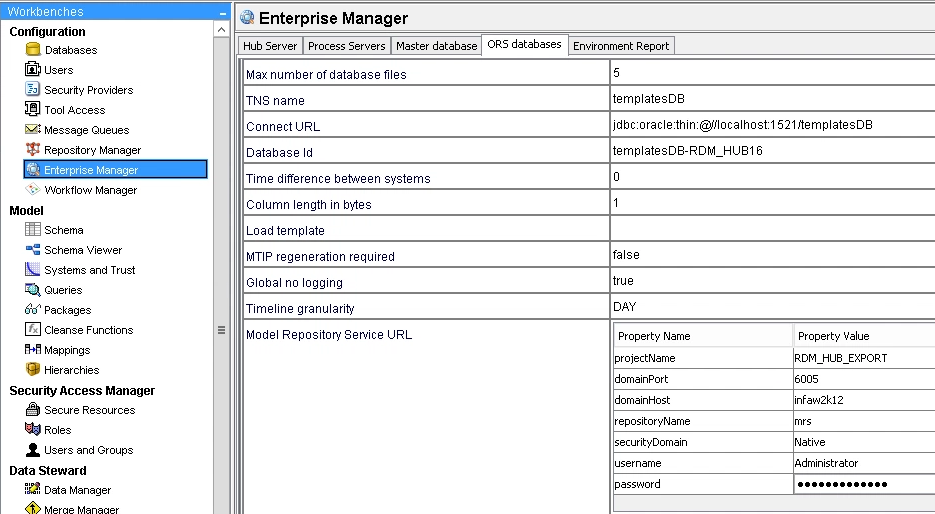




* **Data Loading by ETL** – There is no landing table in the accelerator. Therefore data can be inserted directly to the staging tables and then base objects by IDQ. To load the data directly to the staging tables Model Repository Service (MRS) needs to be configured in the MDM hub. Then mapping has to be created to load the data in the IDQ from source to stage and then BO.

There are few steps to be followed to load data by IDQ directly to staging table:

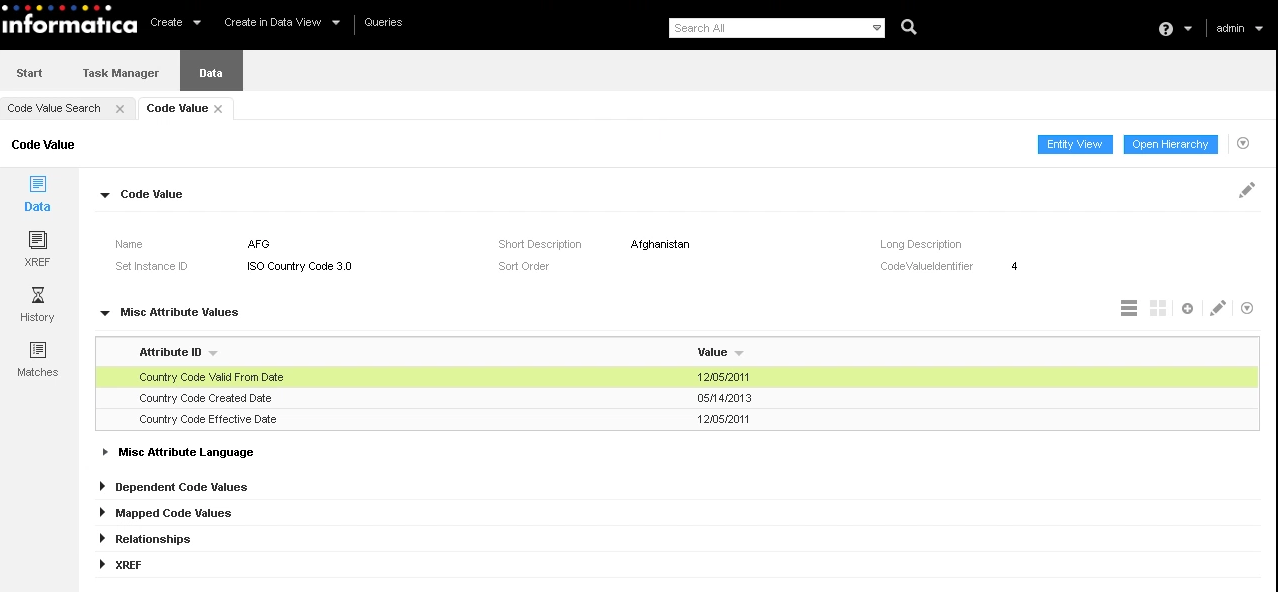
* + Enter the Model Repository Service connection parameters to the MRS\_URL field (as shown in the image below)
  + Enable ‘Informatica platform staging’ of staging tables
  + Enable ‘Synchronize with Model Repository Service’ of staging tables



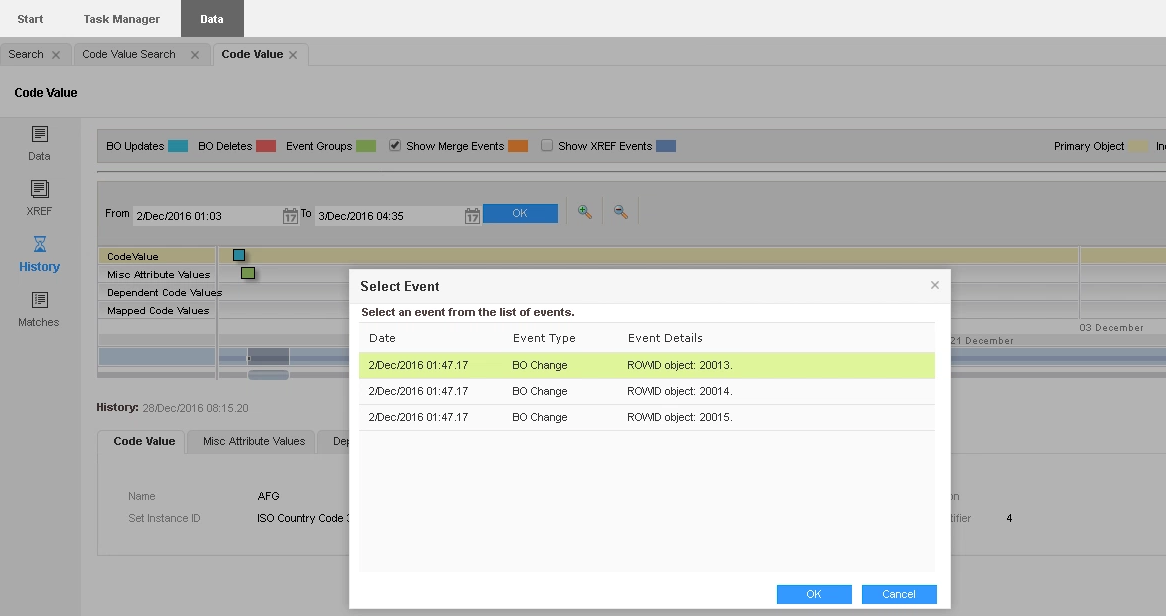
For further details please follow the steps described in the doc.



Final view of a Country Code data in the code value table would be looking like as follows:



For auditing purpose of the data history of any operation done on a particular record can also be seen as follow:



# Publishing Data:

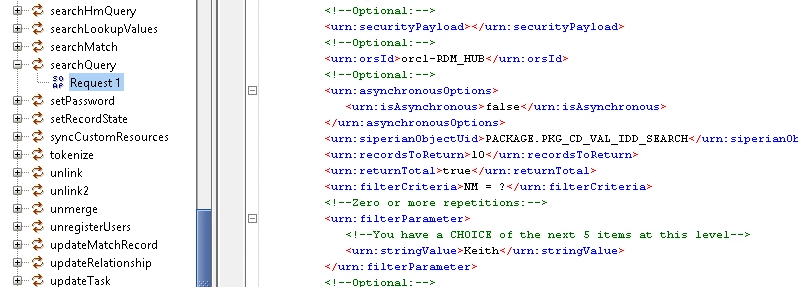
For real-time processing, applications that are external to Informatica MDM Hub invoke Informatica MDM Hub operations using the Services Integration Framework (SIF) interface. SIF provides APIs for various Informatica MDM Hub services.

The SIF API allows to implement the request/response interactions using various access protocols. However, loosely coupled web services using SOAP protocol is one of the best and widely used protocol. Other than the pre-specified packages or quires any customized package or query can also be created and exposed to the external system through SIF.

For C-R-U-D operations following SOAP APIs are used:

* ***get*** – For fetching particular records based on row id
* ***put*** – For inserting or updating any existing record
* ***searchQuery*** – For fetching any records based on business search criteria
* ***delete*** – For deleting any record based on row id

Below is an example of *searchQuery* API request:



# Advantages of RDM Accelerator:

* Very less development effort as most of the MDM Hub development is ready to use
* Often hosted in a cloud platform. So less dependency on installation and configuration
* A predefined, well-structured data model for fitment of most of the reference data
* Shares same platform and concept of Informatica MDM. So any customization needed in the accelerator can be done easily
* If business requires separate data model or additional Landing, Staging, Base Objects can be configured in the same MDM platform