



Vermont Enterprise Architecture Framework (VEAF)

Master Data Management Design

EA APPROVALS

Approving Authority:

<Signature>

<Date>

<Printed Name>

<Position Title>

REVISION HISTORY

Version	Date	Organization/Point of Contact	Description of Changes
0.1	08/10/2015	EA, Seamus Loftus/Arun	Initial Draft – Document Outline
0.2	08/14/2015	CTO, John P Hunt	Need brief para describing the criticality of Data Governance
0.3	08/14/2015	EA, Casey Cleary	Adding brief para describing the criticality of Data Governance
0.4	08/14/2015	CTO, John P Hunt	Added Data Governance entities, tightened up DG functions.
0.5	09/01/2015	Cameron Bradley	Added Graphics, regrouped “Marketing Strategy” into “Advantages” “Considerations” and “Services”

Review History

Version	Date	Organization/Point of Contact	Description of Changes
0.5	9/21/2015	CTO, John P. Hunt	Draft Approval for Discussion

Confidentiality Statement

This document is produced for the State of Vermont and cannot be reproduced or distributed to any third party without prior written consent.

No part of this document may be modified, deleted, or expanded by any process or means without prior written permission from the State of Vermont.

Table of Contents

Revision History	2
1. Introduction	5
1.1. Intended Audience.....	5
1.2. Purpose.....	5
2. Data Governance	5
3. Design	6
4. Advantages of MDM	8
4.1. Single Source of Truth.....	8
4.2. Quality of Service.....	8
4.3. Cost Reduction	8
4.4. Stewardship	9
4.5. Open Standards.....	9
4.6. Auditability.....	9
4.7. Security.....	9
5. MDM Considerations	9
5.1. Data Profiling.....	9
5.2. Rules.....	9
6. MDM Services	10
6.1. Probabilistic Matching.....	10
6.2. Merge/Unmerge.....	10
6.3. De-duplication.....	10
6.4. Cross Reference.....	10
6.5. Address Verification.....	10
6.6. Data Cleansing.....	10
7. Onboarding Process	10

- 7.1. Implementation..... 11
 - 7.1.1. Implementation Team 11
- 7.2. Operational 11
 - 7.2.1. Maintenance & Operations Team 11
- 8. Data Warehouse Integration 12**
- 9. Service Revenue 12**
 - 9.1. Initial..... 12
 - 9.2. Service Surcharge 13

Table of Figures

- Figure 1 MDM Design Example..... 6
- Figure 2 Ideal Information Architecture 7
- Figure 3 Overall Data Context..... 8

DRAFT

1. INTRODUCTION

This document details a Master Data Management (MDM) Design for use within the state of Vermont. There are numerous agencies that can all benefit from MDM, but building a unique MDM solution for each defeats the purpose of MDM. Due to desire to avoid duplicative effort, it is important to consider an MDM solution as an Enterprise-wide solution. The Overall Master Data Management strategy for the State of Vermont has been detailed fully in the Vermont Enterprise Architecture Framework (VEAF).

1.1. Intended Audience

The intended audience of this document is for CIOs, CTOs, Enterprise Architects, Implementation Teams, and Business Leads of various State of Vermont agencies focused on a single solution for MDM within the state.

1.2. Purpose

The purpose of this document is to detail a proposed MDM design for use by the State of Vermont.

1.2.1. Summary

This document describes the General Structure of a Master Data Management Solution for the State of Vermont, the advantages of an MDM solution, potential considerations to be understood before implementing, and services that can be realized through successful Implementation. It then explains, in general terms, the responsibilities of the Onboarding and Maintenance and Operations teams. Finally, discusses the upfront costs and potential advantages of an MDM implementation.

2. DATA GOVERNANCE

MDM design starts with Data Governance.

Data Governance is the specification of permissions and accountability used to encourage desirable behavior in the creation, valuation, storage, use, archival, and deletion of data and information. Data Governance entities exist at two levels.

1. Agency Data Governance Teams – Responsible for the governance of Data within their agency.
2. State of Vermont Data Governance Council – A group that meets quarterly to review policies, standards, and accepted data models used throughout the State of Vermont.

In general, Data Governance adheres to the following guiding principles to ensure the right people have the right information at the right time:

- 1. Integrity** – The Agency Data Governance Team focuses on the drivers, constraints, options, and impacts of data related decisions.
- 2. Transparency** – Both the Agency and State Data Governance Teams document how and when data related decisions and controls were introduced into the processes.
- 3. Auditability** – Data Governance decisions, processes, and controls will be auditable and accompanied by documentation supporting compliance-based and operational auditing requirements per the State of Vermont Data Governance Council.

4. Accountability – There will be accountability for cross-functional data-related decisions, processes, and controls.

5. Stewardship – Individual contributors will be held accountable for data under their stewardship. In general, Data Stewards function within the Agency Data Governance.

7. Standardization – The SoV Data Governance Council is responsible for the repository of data in the form of logical data models, meta-data models, shared XML registries, and the enterprise data dictionary that can be shared among State Agencies.

8. Change Management - Change Management activities must be performed for reference data values and the structure/use of master data and metadata.

3. DESIGN

The design of an enterprise wide solution consists of an MDM software suite installed in a virtualized environment or in the Cloud. The MDM solution would be accessed via an Enterprise Service Bus through standard web services. Agencies and their applications would connect to the SOA with their own web services, which would then make calls to the MDM as needed. The Data Governance Council determines what data should be stored within the MDM solution, these directives are then executed by the Implementation Team. Reporting will be generated by out of the box reports from either the MDM solution or Business Intelligence tools integrated into the solution.

This is detailed out in the figure below:

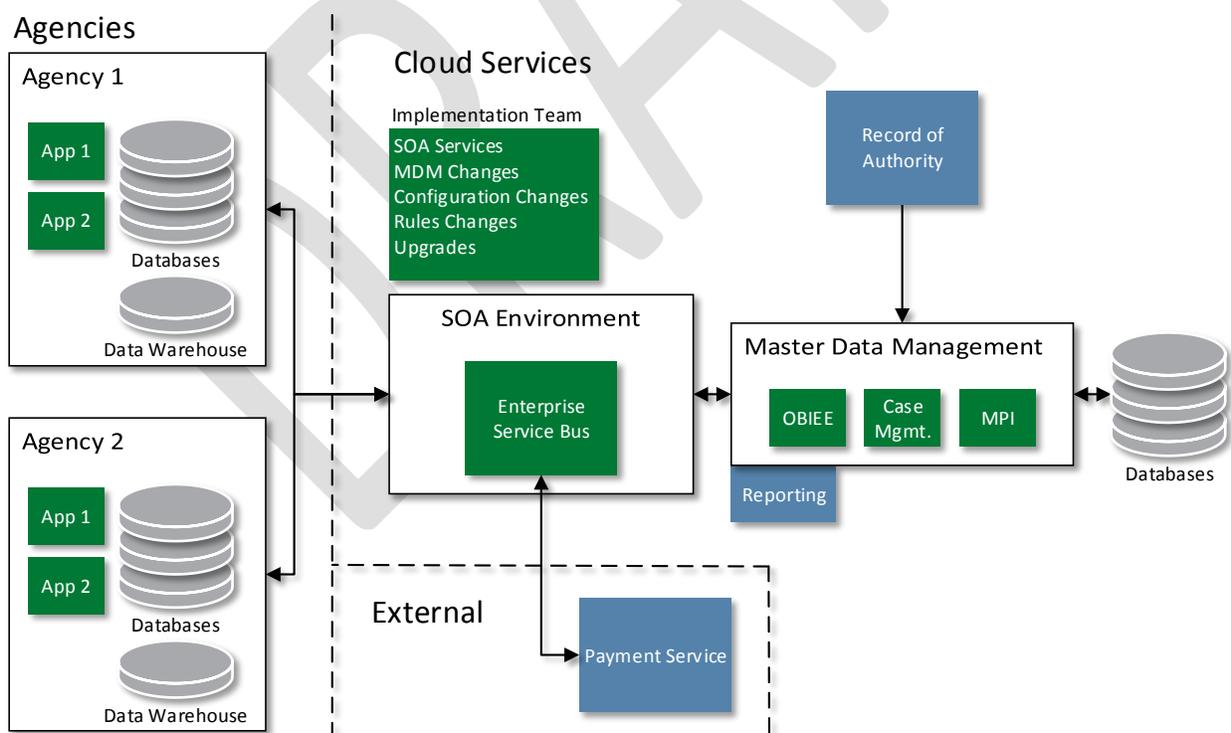


Figure 1 MDM Design Example

The example diagram is made to match an ideal implementation described in the graphic below, from a September 2011 Oracle White Paper.

Ideal Information Architecture

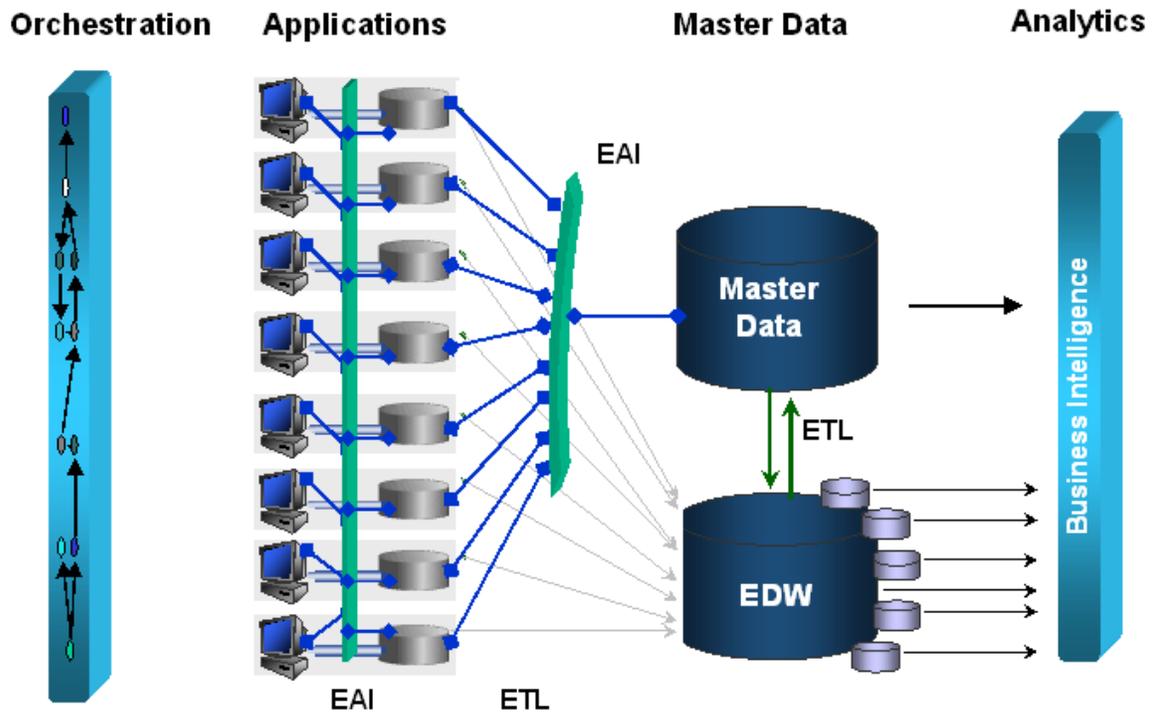


Figure 2 Ideal Information Architecture

Master Person Index is a logical place to start when building an MDM solution. Vermont citizens have a number of attributes including name, address, social security number, phone number, date of birth, and a number of other data fields that can be stored in MDM. There are close to 650,000 residents in the State of Vermont. This volume of citizens should be able to be handled by a single MDM solution. As a new application is onboarded, the data governance council and data stewards will determine if and what additional data may be added to the Master Data Hub.

The image below originates from the VEAF Information Strategy Presentation shows an example of data within an agency, displayed here to discuss data that the entire state may use.

The following is from the VEAF Information Architecture deck.

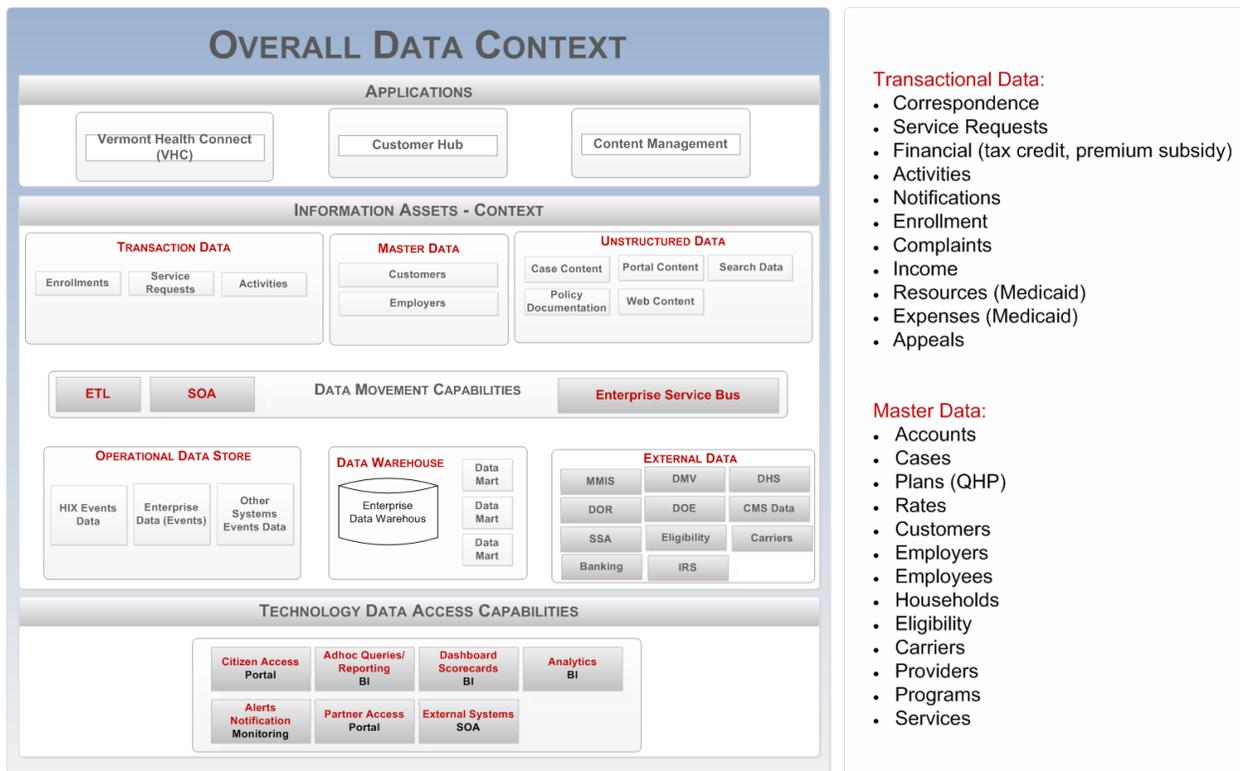


Figure 3 Overall Data Context

4. ADVANTAGES OF MDM

The sections below detail out several advantages a state-wide enterprise solution will bring to the business:

4.1. Single Source of Truth

MDM serves to provide a single source of truth, allowing for data elements to be stored only once on the platform. Currently, applications within agencies have data which may be duplicated elsewhere, creating integrity issues with the data. Older systems may not rely upon relational databases, causing them to generate more data than necessary and causing headaches for those tasked with managing the data.

4.2. Quality of Service

A state-wide solution for MDM can ensure consistent quality of service across all agencies. Data can be accessed quickly and efficiently, instead of being stored across multiple systems requiring complicated and time inefficient integration.

4.3. Cost Reduction

A fully realized MDM solution will reduce costs across all participating agencies. Each agency is working toward increasing the quality of their data, with several seeking to implement their own Master Data solutions. However, one enterprise-wide Master Data solution offers a better and more cost effective solution. This solution would allow agencies to share costs across the enterprise without having to create their own solution.

4.4. Stewardship

Stewardship of data is more manageable with a state-wide MDM solution. Stewards can be given the latitude to define data at the enterprise level and then push out data rules to other agency and department applications.

4.5. Open Standards

An MDM solution using Open Standards allows for easier integration with both new and legacy applications being onboarded. By moving away from proprietary software standards allows for competition between multiple to supply solutions, ultimately reducing the cost to the State.

4.6. Auditability

An MDM solution provides cleaner data, which leads to better auditability of state data assets. This will allow for an audit trail will exist on the data that can be reviewed independently and show that the processes for managing compliance are transparent as well. (Master Data Management by Dave Loshin, pg 77)

4.7. Security

When MDM is integrated with the SOA Service Bus offers it creates a secure layer of defense ensuring that only calls to the Master Data Hub from the SOA Service Bus are allowed. When another application needs to send information to the Hub, it must send it through the service bus, which handles the initial security.

5. MDM CONSIDERATIONS

The following should be considered when designing an MDM solution.

5.1. Data Profiling

Data profiling is the first step of data integration. It is assess the quality of source data, before it is loaded into the Master Data Hub. This ensures that baseline data quality and rules are maintained in the initial Master Data Hub. This is also discussed in the VEAF MDM Strategy, the purpose of Data Profiling is to:

- Understand if data is complete, accurate, and conforms to the State's source system
- Highlight missing, incomplete, and erroneous values in data sets; emphasizing repeating patterns
- Assess the quality of data sources to determine if and how data will be brought into the Master Data Hub
- Understand the risk, if any, associated with data
- Provide insight into the effort required to standardize, cleanse, and transform data

5.2. Master Data Rules

Master Data rules are important, because they allow the Data Governance council to implement standards across the organization's Master Data. The Data Governance Council can then use these established rules to increase the quality and usability of the Master Data.

6. MDM SERVICES

An important part of a contributing to the success of a state wide MDM solution is the cooperation of the various agencies and departments. This is not possible until there is a common understanding of the services that make up an MDM Solution. All groups working together, ensure more successful outcomes of MDM implementations. It is important that all parties using an MDM Solution understand the services offered by an MDM Solution. It is important to clearly detail the benefits of a Statewide MDM solution over individual departmental MDM solutions.

6.1. Probabilistic Matching

Probabilistic matching is also known as a fuzzy search. It assigns weights to fields and then sums the values when matching records. If the sum of the weights are close, there is a higher propensity of the records matching.

6.2. Merge/Unmerge

Merge is an operation that compares two data sets, and merges them into one record. This further enhances the master data. Unmerge is the reversal of merge. This allows multiple agency datasets to be matched for use in the MDM solution.

6.3. De-duplication

A state-wide MDM solution offers the advantage of de-duplication, or an elimination of matching records.

6.4. Cross Reference

Cross-referencing enables the integration layer to identify correlated information across independent applications. This is needed for every integration between applications. Within an MDM solution, there is also a cross-reference on the application level within data hubs to identify source systems for master records. (<http://www.oracle.com/us/products/applications/056910.pdf>)

6.5. Address Verification

An MDM solution offers Address Verification through the Oracle Enterprise Data Quality Server. Many applications require costly address services that can cause a lot of pain and time to update applications with new addresses. A centralized address verification service for the state can help with this, by offering the service out to the integrated applications.

6.6. Data Cleansing

MDM offers a data cleansing service, which removes data in a database that is incorrect or incomplete. This allows for better quality of data and increases the integrity level of the data for the entire state.

7. ONBOARDING PROCESS

As part of the MDM design, it is important to have a plan to first onboard an MDM solution, and then ensure that new applications brought into the State Enterprise are allowed to take advantage of an MDM solution. It is important to recognize two particular phases of an application lifecycle, the Implementation and Operational phases. The implementation phase, requires a great effort to get an application into

place. Once the initial implementation is complete, the application moves from implementation to a steady operational state. It is in the operational phase where organizations will fully realize the benefits of an MDM solution.

7.1. Implementation

The initial process involves taking the application data through Extract, Transform, Load (ETL) and going through a number of activities to improve the quality of the Master Data. It will also require the effort of an Implementation Team. These activities include data profiling, probabilistic matching, merging, de-duplication, and applying data rules to provide a clean set of Master Data. SOA services must then be developed to connect data from the new application to the Enterprise Service Bus, which can then call MDM services, such as Access Combined Contact Service, Merge and Publish Person, Match and Fetch Person, and Synchronize Person.

Additional services can be added as required and defined by the data governance council. A data steward will ensure that the policies set by the data governance council are enforced, such as length of fields, and which data fields were selected for

7.1.1. Implementation Team

An implementation team is required to perform several operations within an MDM environment. They will need to ETL the data from the integrating application to the MDM environment, and perform data profiling, probabilistic matching, merging, de-duplication, and the application of the data rules. The implementation team will need to build the necessary SOA services from their integrating application to the SOA service bus. This team will use open standards to integrate with the SOA through web service calls.

7.2. Operational

Once the implantation has been completed, the application will move into its operational phase. In the operational phase, the advantages of a state-wide enterprise solution will be realized. A data steward will monitor the master data to ensure the quality is high. There may be a need for additional integration at this phase, if a previously integrated legacy system decides they wish to use data from a newly integrating application. This activity would be the responsibility of the M&O team to implement.

7.2.1. Maintenance & Operations Team

The maintenance and operations team will be responsible for the day to day operations of the MDM/SOA environment on a cloud-based platform. They will perform several activities critical to an MDM implementation's success. They will perform server administrator duties to ensure system health. The team will ensure that security and security monitoring are in place to protect the data stored on the MDM solution. They will be responsible for the health of the MDM composites on the SOA Service Bus. They will ensure reports are generated and routed to the correct personnel, such as the data governance council, data stewards and identified business users. They will be responsible for building new composites as needed between SOA, MDM, and OBIEE. They may even need to build composites between the new integrating application and the SOA Bus.

8. DATA WAREHOUSE INTEGRATION

A data warehouse stores historical data, therefore it is different than Master Data Management. It is important to note this differences, as supplied in the figure below, from this online article (<http://www.toadworld.com/platforms/sql-server/b/weblog/archive/2013/09/24/data-warehouse-vs-master-data-managment>):

DATA WAREHOUSE	MASTER DATA MANAGEMENT
Data is historical	Data is current
Contains transactional data as well as master data	Contains only subject area specific master data
Contains only that data deemed necessary for analytical demands	Contains complete information about entities used across the enterprise
Dimensions are flat	Entities are contextual, entity views
Usually a consumer of data, an endpoint	Usually a consumer and a provider, often with a bidirectional relationship
Purpose is to analyze historical data	Purpose is to ensure accurate and updated entity data
Focus is on the aggregate and Reporting	Focus is on the entity
Dimensions not consumed by other LOB applications	Entities consumed by other applications, including data warehouse
Tools are built around analysis	Tools are built around management
"One version of the truth"	"Current snapshot of the world"

An MDM solution can integrate with data warehouses via the Data Quality servers.

9. SERVICE REVENUE

While a state-wide MDM solution would be a great achievement for the state of Vermont, it is important to find a way to finance the solution. The solution will cost the state money, but if the costs can be spread out across various agencies and departments who use the solution, it will be easier to justify. A multi-agency solution will also reduce costs for agencies in times of tight budgets.

9.1. Initial

Agencies and departments would be charged an initial fee to join the state-wide MDM solution. This would cover any integration the Maintenance and Operations team would be responsible for, as well as the initial required MDM items that includes Data Profiling, Probabilistic Matching, applying Rules, Merging, and the de-duplicating of data. The initial charge will also include establishing or onboarding the agency or department

9.2. Service Surcharge

Once an application is onboarded, it is important to determine a way to charge out to customers of the MDM solution. A method of this is to use a service surcharge. Every time the service is used, it can be monitored on which state application made the call to the MDM. DII can charge based on the service calls. This offers a fair advantage to applications that use the system very rarely, versus systems that are constantly using MDM.

The service cost model should be as follows:

Total Number of Service Calls / (Total M&O Cost + Hardware & Software License) = Individual Service Charge

A bill would be provided to each agency based on the volume of services they used multiplied by the individual service charge. This billing model could be by quarter or yearly, as desired.