

**VA Enterprise Design Patterns  
Interoperability and Data Sharing**

# **Enterprise Data Access**

**OFFICE OF TECHNOLOGY STRATEGIES (TS)  
OFFICE OF INFORMATION AND TECHNOLOGY (OI&T)**

**VERSION 2.0**

**DATE ISSUED: AUGUST 2017**



## APPROVAL COORDINATION

Gary E.  
Marshall  
137891

Digitally signed by Gary E.  
Marshall 137891  
DN: dc=gov, dc=va,  
o=internal, ou=people,  
0.9.2342.19200300.100.1.1=g  
ary.marshall@va.gov,  
cn=Gary E. Marshall 137891  
Date: 2017.08.15 14:40:49  
-04'00'

GARY MARSHALL  
DIRECTOR, TECHNOLOGY STRATEGIES, ASD

Deeneen  
U Akeo  
622220

Digitally signed by Deeneen U  
Akeo 622220  
DN: dc=gov, dc=va, o=internal,  
ou=people,  
0.9.2342.19200300.100.1.1=de  
eneen.akeo@va.gov,  
cn=Deeneen U Akeo 622220  
Date: 2017.08.04 16:04:36  
-04'00'

DEENEEN AKEO  
ACTING DIRECTOR, ASD

## REVISION HISTORY

Version	Date	Approver	Notes
1.0	12/15/17	Nicholas Bogden (ASD TS)	Final version for TS leadership approval and signature, including all applicable updates addressing stakeholders' feedback and Section 508 Compliance.
2.0	07/17/2017	Michael Miti- Kavuma (ASD TS)	Document updates include: updated terminology (including "HDA" to "EDA"), streamlined document, incorporated EDA's role in VA's Open Data initiative, updated the TRM table, update VA's database inventory effort status, updated all references of VSA to VIA, incorporated the term data federation and how it can work with the proposed data lake.

## CONTENTS

1	Introduction .....	5
1.1	Business Problem .....	5
1.2	Business Need .....	6
1.3	Business Case .....	6
1.4	Approach .....	7
2	Current Capabilities and Limitations .....	7
2.1	Veterans Information Systems and Technology Architecture Integration Adapter Platform .....	10
2.2	VA Data Access Service .....	12
2.3	Customer Relationship Management (CRM) Unified Desktop .....	13
2.4	Burial Operations Support System Enterprise Database .....	14
2.5	Master Veterans Index .....	15
2.6	VA Data Inventory and Data Architecture Repository .....	16
2.7	Customer Data Integration and the VA Data Architecture .....	16
2.8	Enterprise Information Management Policy .....	17
3	Future Capabilities That Address Current Limitations .....	18
3.1	Pre-Implementation .....	18
3.1.1	Map the VA Data Landscape .....	18
3.1.2	Enterprise Data Management and Governance .....	19
3.1.3	Records Management .....	20
3.1.4	Develop Schema, Metadata, and Transformation Tools .....	20
3.2	Phase One .....	21
3.2.1	Enterprise Create, Read, Update, and Delete (eCRUD) .....	23
3.2.2	The Data Lake .....	23
3.2.3	Change Data Capture for Authoritative Data Sources .....	24
3.3	Phase Two .....	25
3.3.1	Expand Schema to Support New Applications .....	25
3.3.2	Support Additional Types of Data .....	26
3.3.3	Develop Sophisticated Analytics Capabilities .....	26
3.3.4	Consolidate Data into the Lake .....	26
3.4	Enterprise Data Access Constraining Principles and Strategic Guidance .....	27
3.5	Alignment to Veteran-Focused Integration Process (VIP) .....	28
4	Use Cases .....	28
4.1	Use Case 1: Veteran Initiates a Change of Address .....	29
4.1.1	Purpose .....	29
4.1.2	Assumptions .....	29
4.1.3	Use Case Description .....	29
4.1.4	Use Case Context Diagram .....	31
4.2	Use Case 2: Clinician Accesses Veteran's Health Records from Another Region .....	33
4.2.1	Purpose .....	33
4.2.2	Assumptions .....	33
4.2.3	Use Case Description .....	33

4.2.4	Use Case Context Diagram.....	34
4.3	Use Case 3: VA Staffer Checks Veteran’s Status Prior to Outreach .....	35
4.3.1	Purpose .....	35
4.3.2	Assumptions.....	36
4.3.3	Use Case Description .....	36
4.3.4	Use Case Context Diagram.....	37
Appendix A.	Scope .....	38
Appendix B.	Key Concepts.....	41
Appendix C.	Definitions.....	45
Appendix D.	Acronyms.....	50
Appendix E.	References, Standards, and Policies .....	52
Appendix F.	Database Management Systems Used at VA.....	54
Table 1: EDA Constraining Principles and Strategic Guidance.....		27
Figure 1: Illustration of EDA's Position in the VA EA SOA Infrastructure .....		9
Figure 2: EDA and VistA in the SOA Architecture .....		11
Figure 3: Components of the VA EA Data Layer .....		22
Figure 4: EDA Use Case #1 - Veteran Initiates a Change of Address .....		32
Figure 5: EDA Use Case #2 - Clinician Accesses Veteran's Health Records from Another Region .....		35
Figure 6: EDA Use Case #3 - VA Staffer Checks Veteran's Status Prior to Outreach .....		37

## QUICK JUMP

Select an icon to skip to a section.



**Current Capabilities**



**Future Capabilities**



**Use Cases**



**One-VA Technical Reference  
Model**



**The Veteran-Focused  
Integration Process**



**Enterprise Design Pattern  
Scope**

---

## 1 INTRODUCTION

### 1.1 Business Problem

In charting the strategy to achieve the information technology (IT) vision at the Department of Veterans Affairs (VA), the Enterprise Technology Strategic Plan (ETSP) has identified an interoperability gap among the sources and services of common customer data (CCD). CCD is primary among the MyVA business drivers that are aligned to IT investments, VA policies, and the VA Enterprise Architecture (EA) to achieve the VA IT vision. As stated in a 2014 presentation on “MyVA” initiatives by former United States Secretary of Veterans Affairs (VA) Robert A. McDonald, “Assessments informing the [2014-2020] strategic plan told us VA often provides a fragmented, disjointed experience that results in poor customer service and frustrated Veterans and beneficiaries.” The Secretary described how Servicemembers, Veterans, and beneficiaries take on the burden of serving as their own integration point for “multiple VAs” as a result of this data management gap. Veterans and beneficiaries contend with the following issues in their interactions with the Department:

- The need to initiate contact with different parts of VA (through multiple call centers, hotlines, and websites) in order for VA to “know” them and determine discrepancies in how different parts of VA understand them

- A requirement to complete multiple forms to provide the same data multiple times to different parts of VA, without knowing where particular forms should be sent
- An inconsistent visibility into their own data at VA (including benefits/services history), and no single way to “write back” updates, corrections, or feedback to VA data sources
- The receipt of uncoordinated, duplicative, and at times, inappropriate outreach from different parts of VA

## **1.2 Business Need**

The following issues stem from fundamental, systemic problems surrounding data management and sharing at VA (and many other large organizations):

1. VA does not have a complete understanding of the number of databases in the enterprise, where they are located, and a description of the data they contain.
2. Central policies and standards that surround data and databases are lacking. Since siloed data stores cannot “talk” to each other, barriers to data sharing exist.
3. Data silos containing the same data fields (e.g., address fields) require manual updates on an individual basis. Manual updates risk the quality of data, especially in terms of data timeliness and data accuracy.
4. There are no enterprise-wide data quality requirements or “official” definitive data records. Inconsistencies in data quality and data records are difficult to catch and almost impossible to correct.
5. Scattered Veteran records had not been tied to a single VA-wide identity (until recently), making it difficult to obtain a unified view of Veterans across VA.

A business and technical solution that integrates and provides standardized, transparent access to VA’s data stores can help address the data integration issues that are experienced by Veterans and the VA staff who serve them.

The future capabilities are specified in Section 3 to describe how this enterprise design pattern (EDP) contributes to addressing data management problems at VA.

## **1.3 Business Case**

Enterprise Data Access (EDA) is a construct that incorporates Authoritative Data Sources (ADSs), common data exchange standards, and data governance mechanisms to support standardized IT capabilities. EDA will reduce the burden of data integration and manual data management for Servicemembers, Veterans, beneficiaries, and other VA stakeholders.

The following represents how EDA benefits Servicemembers, Veterans, and beneficiaries:

- Supports access to health and benefits information through a variety of web interfaces, applications, and mobile devices, using a single portal
- Allows Veterans to update their information in a single place at a single point in time, rather than requiring them to submit the same information to multiple VA offices
- Ensures that Veterans and beneficiaries receive coordinated services and timely, appropriate outreach from VA

The following represents how EDA benefits VA employees:

- Reduces redundancy, time, and effort to update Veteran records in individual data stores
- Facilitates the enterprise-wide use of designated ADSs by serving as a standardized, scalable, and centrally managed integration point
- Ensures the enterprise-wide availability of accurate and up-to-date records, allowing VA to streamline back-end operations and provide VA employees with additional time and effort to devote to other tasks

#### 1.4 Approach

VA's data layer, with EDA capability, will meet VA's current data integration needs and serve as a foundation for advanced capabilities. These capabilities include processing streaming data from wearable devices and trend analysis and automatic enrollment in VA benefit programs for Servicemembers who are separating or retiring. The implementation of the EDA capability into the data layer will be implemented in two phases:

- Phase One: A data integration strategy that enables sharing common, high-value customer data
- Phase Two: A big data strategy that supports sophisticated service delivery and analytic capabilities, including automatic enrollment of separating Servicemembers in VA benefit programs, and near real-time processing of streaming data from wearables and medical devices



## 2 CURRENT CAPABILITIES AND LIMITATIONS

VA is a large, diverse, geographically distributed organization in which individual lines of business (LOBs), regional offices, and operating units have traditionally developed and maintained their own IT systems.<sup>1</sup> This means that a significant number of systems in VA's inventory consist of siloed databases. Siloed databases are tightly coupled with business

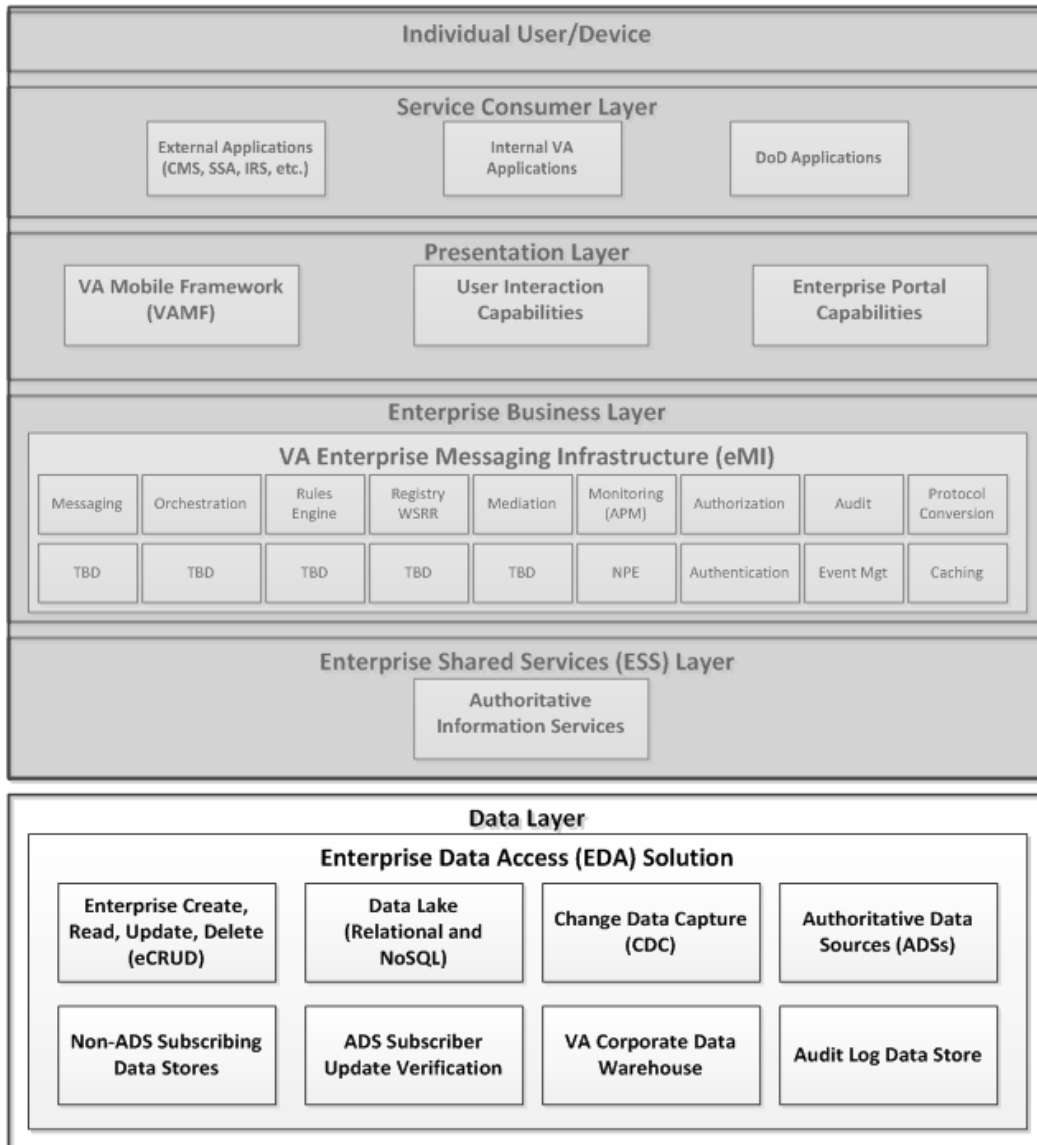
---

<sup>1</sup> For purposes of this document, the word "system" means a FISMA-reportable system, unless otherwise noted.

applications. Siloed databases rely on customized and/or proprietary software that do not anticipate service-oriented computing or shared services.

VA has launched an enterprise-wide effort to transform its IT architecture from a set of stove-piped systems to an integrated, modern, service-oriented architecture (SOA) environment. This effort includes modernizing existing applications and standardizing implementations of future applications, using shared enterprise services and data within the VA Enterprise Architecture (VA EA) SOA infrastructure. This modernization of the current IT architecture will promote data integration and help VA realize its Open Data initiative. EDA will form the data layer of this infrastructure. The simplified VA SOA architecture diagram in Figure 1 focuses on the data layer, and highlights some of its internal components.





**FIGURE 1: ILLUSTRATION OF EDA'S POSITION IN THE VA EA SOA INFRASTRUCTURE**

*Figure 1 Description: The architecture consists of six layers, which are (from top to bottom): Individual User/Device, Service Consumer Layer, Presentation Layer, Enterprise Business Layer, Enterprise Shared Services (ESS) Layer, and Data Layer. Since the presentation focuses on the Data Layer, the layers above it are grayed out. Within the Data is a component called EDA. EDA contains the following capabilities/components: Enterprise Create, Read, Update, Delete (eCRUD), Data Lake (Relational and NoSQL), Change Date Capture (CDC), Authoritative Data Sources (ADSs), Non-ADS Subscribing Data Stores, ADS Subscriber Update Verification, VA Data Warehouses, and Audit Log Data Store.*

Within the data layer, EDA will aggregate and standardize data from across VA. EDA will support a transparent connection point that provides systems, applications, and users with on-demand

access to VA's enterprise data. EDA will also simplify and accelerate application development by providing a standardized portal for data access.

The sections below highlight efforts to address data silos and management issues at VA, including descriptions of specific issues and where the solution effort contributes to this EDP.<sup>2</sup> It is important to note that some of the problems described are within historical cases. Some of the efforts to address historical cases are notional or in development.

## **2.1 Veterans Information Systems and Technology Architecture Integration Adapter Platform**

The Veterans Information Systems and Technology Architecture (VistA) is VA's largest IT system. VistA supports a variety of core VA functions related to healthcare, financial transactions, and benefits. In addition, it is the primary data store for Veterans Health Administration (VHA) patient records and the back-end for the Computerized Patient Record System (CPRS).

Instances of VistA are siloed, based on the VA Medical Center (VAMC) in which they are physically housed. The CPRS allows clinicians to have read-only access to patient medical records from a different VistA instance. If clinicians want write access to patient records from a different VistA, they must use a different set of applications than the ones they use on a daily basis. Unfortunately, many clinicians do not complete the training necessary to use these applications, making patient data difficult to access as they change VAMCs.

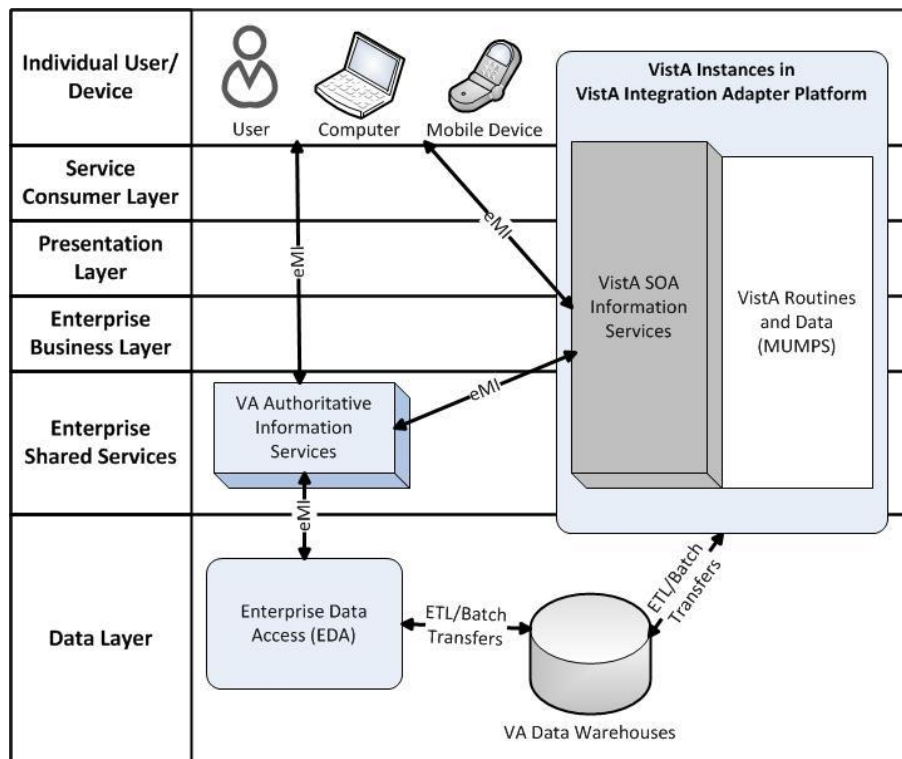
This means that while the capability exists to move Veteran medical records from one VistA to another, that capability is rarely, if ever, used. Veterans are responsible for transferring their own records between VistAs, either as physical copies, or as electronic files, using My HealtheVet's "Blue Button" functions to send the Portable Document Format (PDF). As with records from other VistAs retrieved using CPRS, Veteran health records must be manually entered into the target VistA to allow read/write access. This manual input impedes clinician access to important health information and creates significant potential for error and information loss.

To solve this problem, VHA is transitioning its VistA instances to SOA architecture.<sup>3</sup> The VistA Integration Adapter (VIA) platform that is currently in development will encapsulate VistA instances within one interface. The VIA platform gives VA clinicians access to data in any VistA instance through the same set of applications they use on a daily basis, regardless of their physical location.

---

<sup>2</sup> This does not represent a comprehensive list of all VA data integration and/or management solutions.

<sup>3</sup> For additional details on this effort, refer to the VistA Evolution SOA EDP.



**FIGURE 2: EDA AND VISTA IN THE SOA ARCHITECTURE**

Figure 2 Description: The Veteran, or a VA Staffer acting on the Veteran's behalf, initiates a change of address through a web application to reflect the Veteran's move from the Washington, DC area to the San Francisco, CA area. A data message is sent from the application through authoritative information services down to the data layer. This data message contains the following elements: The Identity and Access Management (IAM) security ID of the user responsible for the address change (Veteran or VA Staffer); Veteran's Master Veterans Index (MVI) Integrated Control Number (ICN); Veteran's new address; Veteran's previous address; when the new address record was created, and whether it was through an internal or external user application. The message is processed by the eCRUD service, which writes the address change to the data lake and also to the ADS for Veteran addresses. The change of address in the data lake is detected by change data capture (CDC) tools, triggering multiple preprogrammed actions (described below). The change of address is replicated to target data stores that subscribe to updates of master records in the ADS for Veteran addresses. All but one of the subscribing data stores sends an acknowledgement to the data lake indicating that the address update was received and written successfully. The data lake sends a second update message to the target data store that failed to acknowledge the first message, and waits for a reply. The data store at issue does not acknowledge the second update message, so the data lake sends an alert to the owners of the subscribing data store and the data lake stewards informing them of the issue. The data lake sends instructions to automated outreach systems to send follow-up physical and/or electronic email to the Veteran that a) gives the Veteran information on the VA Health and Benefits centers closest to his/her new address (along with links to their websites; b) informs the Veteran that his/her dependents (listed individually by name) will have their contact information changed to match the Veteran's and c) Supplies contact information for the appropriate VA offices the Veteran can reach out to if he/she has any issues associated with his/her move to a new address. The data lake ensures that the Veteran's change of address is extended to the Veteran's dependents by taking the following actions: Queries BIRLS for the

*MVI ICNs of dependents associated with the Veteran (e.g., spouses, children); Upon receiving requested ICNs from BIRLS, queries its own database for records that: Have the ICNs returned by BIRLS; Have a physical address and home phone number matching the Veteran's previous address and phone number; Updates the contact information of the Veteran's dependents in its own data stores and sends the update to the appropriate ADS and ADS subscribers. This time it also expects an acknowledgement from the ADS. It receives acknowledgements from all subscribers except for the one that was experiencing problems before. Uses authoritative information services to notify VistA that the Veteran's address has changed and he/she and his/her dependents should be aligned to a new VAMC on the basis of his/her new ZIP code.*

This encapsulation means that VistA data stores will not be integrated into the VA EA data layer, but will be accessible through shared authoritative information services. Figure 2 illustrates the relationship between the data layer, the encapsulated VistA instances, and shared authoritative information services in the VA EA framework. Both the data layer and VistA feed directly into VA data warehouses.

## **2.2 VA Data Access Service**

VistA alone cannot fulfill all of VHA's patient-related data needs. It is not flexible enough to store all of the VHA data varieties, including the following:

- Electronic case files from the Department of Defense (DoD)
- Unstructured document data (e.g., questionnaires, scanned treatment records)
- Patient-generated data (PGD) from Veteran-facing applications

Similarly, VistA is not scalable or malleable enough to serve as the foundation of a SOA application stack while supporting the functionality that Veterans expect from web and mobile applications.

VHA implemented the Data Access Service (DAS) to serve as the back-end to its own SOA architecture to address these gaps. The core of DAS is a group of virtualized, scalable Not Only SQL (NoSQL) data stores that contain data in both structured and unstructured formats.

The data stores are encapsulated in a single logical object by an eCRUD logical wrapper. The eCRUD provides and mediates access through standardized, secure Representational State Transfer (REST)-based Application Programming Interfaces (APIs). It is responsible for data ingest, access control, data persistence, logging, and auditing functionality for the DAS data stores. Other VA programs are developing or have already implemented their own versions of DAS and its related components (including eCRUD).

The proposed VA EA data layer with the EDA capability draws many elements from DAS, particularly eCRUD and the MongoDB data stores (the latter can be directly incorporated into

the data lake). The development and operation of DAS also provides best practices for using and accessing a NoSQL data store.

### **2.3 Customer Relationship Management (CRM) Unified Desktop**

As a service-oriented organization, VA has a large number of call centers belonging to various operating units within different LOBs, while sharing common business processes and resource requirements. There are many scenarios in which a Veteran or beneficiary may need to interact with multiple call centers in different LOBs to address a complex problem, such as applying for disability benefits.

Historically, VA call centers have operated in isolation from each other, leading to a variety of issues that negatively impacted internal VA operations and service delivery to Veterans:

- Redundant expenditures of cost and effort on similar data stores, IT infrastructure, and applications for individual call centers
- Decreased visibility into a complete Veteran customer service history, creating delays and difficulties in resolving Veteran issues
- Lack of communication and information sharing between call centers, including those in the same line of business, creating scenarios in which Veterans were required to provide the same information multiple times

To address these issues, the Veteran Relationship Management (VRM) program developed the Customer Relationship Management (CRM) Unified Desktop, based on the Microsoft Dynamics CRM application. The CRM integrates desktop applications to support capabilities that include the following:

- A single consolidated view of the Veteran/Beneficiary for VA call center staff, ensuring that all staff have access to up-to-date information on callers who seek assistance
- Ability to collect, preserve, and track the context of all interactions with Veterans and beneficiaries, giving VA customer service representatives immediate access to details on caller issues and efforts made to address them
- Access to a set of shared resources, significantly reducing duplicative effort and cost

While the unified desktop was developed and implemented by the Veterans Benefits Administration (VBA), the solution is used by VA call centers across all LOBs.

CRM's information flow models and business processes can be a source of best practices and lessons learned for the proposed data layer. The CRM Reference Architecture designates ADSs (beyond those already named by VA) for particular elements of Veteran and beneficiary data,

providing criteria for selecting ADSs to be used in an enterprise-wide solution.<sup>4</sup> Finally, the forthcoming CRM Enterprise (CRMe) release includes a data integration layer that will facilitate Veteran data access for customer service purposes, a role that the EDA capability can serve.

## **2.4 Burial Operations Support System Enterprise Database**

The National Cemetery Administration (NCA) maintained exclusively paper-based records of burial locations and memorial inscriptions up until the early 2000s. These records were scattered across more than one hundred national cemeteries, monument sites, and regional offices. Similarly, NCA's business processes were entirely paper-based and manual. The manual processes included those that handled memorial and burial applications and provided information on where particular individuals were buried. NCA needed a database solution to consolidate, organize, and provide easy access to all its records for both internal and external users.

NCA's solution is the Burial Operations Support System (BOSS) Enterprise Database, a virtualized Oracle data store that is hosted in the Quantico Information Technology Center. Due to resource constraints, the staff responsible for developing and maintaining the NCA systems found it necessary to modify and expand NCA's single data store. The purpose of the expansion was to support new applications and capabilities, rather than setting up additional databases. As a result, NCA effectively uses its BOSS Enterprise Database as a shared data service and does not suffer from data silo issues found in other LOBs at VA. In addition, since all NCA processes and applications use the same set of data, NCA has been able to implement many automatic or "pass-through" processes. Data output from one workflow triggers and provides input to another workflow.

The BOSS Enterprise Database serves as a model of practices that can be implemented in the data layer, such as:

- Ensuring the availability and consistency of data for a geographically distributed population of internal and external users
- Leveraging database virtualization for scalability, flexibility, performance improvements, and continuity of operations planning (COOP)
- Using a shared data store to support pass-through processing and automated functions

---

<sup>4</sup> Refer to the Microsoft Dynamics CRM Platform Reference Architecture v5.1 for further details.

## 2.5 Master Veterans Index

Historically, VA did not have standardized enterprise-wide unique identifiers for individual Veterans and beneficiaries. That is, there was no identifier to serve as an organizing principle for individual records across multiple siloed VA data stores. Under those circumstances, identity confusion and subsequent recordkeeping errors were common. It was also impossible for VA to support a secure single sign-on (SSO) capability for external users that would have provided access to all VA applications.

The Office of Information and Technology (OI&T) addressed this problem by adapting the VHA Master Patient Index (once part of VistA) into an overarching Master Veterans Index (MVI). The MVI acts as the mandated ADS for identity information that is related to over 22 million Veterans, beneficiaries, and other persons of interest to VA. Each person is assigned a globally unique MVI Integrated Control Number (ICN) that corresponds to the “Primary View” or definitive “gold copy” of their identity data. Identities in MVI form the basis of the user accounts that VA external users employ for SSO to all VA applications.

VA applications are mandated to integrate with MVI and use it to locate and retrieve records associated with a Veteran or beneficiary. The MVI Identity Service (IdS) correlates a patient’s identity across the VA enterprise. The MVI IdS broadcasts identity trait updates to systems for which the identity record is correlated. The VHA Data Quality Healthcare Identity Management (HC IdM) program is MVI’s data steward. The HC IdM program is responsible for ensuring the quality and correctness of MVI’s data and for performing support and maintenance activities, such as establishing interoperability with legacy records.

The MVI underlying system was not designed to support Enterprise Shared Services (ESS). This necessitated new interfaces to MVI, using a set of adaptable integration design patterns. The interface owner must coordinate with the owners of other VA systems and applications. Coordination can be challenging when both parties have conflicting technical, business, and security requirements.

This issue has been mitigated for future IT projects through VA’s requirement for MVI integration. It is expected to be addressed early in the design phase of new IT projects.<sup>5</sup> The adaptable integration patterns for MVI can be leveraged for the proposed data layer. The standards and processes used to ensure data quality for MVI can be applied to the EDA data lake. The documentation of challenges experienced by VA owners when integrating technical, cultural, and business processes with MVI can provide a valuable resource. By highlighting the

---

<sup>5</sup> This requirement was established in 2010.

potential pitfalls of implementing a shared service, VA owners can be adequately be prepared for integration with the existing technology and business environment.

## **2.6 VA Data Inventory and Data Architecture Repository**

VA does not currently have a single comprehensive inventory of its data stores, their contents, and their physical schemas. A data call conducted in the second half of 2016 indicated that VA has a comprehensive inventory of about 20% of its systems. Without complete knowledge of its data assets, VA is unable to develop or implement feasible, effective mechanisms for central data management or enterprise data integration. To facilitate its enterprise data and SOA development efforts, VA OI&T's EA group is launching an effort to develop an enterprise-wide data inventory and create the first complete map of the VA data landscape.

The VA EA group is expanding the VA Enterprise Architecture Repository (VEAR) to include an authoritative VA Data Inventory (VADI), building upon the existing VA Systems Inventory (VASI) and the VA Data Architecture Repository (DAR). VASI is a searchable enterprise-wide authoritative systems inventory that includes information on the databases that is associated with each system. The DAR, whose content includes database designs for a subset of existing VA databases, is planned to be absorbed into the VEAR. The DAR includes VistA, the Administrative Data Repository [ADR], Beneficiary Identification Record Locator System [BIRLS], VBA Corporate Database [CORPDB], and BOSS. Current DAR capabilities include a VA-wide metadata catalog that allows users to search, report, and manage characteristics of data and data stewardship across platforms.

The VADI will contain a complete list of VA data stores, their physical schemas, and their relationships to data model concepts and data stewardship metadata. The VADI will help stakeholders understand the persisted data environment by enabling the following capabilities:

- Identify authoritative data stores
- Reduce redundant data storage
- Promote data standardization
- Facilitate enterprise data integration and centralized data management

## **2.7 Customer Data Integration and the VA Data Architecture**

Certain Veteran and beneficiary data are used throughout VA and across LOBs (e.g., name, date of birth, Social Security number, address, military service history). While this data can be used across the enterprise in similar ways, the data fields are not formatted in a consistent manner, making it impossible to share the data.



OI&T is taking steps to make sharing possible, starting with the Customer Data Integration (CDI) initiative. The CDI initiative addresses the lack of consistent, reliable, and authoritative data that exists for Veterans and their beneficiaries. The CDI will leverage existing efforts and strategies, including DAR, to provide authoritative Veteran data in cases where information is captured once and used many times. CDI will enable VA to maintain authoritative common data on each customer and securely share the data across the enterprise. The CDI initiative reduces the burden to the Veteran, their family, and VA staff, and allows VA to dramatically improve its delivery of benefits and services. CDI has developed a formal inventory of CCD elements that are sorted into four categories: Identity, Military Service History, Contact Information, and Demographic and Socioeconomic Data.

The VA Data Architecture Framework provides an integrated approach to describe, understand, address, and align VA data. The Framework is comprised of conceptual, logical and physical data models. The VA Conceptual Data Model (CDM) categorizes and defines VA-wide common data concepts, and the data concepts that are unique to each LOB. VA's Enterprise Logical Data Model (ELDM) documents data entities and their relationships and attributes, extending the description of the information identified in the VA CDM v1.0. The ELDM, along with the CDM, informs and promotes VA data requirements and consistency with each LOB's data concepts. The EA Office will work with the CDI team to support the analysis of enterprise-wide standards around common VA customer data. The goal of this effort is to use the ELDM as the basis for enterprise-wide standards around common VA customer data, which will inform the design of a technical customer data sharing solution.

One of the purposes of the EDA EDP is to describe how the ELDM will become operational. It will describe how the ELDM will be used as the common VA model for formatting and structuring the common high-value customer data that is identified by CDI. The use of schema-on-read functionality in the proposed solution will make it possible to use defined and approved ELDM entities and attributes in a production environment (with their associated business rules), as they become available. At this time, considerable development is required before the ELDM is mature and sufficiently articulated in order for it to be used as the basis of operations in the data layer.

## **2.8 Enterprise Information Management Policy**

As the authoritative source for customer identity data, MVI stands alone as the official data source for Veteran and beneficiary data. The MVI is also the official source for providing enterprise-wide quality data standards. VA needs to select and mandate the use of master records for customer data to enable data sharing and to establish policies for maintaining master record quality.

To accomplish this, the VA Assistant Secretary for Information and Technology signed the Enterprise Information Management (EIM) policy in early 2015. EIM will drive efforts to establish ADSs for the entities defined in the ELDM. The policy requires VA LOBs to coordinate with OI&T to designate additional ADSs and data stewards to maintain them. The policy has not yet been implemented, and a timeline for implementation has not been established.

The EIM policy creates the necessary preconditions for an enterprise-wide data solution, while EDA will become a mechanism for enabling compliance with EIM within the data layer. The mechanism by which EDA will publish ADS records to subscribed VA data sources is described in Section 3.2.3.



### **3 FUTURE CAPABILITIES THAT ADDRESS CURRENT LIMITATIONS**

As described in Section 1.4, the approach to build the EDA capability into the VA EA data layer will be accomplished in two phases. The capacity to reach Phase One, and subsequently evolve to Phase Two, will depend on the data lake's ability to scale and store unstructured data, and support "schema-on-read" functionality. A full description of the required capabilities for the data lake is provided in Section 3.2.2.

Aspects of the data layer and EDA are described in Sections 3.1 and 3.2 and are each tied to the relevant data problems that are outlined in Section 1.1.

#### **3.1 Pre-Implementation**

In order to ensure that the proposed EDA capability and the data layer provide sufficient return on investment, the following issues must be addressed prior to deploying the solution into production:

- Develop a comprehensive inventory of VA data stores, their contents, and their schema
- Implement the EIM policy across the enterprise
- Develop enterprise-wide records management strategies for:
  - Enterprise data retention and persistence
  - Creation, management, and retention of audit logs
- Establish common enterprise metadata, schema, and related tools to support data access and consumption by EA enterprise services

##### ***3.1.1 Map the VA Data Landscape***

A full map of the data landscape must be built to support the enterprise data solution. VA requires an inventory of data stores detailing the owner, location, current database

management system (DBMS), and data types being stored. VASI contains a partial list, which VADI will build on to establish a complete record of VA data stores.<sup>6</sup> The VADI will also clearly indicate encapsulated (i.e., federated or virtualized) groups of databases.

EDA and the data layer will leverage the VADI, which will initially focus on data stores containing customer data so that these stores can subscribe to ADSs. This does not mean that every data store in the inventory will need to be integrated. If a system is slated for transition/decommissioning, or the system uses a DBMS that is not approved by the One-VA Technical Reference Model (TRM), it may be practical to decommission the system.

## **VA Data Problems Addressed: #1**

### ***3.1.2 Enterprise Data Management and Governance***

VA needs formalized governance processes and modeling concepts for enterprise shared data to shape the operations of the data layer. Otherwise, integration in the data layer would accomplish nothing more than automating chaos. Fortunately, some of the governance and modeling elements that are necessary to make the data layer function are already in development at VA.

One such element is the EIM policy that is mentioned in Section 2.8.<sup>7</sup> A full inventory of data elements will guide the Data Governance Council (DGC) in identifying, evaluating, and selecting potential ADS candidates. Once ADSs have been designated, VA must coordinate with data stewards to establish enterprise business requirements and rules for CCD.

The Enterprise Architecture Council (EAC), in partnership with DGC, must further develop the ELDM by defining entity attributes, business rules, and associations. Definitions are completed before the ELDM can be used as the basis for operations in the data layer. In addition, data obligations (i.e., rules around eCRUD for particular types of data elements) must be articulated to be operationalized and enforced in the data layer. The VHA Business Information Architecture (BIA) program has a large collection of tools, repeatable processes, and best practices that can be leveraged to further develop the ELDM and its related business rules and obligations. The DGC should establish the priorities for the evolution of the ELDM to meet operational requirements.

## **VA Data Problems Addressed: #2, #4**

---

<sup>6</sup> A list of DBMSs currently in use at VA, originally compiled by the One-VA TRM team, can be found in Appendix F.

<sup>7</sup> The EIM policy mandates that VA formally designate ADSs that will host the master records for common customer data (CCD) elements (as mapped and defined by the ELDM).

### **3.1.3 Records Management**

A records management strategy is essential for governing and managing an enterprise data solution. Common enterprise customer data (defined by the ELDM and published by ADSs to subscribers) will persist in the data lake. Other types of data can be retained if there is a valid business case. Any type of data that is not explicitly targeted for persistence in the data lake will be overwritten.

A records management strategy for the creation, use, and retention of logs is required. Lessons learned from DAS demonstrate that the rapidly growing volume of log files will become unmanageable and prohibitively expensive to store without such a strategy. A practical and effective strategy answers the following questions:

- Why is the organization conducting audits (e.g., performance monitoring, security)?
- What are the requirements for audit information?
- What utility will the organization realize from its logs (i.e., what is planned for them and what benefit will the actions provide)?
- When does a log qualify as an official record, personally identifiable information (PII), or protected health information (PHI)? In such cases, what are the legal requirements for retaining/protecting the log?
- What are the forensic requirements for security logs? What are the requirements for logs to serve as evidence in a court of law?
- If a security breach occurs, what is the anticipated duration of time that is necessary for security log records to be maintained and archived in order for organizational security personnel to be able to discover the root cause of the breach?

The records management strategy will address archiving for both operational data and log data. The trigger for moving data from “hot” on-demand storage to archival storage (e.g., time-based vs. event-based) will differ. The differences depend upon such factors as the particular type of data and how it is used.

## **VA Data Problems Addressed: #2**

### **3.1.4 Develop Schema, Metadata, and Transformation Tools**

When ADSs, enterprise business rules, and auditing policies have been established, VA can begin to develop a common enterprise schema for customer data that is based on approved entities and definitions from the ELDM. Because the data lake will use a schema-on-read approach, it is feasible to use one or more “starter schema” and make incremental modifications as needed, with little or no disruption.

VA needs to develop metadata that is consistent with its business rules for enterprise data. The metadata will provide the capability to attach provenance and lineage information to each data record and/or message. The metadata will also include additional information that is needed to support logging, auditing, and performance monitoring.

VA will need to configure the transformation capabilities of the data layer. EDA will require the capacity to transform data from any format used in the VA enterprise to any other, including a common enterprise data format that is consumable by all VA SOA applications.

All of the above elements – the schema, metadata, data lake, and transformation capabilities – can be prototyped and refined before production in a secure, easily configurable cloud environment, using historical and anonymized warehouse data.

### **VA Data Problems Addressed: #2, #3**

#### **3.2 Phase One**

The results of the pre-work described in Section 3.1 will inform the configuration and implementation of the data lake and its associated components. Phase One will accomplish the following goals:

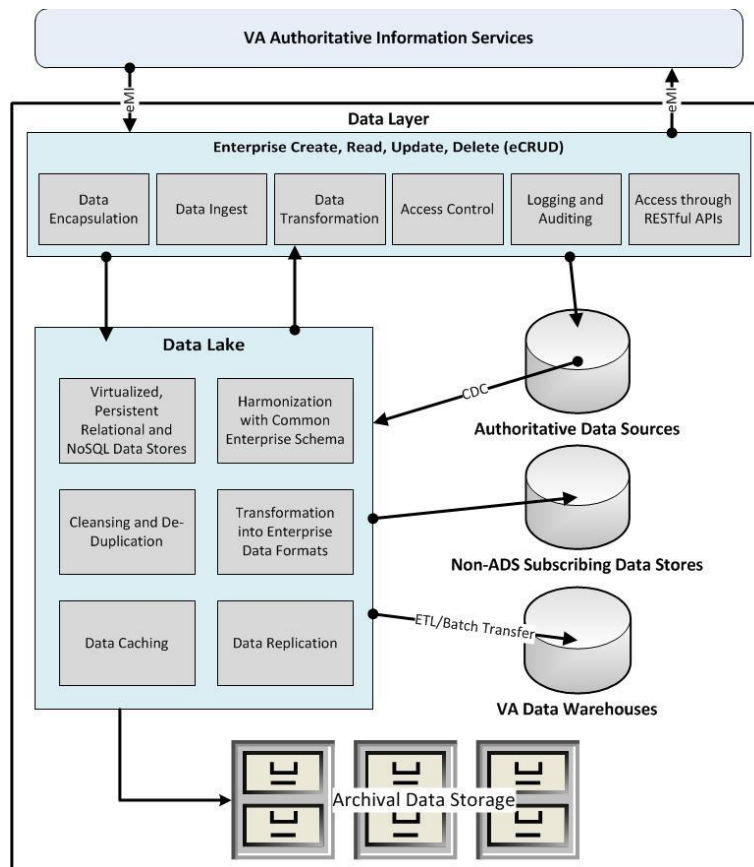
- Implement the EIM policy by publishing customer records from designated ADSs to designated subscribers in the VA enterprise. Provide all VA users (internal and external) with access to current, correct Veteran and beneficiary records. Through eCRUD, provide a single common access point for enterprise data to be used by VA ESS, SOA applications, partner organizations, and authorized third-party applications.
- Implement the necessary capabilities for authorized internal and external users to easily view, update, and provide feedback on data relating to a particular Veteran or beneficiary.
- Support the continued operation of existing VA data stores (including data warehouses) and their associated applications and systems.

The three key components of the data layer will provide the following capabilities:

- eCRUD
  - Ingest for the data lake
  - Provision of unified, standardized, secure read and write-back access to the VA data layer (specifically the data lake and ADSs) through RESTful APIs
  - Auditing and logging for operations in the data layer
- The Data Lake

- Aggregation and persistence of enterprise data in multiple formats
- Publication of master records from ADSs to subscribing data stores
- Harmonization of data with established common enterprise schema based on ELDM
- Provision of data to consuming VA enterprise services
- Change Data Capture
  - Detection and propagation of updates to master records in ADSs
  - Triggering of automated workflows based on creation of or updates to records in VA data stores

Figure 3 illustrates the components of the data layer and the flow of information between them, including ADSs, non-ADS data stores, VA data warehouses, and archival data stores.



**FIGURE 3: COMPONENTS OF THE VA EA DATA LAYER**

*Figure 3 Description: Figure depicts the interaction of VistA and the Data Layer/EDA in the proposed VA EA. The layers of the VA EA are, from top to bottom: Individual User/Device, Service Consumer Layer, Presentation Layer, Enterprise Business Layer, Enterprise Shared Services, and Data Layer. There are five components of interest in the diagram: Users, computers, and mobile devices in the Individual User/Device layer (top); VistA Instances in the VistA Integration Adapter Platform (right), which include*

*VistA SOA Information Service and VistA Routines and Data (MUMPS); VA Authoritative Information Services in the Enterprise Shared Services layer (middle); Enterprise Data Access (EDA) in the Data Layer (bottom); and VA Data Warehouses (bottom). Users, computers, and mobile devices in the top layer send messages to and receive messages from both the VistA instances and VA Authoritative Information services via the Enterprise Messaging Infrastructure (eMI). The VistA Instances also communicate with the VA Authoritative Information Services via eMI in order to exchange data/messages with EDA in the Data Layer. Warehouses in the Data Layer also receive batch transfers from both the VistA Instances and EDA. In sum, the VA Authoritative Information Services mediate user and device access to the Data Layer. They also mediate interactions between VistA systems and operational systems in EDA.*

### **3.2.1 Enterprise Create, Read, Update, and Delete (eCRUD)**

Although there are multiple versions of eCRUD used in various SOA application stacks within VA, the canonical, definitive version is associated with the DAS data store that is owned by VHA. Currently, eCRUD is used on a relatively small scale and is specifically intended for use with a MongoDB data store, but it can potentially be scaled up, extended, and otherwise modified to serve as the provider and mediator of access to an enterprise data lake.

eCRUD will perform the same key functions as it does for DAS, such as data ingest, logging/auditing, enforcing access control policies, and queuing data messages. By exposing the data lake and ADS data stores as standardized, RESTful APIs, eCRUD will facilitate read and write-back access to the data layer for VA information services, VA partner organizations, and Veteran-owned third-party applications. VA will also utilize these standardized RESTful APIs to make VA datasets accessible and machine-readable, as part of VA's Open Data initiative. These standardized RESTful APIs must be compliant with OMB M-13-13 (Open Data Policy – Managing Information as an Asset).

eCRUD will perform data cleansing and de-duplication on write operations from these sources in order to maintain data integrity for the data lake and ADSs. eCRUD will also handle writes that are “internal” to the data layer (i.e., from ADSs to the data lake). The collection of adaptable interface patterns created by VA Identity Services Integrated Technical Team (Ids ITT) for MVI is planned for use as a model for interfaces between eCRUD and ADSs.

### **VA Data Problems Addressed: #3, #5**

#### **3.2.2 The Data Lake**

To provide sufficient scalability, maximize availability, and ensure continuity of operations, each instance of the data lake will be virtualized on clusters of co-located machines. There will be at least two instances, each in different geographic locations. The data lake will have scalable and extensible storage capacity to accommodate any structured or unstructured data from authorized sources. It will contain local caches of frequently used data from preexisting VA data

stores, particularly ADSs. The data lake will support persistent virtualization of additional databases, particularly virtualized instances of databases from systems that have been decommissioned and retired from service. It will have sufficient capacity to process high volumes of incoming and outgoing data messages at high speed.

The data lake will serve as the (logical) halfway point between data stores in the enterprise. As with eCRUD, the collection of adaptable interface patterns created by Ids ITT for MVI will be used as a model for data lake interfaces to other VA data stores. Where there are existing database integration solutions (such as virtualized or federated database, or data federation), the data lake will by default be connected to their designated interfaces, rather than to the individual data stores that compose them. Through these connections, the data lake will replicate published ADS updates to subscribing data stores and evaluate for possible batch transfers to VA data warehouses.<sup>8</sup>

In addition, the data lake will read newly ingested and loaded “raw” input, using transformation tools to harmonize it, if possible, with the common enterprise schema (singular or plural). The harmonized data will be made available for consumption by enterprise SOA services.<sup>9</sup> This data can also be transformed and replicated to subscribing data stores. The data lake will include cleansing and de-duplication capabilities (or at a minimum, the capacity to add those capabilities) to support large-scale processing, merging of data sets, and quality assurance for any data stores that are hosted exclusively in the data lake.

While the data lake will perform many functions in VA’s enterprise data solution, it will not be used as a data warehouse (i.e., to support analytics). It will instead replicate data (raw, transformed, or both, as desired) to existing data warehouses for analysis. Likewise, audit logs for the data layer as a whole will be stored separately from the data lake.

## **VA Data Problems Addressed: #2, #3**

### ***3.2.3 Change Data Capture for Authoritative Data Sources***

Changes to definitive ADS records by SOA applications will be routed through eCRUD. Legacy applications that are directly tied to existing data stores will not use eCRUD (they cannot be modified to use an enterprise SOA infrastructure), but will still be supported by the data layer. If a legacy application (as opposed to a SOA application) writes to an ADS, the resulting change is detected by a change data capture (CDC) tool, which will trigger replication of the update to a

---

<sup>8</sup> This will require coordination with the owners of those warehouses.

<sup>9</sup> Any data cleansing and de-duplication conducted within VA’s enterprise data solution must be consistent with applicable data policies (e.g., those related to the legal Electronic Health Record).



virtualized “mirror” of the ADS in the data lake.<sup>10</sup> Changes in the ADS mirror will trigger additional replication to subscribing systems, including transformation of the updated data into the format used by those systems.

Subscribers to ADS updates will have the capability to verify and validate ADS updates sent to them from the data lake, and also to indicate whether or not those updates were successful. If the data lake pushes an update to a subscriber and does not receive an acknowledgement of the update and/or the acknowledgement indicates the update was not written correctly, the data lake will attempt to push the update again. Repeated failures will trigger an alert to the owners of both the data lake and the subscribing system at issue.

Using CDC tools to route updates from ADSs to the data lake will introduce some delay in publishing those updates. However, it will make it possible for ADS owners to offer their data as a shared enterprise resource, without taking on the burden of turning their own systems into shared services, as IdS ITT had to do with MVI. ADS owners can simply ensure that they have a working interface to one data store – the data lake – instead of establishing and maintaining interfaces to dozens of subscribers across VA. Similarly, the owners of subscribing systems can connect to the data lake and rely on it as their source for ADS updates, rather than having to integrate directly with the ADSs.

## **VA Data Problems Addressed: #3, #4, and #5**

### **3.3 Phase Two**

#### ***3.3.1 Expand Schema to Support New Applications***

The use of schema-on-read will allow adding new data elements to the common enterprise schema as needed, with minimal time, effort, or disruption. However, adequate change management processes must still be applied to schema updates. Testing of proposed updates, prior to their implementation in production systems, is strongly recommended.

The stewards for the data lake will coordinate with the ELDM team to ensure that the data layer’s common enterprise schema is consistent with the most recent release of the ELDM. When a new version of the ELDM is released, corresponding updates to the common enterprise schema will be made within a reasonable amount of time. If the data lake stewards perceive demand for changes to the common enterprise schema, they will work with the ELDM team to

---

<sup>10</sup> The same security policies enforced in an ADS or other data store (particularly with regards to access control) must be enforced for its mirror and/or caches in the data lake.

incorporate those changes. The common enterprise schema in the data lake will always be “synchronized” with the ELDM.

### ***3.3.2 Support Additional Types of Data***

While the initial purpose of the data layer is to facilitate sharing of and access to CCD, it can (and should) eventually support other types of data as well. This will require expanded data transformation capabilities and new capabilities within the data lake itself (e.g., a terminology service for clinical content mapping and harmonization within the data lake).

### ***3.3.3 Develop Sophisticated Analytics Capabilities***

As the number of VA data producers and consumers connected to the data layer increases, eCRUD can be scaled to meet demand. It will need to be modified and extended to support new, advanced capabilities, such as processing streaming data from medical devices and performing near-real-time analytics.

### ***3.3.4 Consolidate Data into the Lake***

In the long term, VA will use the data lake and/or the underlying virtual platform to consolidate and virtualize its (non-VistA) operational databases. In the nearer term, the data lake can serve as a long-term archive for static data from legacy databases that are still in operation. The data lake can continue to support the non-SOA applications that are connected to them, with a minimal amount of dedicated infrastructure.

Ultimately, legacy data stores (or at least the datasets they contain) will be transitioned entirely into the data lake itself. Any new operational VA data stores will be created there (unless they have technical, security, or other requirements that cannot be met by the data lake). This means that the data lake can potentially become an ADS by hosting a set of persistent “master record” data.

Stewards of data stores that have been transitioned or created in the data lake will have the access rights and capabilities necessary to perform their duties. In particular, stewards of data stores will maintain the quality of the data supporting their business processes. Stewards will modify their individual data store as required, as long as the effects of their modifications do not impact the functioning of the data lake and/or data layer as a whole. Access control and other security policies will continue to be enforced, just as they were in the original database. In that respect, the data lake will have similarities to the NCA BOSS enterprise database. That is, the underlying DBMS, platform, and infrastructure are managed by OI&T, but users in NCA offices and cemeteries have full, transparent access to the data itself.

Consolidating the VA data stores into a virtual environment will facilitate central management and governance, while giving system and application owners the flexibility needed to generate and use data for business processes. It is also compatible with VA's long-term strategic goal of transitioning to a cloud-based technology environment.

### 3.4 Enterprise Data Access Constraining Principles and Strategic Guidance

The following table summarizes the constraining principles associated with EDA/data layer concepts and architectural attributes. These principles will be used to drive implementation guidance for the EDA capability, the data layer, and any components (services, applications, and data stores) that will integrate with it.

**TABLE 1: EDA CONSTRAINING PRINCIPLES AND STRATEGIC GUIDANCE**

#	Data Layer Concept/Attribute	Principles & Strategic Guidance
1	Data Access	Enterprise data will be accessible through non-proprietary protocols and communication methods.
2	Data Access	The data layer will support the use of dynamic schema (also called "schema-on-read") to support extensibility and fault tolerance.
3	Data Access	Proposed modifications to common enterprise schema will be tested before being implemented in a production environment.
4	Data Aggregation	The data layer will provide functions such as data aggregation, data de-duplication/rationalization, and data synchronization through the use of the data lake, eCRUD, and CDC tools.
5	Data Aggregation	All common enterprise schema and business processes used within the data layer will correspond with OI&T's ELDM.
6	Data Aggregation	The underlying platform of the data lake within the data layer will support virtualization and continued use of databases from systems being transitioned/divested.
7	Data Aggregation	Development and sustainment efforts associated with the data layer and its components will encourage and facilitate the consolidation and standardization of both new and existing data stores.
8	Data Aggregation	The data lake will not be used as a data warehouse for historical analytics purposes, nor will it contain security/audit logs or archives: it is solely intended for use as an operational/transactional data store.

#	Data Layer Concept/Attribute	Principles & Strategic Guidance
9	Data Security	Upon the creation of records or collection of records from joins, federation, or aggregation, the security and access levels will raise to the level of the data element with the highest security classification.
10	Data Security	Implemented data services and enterprise data stores will provide full, end-to-end FIPS 140-2-compliant encryption modules to ensure that data is encrypted both in transit and at rest.
11	Data Security	Data layer and data lake security, including database security and exception handling, will comply with VA 6500 Handbook, National Institute of Standards and Technology (NIST) and Defense Information Systems Agency (DISA) Security Technical Implementation Guide (STIG) security policies.
12	Data Security	At the level of PGD encryption (i.e., PII/PHI), data at rest and in transit will comply with HIPAA controls in addition to VA 6500 Handbook, NIST, and DISA STIG security policies.
13	Data Security	Within the data layer, the data lake will only accept data messages from, or replicate to, predetermined and explicitly authorized (i.e., whitelisted) systems and software.
14	Scalability	The data layer and data lake will provide a high level of availability for service consumers and facilitate outward scalability.



### 3.5 Alignment to Veteran-Focused Integration Process (VIP)

All mobile solutions that are subject to the Veteran-Focused Integration Process (VIP) will use approved products from the One-VA TRM; following VIP guidelines, all mobile devices will leverage the approved EMM solution. VIP is a Lean-Agile framework that unifies and streamlines information technology (IT) delivery oversight; VIP will deliver IT products more efficiently, securely, and predictably. VIP is the follow-on framework from the Project Management Accountability System (PMAS) for the development and management of IT projects. VIP will propel the Department with even more rigor toward Veteran-focused delivery of IT capabilities.



## 4 USE CASES

The three (3) use case scenarios in this section involve user access to VA enterprise data in the proposed VA EA data layer. These use cases all take place in the context of a larger scenario in

which a Veteran moves from Washington, DC to San Francisco, CA, which leads to not only a change of address, but a change from one VA Veterans Integrated Service Network (VISN) to another. The first use case scenario describes the initial change of address for the Veteran, which creates the conditions for the two subsequent use cases.

#### **4.1 Use Case 1: Veteran Initiates a Change of Address**

##### **4.1.1 Purpose**

In the VA “as-is” environment, a Veteran’s single interaction to change an address will not propagate across all the databases belonging to all the VA organizational units that provide services and benefits to the Veteran. Instead, the Veteran must notify multiple VA organizational units of his or her change of address. This is burdensome for the Veteran and creates significant opportunities for errors.

This use case provides an example of how the proposed capabilities of the VA EA data layer will allow a Veteran (or VA staff, acting on the Veteran’s behalf) to update his or her address once, and have the update automatically propagated across VA, without any further action on the Veteran’s part.

##### **4.1.2 Assumptions**

Assumptions for this use case include:

- Regardless of the web service or application employed by the user, the same core set of services are used for the address change.
- User authentication and authorization processes have already been completed successfully, allowing the Veteran, or a VA staffer on behalf of the Veteran, to update his/her contact information.
- BIRLS is the ADS for locating the beneficiaries of a particular Veteran.

##### **4.1.3 Use Case Description**

###### **Basic Flow of Events**

1. The Veteran, or a VA staff member who is acting on the Veteran’s behalf, initiates a change of address through a web application to reflect the Veteran’s relocation from Washington, DC to San Francisco, CA.
2. A data message is sent from the application through authoritative information services down to the data layer. This data message contains the following elements:

- a. The Identity and Access Management (IAM) security ID of the user responsible for the address change (Veteran or VA Staffer)
  - b. Veteran's MVI ICN
  - c. Veteran's previous address
  - d. When the new address record was created, and whether it was through an internal or external user application
3. The message is processed by the eCRUD service, which writes the address change to the data lake and also to the ADS for Veteran addresses.
  - a. Because the address update was performed by a SOA application and is already in a format consistent with a common enterprise schema, the data lake does not have to harmonize the data.
  - b. The records in both data stores will now reflect the Veteran's new address.
4. The change of address in the data lake is detected by CDC capabilities, triggering multiple preprogrammed actions (described below).
5. The change of address is replicated to target data stores that subscribe to updates of master records in the ADS for Veteran addresses.
  - a. The data lake waits for a specified amount of time to receive acknowledgement of the update from the subscribing data stores.
  - b. All but one of the subscribing data stores sends an acknowledgement to the data lake indicating that the address update was received and written successfully.
  - c. The data lake sends a second update message to the target data store that failed to acknowledge the first message, and waits for a reply.
  - d. The data store at issue does not acknowledge the second update message, so the data lake sends an alert to the owners of the subscribing data store and the data lake stewards informing them of the issue.
6. The data lake sends instructions to automated outreach systems to send follow-up physical and/or electronic email to the Veteran that:
  - a. Provides the Veteran information to the VA Health and Benefits Centers closest to the Veteran's new address (along with links to their websites).
  - b. Informs the Veteran that his/her dependents (listed individually by name) will have their contact information changed to match the Veteran's contact information.
  - c. Supplies the Veteran with contact information for the appropriate VA offices for resolution of any issues that may be associated with the move to a new address.
7. The data lake ensures that the Veteran's change of address is extended to the Veteran's dependents by taking the following actions:
  - a. Queries BIRLS for the MVI ICNs of dependents associated with the Veteran (e.g., spouses, children)

- b. Upon receiving requested ICNs from BIRLS, queries its own database for records that:
  - i. Have the ICNs returned by BIRLS
  - ii. Have a physical address and home phone number matching the Veteran's previous address and phone number
- c. Updates the contact information of the Veteran's dependents in its own data stores and sends the update to the appropriate ADS and ADS subscribers; again, provenance and lineage metadata is attached to indicate the nature, time, and cause of the change(s)
  - i. As in Step 5, the data lake waits for acknowledgement of the update from the subscribers; this time it also expects an acknowledgement from the ADS. It receives acknowledgements from all subscribers, except for the one that was experiencing problems before.
- d. Uses authoritative information services to notify VistA<sup>11</sup> that the Veteran's address has changed and that the Veteran and dependents should be aligned to a new VAMC on the basis of his/her new ZIP code

## Results

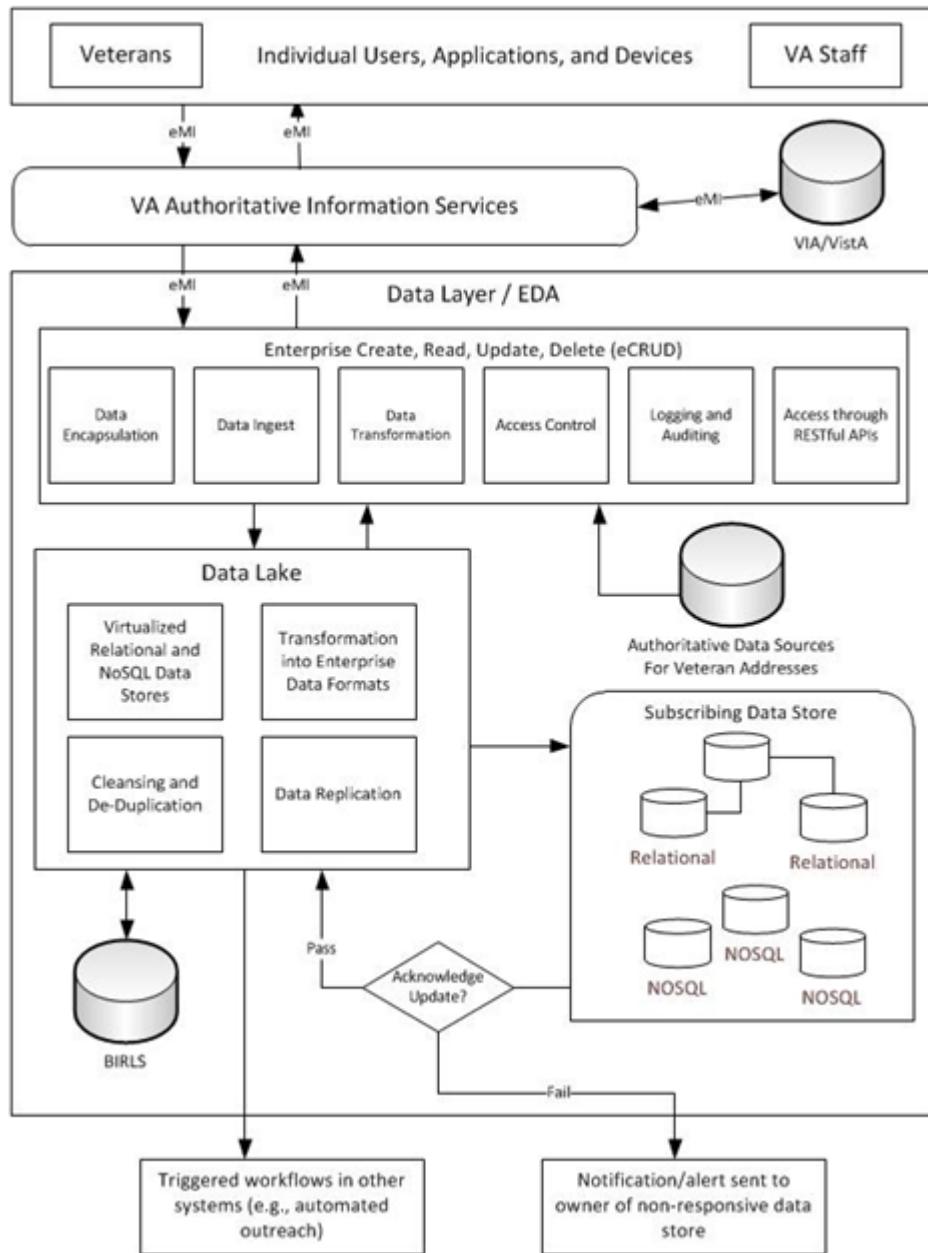
VA benefits and records travel seamlessly with the Veteran and his/her dependents. The Veteran receives mail and phone calls from VA at his/her new address and phone number in San Francisco, and no communications are mistakenly sent to the Veteran's previous address. The Veteran's complete records are easily accessible to VA clinicians and benefits' counselors in the VISN, where the Veteran records now reside.

### 4.1.4 Use Case Context Diagram

Figure 4 illustrates the flow of information through the components of the VA EA architecture and the data layer, as described above.

---

<sup>11</sup> Refer to the *Enterprise Messaging Infrastructure (eMI) EDP* for details on the interaction between services in the service layer. See the *VistA Evolution SOA EDP* for specifics on data access within VistA systems.



**FIGURE 4: EDA USE CASE #1 - VETERAN INITIATES A CHANGE OF ADDRESS**

*Figure 4 Description: The Veteran, or a VA Staffer acting on the Veteran's behalf, initiates a change of address through a web application to reflect the Veteran's move from the Washington, DC area to the San Francisco, CA area. Follows the same data flow as described in Figure 2.*



## **4.2 Use Case 2: Clinician Accesses Veteran’s Health Records from Another Region**

### **4.2.1 Purpose**

In the “to-be” state of the VA IT environment, instances of VistA and their associated data will be encapsulated using VIA, as described in Section 2.1.<sup>12</sup> VIA and the VA EA data layer will be accessible (to applications, services, and each other) through shared authoritative information services. This use case describes a clinician using an application that, through those services, accesses data from both VistA and the data layer, providing contextual information that helps ensure continuity of care when a Veteran moves between different VAMCs.

### **4.2.2 Assumptions**

Assumptions for this use case include:

- The Veteran has moved his or her residence from one state to another state – this can be a permanent move or a seasonal “snowbird” move.
- The clinician who accesses the Veteran’s medical records is doing so because he/she will be meeting the Veteran for the first time during a routine checkup or consultation appointment.
- The clinician who accesses the Veteran’s medical records has successfully completed the user authentication and authorization process for the web interface or application that provides access to the Veteran’s records.

### **4.2.3 Use Case Description**

#### **Basic Flow of Events**

1. Clinician opens an application that provides access to his/her appointment schedule. One of the listed appointments is flagged with an indicator showing that the patient has not been to Clinician’s VAMC before, based on the Veteran’s change of address (in the first use case) and resulting alignment from a VAMC in the Washington, DC area to the current VA Medical Center in the San Francisco, CA area.
2. Clinician selects the appointment in order to access a “Person View” of information pertaining to the new patient, such as name, age, branch, and dates of service, etc. The Person View contains links or buttons that clinician can use to access medical information about the Veteran.

---

<sup>12</sup> For additional details, refer to the *VistA Evolution SOA EDP*.

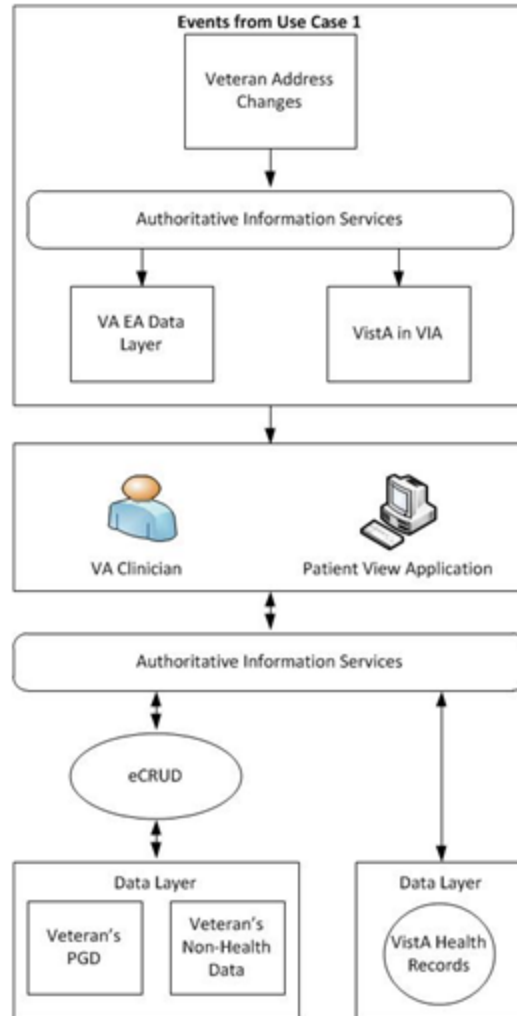
3. Clinician selects option to view prescription information on the patient. Application calls on authoritative information services to access VistA for prescription records associated with the Veteran's VHA PatientID. Clinician reviews Veteran's current prescriptions.
4. Clinician returns to the Person View and selects an option to view Veteran's notes entered into MyHealtheVet. Application calls on services with access to EDA, which retrieves Veteran's PGD stored in the best available instance of the data lake for clinician to review.
5. By accessing both VistA and non-VistA data through the application, clinician is able to get an overall picture of Veteran's current condition, medical history, and health-related concerns.

## **Results**

By accessing both VistA and non-VistA data through the application, clinician is able to get an overall picture of Veteran's current condition, medical history, and health-related concerns, ensuring that the clinician can provide continuity of care.

### ***4.2.4 Use Case Context Diagram***

Figure 5 illustrates how the events of Use Case #1 affect what the Clinician sees in his/her application, and how that application accesses data from both the VA EA data layer and VistA in the VIA platform.



**FIGURE 5: EDA USE CASE #2 - CLINICIAN ACCESSES VETERAN'S HEALTH RECORDS FROM ANOTHER REGION**

*Figure 5 Description: Depicts process for clinician in accessing Veteran's health records from another region. A change in a Veteran's address occurs, as described in Use Case #1. VA staffer initiates "POC Check." VA staffer reviews recipients for a Washington, DC area mailing list. Staffer notices Veteran's name is flagged. VA Staffer reviews "flagged" entry. VA staffer checks investigates flagged Veteran's name. VA staffer sees that Veteran's address is no longer in the Washington, DC area. VA staffer removes "out of range" Veteran from the mailing list. Veteran will no longer receive outreach intended for Washington, DC area residents*

### 4.3 Use Case 3: VA Staffer Checks Veteran's Status Prior to Outreach

#### 4.3.1 Purpose

Uncoordinated or inappropriate outreach is a persistent problem created by VA data silos. Some examples of outreach issues caused by incorrect or outdated information regarding a particular Veteran include:

- Inability to relay critical information in a timely fashion
- Sending sensitive documents to the wrong address
- Attempting to make contact with a Veteran after he or she has died

The proposed capabilities for the VA EA data layer include a mechanism to propagate changes to the master records in ADSs across all other connected VA databases, ensuring that all instances of the record within the data stores are consistent and correct. This use case describes how automatic change propagation and related triggers will allow VA staff to provide better coordinated outreach and services to Veterans.

#### **4.3.2 Assumptions**

Assumptions for this use case include:

- The Veteran has confirmed a change of address (as described in the first use case), updating his/her record across VA and setting off pre-programmed triggers and alerts.
- The VA staffer reviewing the Veteran's contact information has successfully completed the user authentication and authorization process for the VRM application he/she is using.

#### **4.3.3 Use Case Description**

##### **Basic Flow of Events**

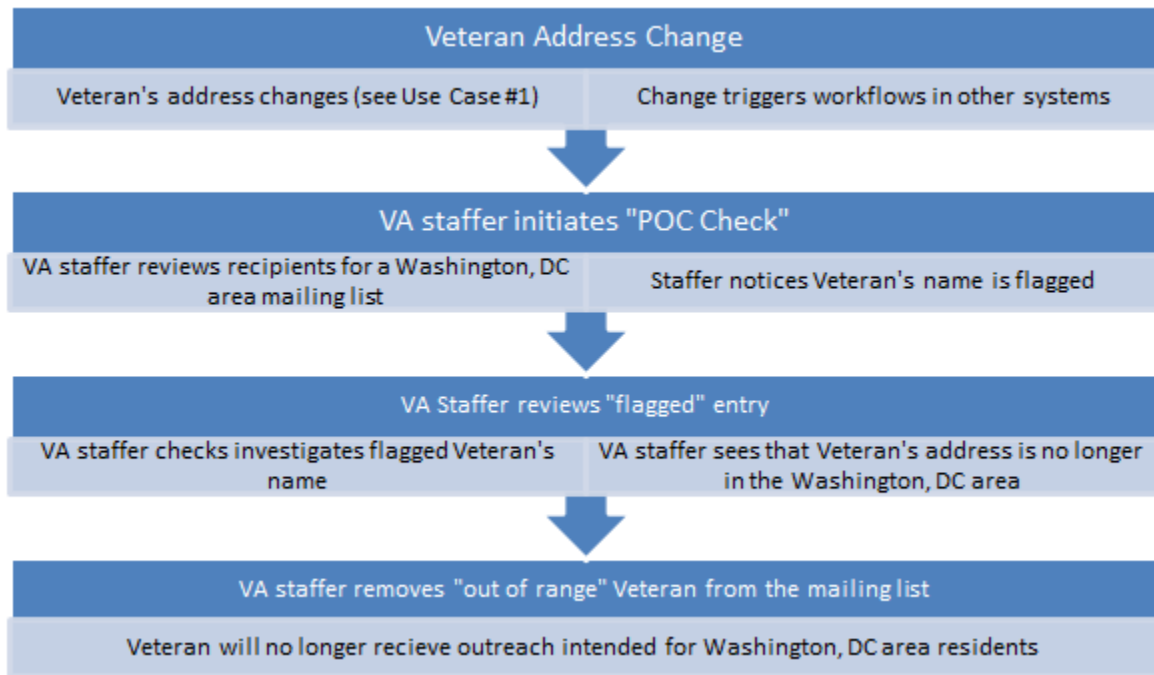
1. VA staffer selects a mailing list associated with mass e-mails for a VA program in the Washington, DC metropolitan area.
2. VA staffer initiates a "POC Check" function that will query various tables in the data lake (including those with the official records for Veteran/Beneficiary contact information) to find entries on the mailing list that meet certain preprogrammed conditions.
3. Based on the results of the query, the VRM application displays a subset of individuals on the mailing list who have been flagged because they meet the preprogrammed conditions specified for the POC Check function.
4. VA staffer reviews details on why each individual is flagged. He/she sees that some individuals on the mailing list are flagged because their physical addresses on file are not in the Washington, DC metropolitan area.
5. VA staffer unsubscribes "out of range" individuals from the e-mail list.

##### **Results**

Without having to unsubscribe or ask to be removed from a mailing list for Veterans in the Washington, DC metropolitan area, the Veteran who changed his/her address in the first use case will no longer receive mail from this list, which is no longer relevant to the Veteran.

#### 4.3.4 Use Case Context Diagram

Figure 6 illustrates the sequence of events as the VA staffer checks and acts on the status of flagged mailing list recipients in an outreach application.



**FIGURE 6: EDA USE CASE #3 - VA STAFFER CHECKS VETERAN'S STATUS PRIOR TO OUTREACH**

*Figure 6 Description: This diagram illustrates the flow of events previously described for Hybrid Data Access Use Case #3, "VA Staffer Checks Veteran's Status Prior to Outreach." Step 1: A change in a Veteran's address occurs, as described in Use Case #1. Step 2: VA staffer initiates "POC Check". Step 2a: VA staffer reviews recipients for a Washington, DC area mailing list. Step 2b: Staffer notices Veteran's name is flagged Step 3: VA Staffer reviews "flagged" entry. Step 3a: VA staffer checks investigates flagged Veteran's name. Step 3b: VA staffer sees that Veteran's address is no longer in the Washington, DC area. Step 4: VA staffer removes "out of range" Veteran from the mailing list. Step 4a: Veteran will no longer receive outreach intended for Washington, DC area residents.*



## APPENDIX A. SCOPE

This appendix contains the following high-level information for the EDA EDP:

- Document scope
- Intended audience
- How to use this document
- Document development and maintenance

### Scope

This increment of the EDA EDP document will expand on the concepts and constraints described in the *Data-as-a-Service (DaaS) EDP* by describing how they will be applied in a VA enterprise data solution. Specifically, this document addresses:

- Interactions between middleware and different types of databases (e.g., Relational, NoSQL).
- Aggregation and processing of information retrieved from multiple data sources of different types into a single message consumable by services/applications connected to EDA.
- Ensuring data can be accessed and/or edited quickly (availability) while maintaining its correctness and integrity (consistency) across data sources.
- Providing access to VistA records through the VIA, the future replacement for the current access service (Medical Domain Web Services or MDWS).<sup>13</sup>

The following concepts are outside the scope of this document:

- Data messaging security and authenticity
- Access control and authorization decisions
- Specifics of applications/services used to access EDA
- Infrastructure and hardware design specifications
- Specifics of services and applications that will support EDA
- Vendor-specific technologies

The use cases associated with this EDP can be found in the companion document, *Enterprise Data Access Increment 2 Use Cases*.

---

<sup>13</sup> For a detailed description of VIA, refer to the VistA Evolution SOA EDP.

## Purpose

The purpose of this document is to provide strategic direction and enterprise-level capability guidance that identifies best practices for developing an EDA capability that provides seamless, unified, read and write-back access to VA's data stores within the VA EA data layer.

EDA focuses on sharing and providing access to high-value customer information about individual Veterans and beneficiaries. This type of information is used across VA Administrations and LOBs, but not in a consistent, enterprise-wide format. Data elements of high-value customer information include the following:

- First name
- Last name
- Date of birth
- Military service history
- Address<sup>14</sup>
- Contact information

Most of the data at issue is relatively static in nature. This means that once a record is created, that record is likely to remain unchanged for weeks, months, or years at a time. It is critical, and currently problematic for VA, that when these changes *do* occur, the changes are reflected in all the records used by all Administrations and LOBs in the enterprise.

Integrating and standardizing this high-value customer data is the key to addressing the systemic data access and management problems identified in the previous section. Maintaining a single data access point across VA enables the Department to provide greater accessibility and transparency for Servicemembers, Veterans, beneficiaries, and partner organizations. Integrating VA's data stores will also reduce redundancy in both organizational data and processes.

## Intended Audience

All programs will use this document to guide efforts toward using enterprise shared services (ESS) for standardized, enterprise-wide access to enterprise data. This will help programs meet data sharing requirements utilizing enterprise data stores while:

- Developing new VA applications

---

<sup>14</sup> A single individual may have multiple types of addresses, for example, a permanent home address, a temporary address, and/or a prescription address. This document uses the term "address" in the sense of one's permanent residence/primary address.

- Modifying existing production systems
- Acquiring and integrating Commercial Off the Shelf (COTS) (including open-source) applications

### **How to Use This Document**

This document can be used to guide and inform the following related activities:

- Development, acquisition, and/or implementation of products and services with the capabilities required for VA enterprise data integration and analytics
- Creation and use of data management and governance policies, standards, and mechanisms
- Definition, articulation, and enactment of requirements and business processes around enterprise data use and management
- Long-term development of sophisticated enterprise analytics capabilities

### **Document Development and Maintenance**

This document was developed collaboratively with internal stakeholders from across VA, with participation from OI&T, Enterprise Program Management Office (EPMO), Office of Information Security (OIS), Architecture, Strategy, and Design (ASD), and Information Technology Operations and Service (ITOPS). In addition, the development effort included engagements with industry experts to review, provide input, and comment on the proposed pattern. This document contains a revision history and revision approval logs to track all changes. Updates will be coordinated with the Government lead for this document, who will facilitate stakeholder coordination and subsequent re-approval depending on the significance of the change.



## **APPENDIX B. KEY CONCEPTS**

This appendix provides high-level overviews of key concepts associated with EDA (and the data layer) that apply to solving recurring problems within the current state of the VA IT environment, such as:

- Lack of a unified, consistent view into records for a Veteran or beneficiary
- Necessity for Veterans and beneficiaries to provide the same information (e.g., address) multiple times to different parts of VA
- Manual, piecemeal write/update processes for siloed data stores, resulting in long delays and significant potential for error
- Duplicative or inappropriate outreach based on incorrect or outdated information

The concepts outlined below provide the context for the data layer “to-be” architectural attributes that are described in Section 3. These attributes will establish the design constraints to be applied to solution architectures that are developed at VA.

Related database concepts (with definitions) that are not included in this section can be found in Appendix C: Definitions.

### **Availability and Consistency**

Availability is concerned with the accessibility of data to authorized consumers (e.g., applications, services). This includes performance considerations, including the amount of time required for a query/response to complete successfully.

Consistency is concerned with the quality, accuracy, and “up-to-date-ness” of data. Achieving consistency throughout an enterprise with numerous distributed data stores is particularly challenging. This requires synchronizing records in data stores across the enterprise, with authoritative sources as frequently and rapidly as possible.

Data consistency and data availability can be mutually exclusive under some circumstances, because ensuring consistency reduces the speed at which data can be delivered. An effective enterprise data solution is balanced between these two properties, which is consistent with the enterprise’s mission.

### **Data Management and Governance**

The following are key concepts within data management and governance: data stewards, data models and schema, and provenance and lineage.

## **Data Stewards**

ADSs are sources of data or information that are officially recognized as trusted, timely, and secure, and used within VA's information environment to support VA business processes. The DGC designates these sources and OI&T develops and maintains technology solutions (e.g. services) that use these sources.

A data steward is a designated role responsible for maintaining the quality of data in an ADS. The data steward is typically an organizational unit that is synonymous with or designated by the business owner of the ADS. Data stewards develop and implement rules and processes around how data in ADSs should be entered, changed, audited, or otherwise handled in order to ensure its quality.

## **Data Models and Schema**

The enterprise data architecture (which also applies to metadata) documents VA data requirements and establishes a source for their data entity definitions, attribute definitions, and characteristics. A high-level conceptual data model describes general groupings of data in terms of entities<sup>15</sup> and the relationships between those entities. It informs a logical data model which describes data elements, their attributes, and their associated rules. The logical data model forms the basis for a common schema or blueprint for the structure of data elements used by enterprise services.

Traditional relational databases use a “schema-on-write” approach to input. Any input to the database must be in a predetermined format recognized by the database before it can actually be written. If the structure of the input is not what the database “expects” – i.e., organized using the database schema – the write operation will fail and errors will result. NoSQL databases can employ a “schema-on-read” approach, accepting any input regardless of its structure (or lack thereof), transform any elements of the input they recognize, and ignore everything else.

## **Provenance and Lineage**

An important aspect of data management and governance is using metadata to track the origin and history – or provenance and lineage – of records and changes to those records. All records should be bundled with metadata indicating their provenance and lineage.

Tracing the provenance of records and updates is essential to preserving data quality in the context of an enterprise data service. Provenance supports data integrity by ensuring only

---

<sup>15</sup> The Unified Modeling Language (UML) is the prevailing language for developing health information models.

changes made by authorized users, processes, applications, and/or systems are propagated through the enterprise data service and committed to ADSs and local data stores. It should be detailed enough to trace the complete “chain of custody” of a record back to its point of origin.

Data lineage is the history of a data record. It tracks deltas (i.e., changes) between the current state of the record and the previous version or versions, with timestamps to indicate when changes were made. Lineage is essential for rollback functionality and for queuing changes to data elements in a record to ensure the data are committed in the correct order.

## **Technical Components of an Enterprise Data Solution**

In addition to the data stores, the proposed EDA capability will include three key components: a CDC, a data lake, and a data ingester.

### **Change Data Capture Tool**

Tools or applications with CDC functionality can detect changes to a data store or record and initiate preprogrammed actions as a result of that change. In the proposed EDA capability, CDC is the preferred mechanism for maintaining consistency between publishers (in this case, ADSs) and subscribers throughout VA. A CDC-capable tool will “listen” for ADS updates (i.e., new records or changes to existing records) and push those updates to one or more selected targets. The CDC tool can be used to initiate other actions that are based on changes to data, including sending alerts or automatically triggering workflows.

### **Data Lakes**

Data lakes<sup>16</sup> are storage repositories that hold vast amounts of data in native formats until needed. A data lake typically contains multiple structured and unstructured data stores. These databases are encapsulated to present a transparent logical interface to the services and applications that interact with them. Instances of data lakes are hosted in clusters of co-located servers. Duplicate instances can be distributed geographically to facilitate local data access and ensure continuity of operations.

Data in the data lake can be cleansed (i.e., standardized, de-duplicated, and corrected) and indexed to support an enterprise-wide search capability. Some or all of the data can be harmonized by transforming it from its original format into a format that is consistent with one or more common enterprise schema. Harmonized data can be served to consuming enterprise services and applications, replicated to a data warehouse for analytics, or both.

---

<sup>16</sup> Equivalent terms include data hubs, data reservoirs, landing zones, or operational data warehouses.

Data lakes can serve as the target store for virtualized databases from decommissioned systems, allowing those systems to be taken offline, without disrupting the applications and services that depend on them. This capability facilitates long-term transitions from a traditional IT environment to a virtualized and/or cloud-based environment.

The data lake should *not* contain a data warehouse or its own security audit logs. It is an industry best practice for enterprise data solutions to use separate, dedicated data stores for workloads associated with operations, analytics, and logging, since they have very different (and sometimes conflicting) security, performance, and resource requirements.

### **Data Ingestion and Data Ingesters**

Data ingestion is a function (or series of related functions) that serves as the intake mechanism for a data lake. The data ingestion function handles large volumes of data messages from various sources across an enterprise. The data ingestion function queues data messages to ensure they are committed in the correct order, and performs preprocessing before writing the data into the data lake's storage.

In some enterprise data solutions, data ingestion is a capability of the data lake itself. Other data solutions decouple data ingestion from the data lake into a “data ingester” component, with its own dedicated cluster of hardware. EDA employs a data ingester, partly because it can support write-back from SOA applications to ADSs (not just the data lake). A data ingester can also be scaled to accommodate growing data message volume, without impacting the performance of the data lake. Data ingesters can be extended to support advanced capabilities, including processing streaming data from wearables and medical devices.

To some degree, a data ingester can also perform data transformation. It can transform some data messages in route to the data lake, while letting others pass through to be transformed after they are loaded into storage. The proportion of total transformation work performed by the ingester, versus the data lake itself, depends upon the characteristics of the IT environment and the requirements or business processes of the enterprise.

## APPENDIX C. DEFINITIONS

This appendix provides definitions for terms used in this document, particularly those related to databases, database management, and data integration.

Key Term	Definition
Atomicity, Consistency, Isolation, Durability (ACID)	This set of properties guarantees that database transactions are processed reliably. Some DBMSs provide better guarantees of ACID properties than others.
Authoritative Data Source (ADS)	This source of data or information is designated and recognized as officially trusted, timely, secure, and used within VA's information environment in support of VA business processes. Administrations and staff offices designate these sources within domains for which they are the stewards (i.e., owners/responsible parties). OI&T develops and maintains technology solutions (e.g. services) that use these sources.
Change Data Capture (CDC)	The capability is used to initiate certain preprogrammed actions that are based on changes to data stores and/or data records; e.g., "When a new record is created in Database A, create an associated record in Database B."
Common Customer Data (CCD)	CCD is normalized across the enterprise and made available to VA application software as core data elements, such as identity, military service record, and contact information, which are necessary for interaction with Veterans.
Conceptual Data Model	This model is presented at a high level of abstraction, hiding the underlying details, and making it easier for users to comprehend.
Consistency	In the context of data and data management, consistency is concerned with the quality, accuracy, and timeliness of data. Achieving consistency throughout an enterprise with many distributed data stores can be particularly challenging. It requires synchronizing records with authoritative sources as frequently and as rapidly as possible in data stores across the enterprise.

Key Term	Definition
Data Federation	This concept relates to aggregating data from disparate sources in a virtual database so it can be used for business intelligence or other analysis. The virtual database created by data federation software does not contain the data itself. Instead, it contains information about the actual data and its location.
Data Ingestor	An ingester is a component in an enterprise data solution that serves as the intake pipeline for a data lake. It takes in large volumes of data messages from various sources across an enterprise, queues them to ensure that they are committed in the correct order, and performs some level of preprocessing before writing them into the data lake's storage. The data ingester may perform data transformation on some incoming data messages, while others are transformed in the data lake itself. The proportion of total transformation work performed by the ingester, versus the data lake itself, depends upon the characteristics of the IT environment and the requirements or business processes of the enterprise.
Data Lake	A lake is a type of data store that is a critical component of enterprise data integration and/or big data solutions. It serves as the aggregation, integration, and distribution center for data from sources across the enterprise. A data lake makes it available to consume services and applications. It is also called a data sink, data reservoir, or landing zone.
Data Model	See <i>Conceptual data model</i> and <i>Logical data model</i> .
Data Store	Often used interchangeably with "database," a data store also encompasses types of data repositories that are not strictly databases, such as data warehouses and data lakes.
Data Warehouse	A warehouse is an archival data store where enterprise data is replicated and processed for analytics and the creation of test data. Unlike a database, a data warehouse does not hold "live" production data.

Key Term	Definition
Database Federation	Federation is an established method of encapsulating independent databases into a single logical entity with a common interface. Queries coming through the interface are “broken” into subqueries, which are sent to the appropriate databases in the federation. Responses to the queries are reassembled into a single message for whatever application/service initiated the original query.
Database Virtualization	Virtualization is the practice of hosting databases in a virtual environment. Like database federation, this can be used to encapsulate multiple databases into a single logical entity. Virtualization provides other advantages, such as faster query/response performance, platform independence, centralized management/governance, and analytics capabilities.
Enterprise Create, Read, Update, Delete (eCRUD)	The eCRUD service was initially created as part of the DAS project. eCRUD provides an interface that allows enterprise services to perform create, read, update or delete (CRUD) operations on data in the VA SOA data access layer/EDA solution. It also supports numerous adapters for data transformation, notification of data changes, and custom event handlers.
Enterprise Messaging Infrastructure (eMI)	The eMI is a communications infrastructure that supports transactions in the VA SOA environment.
Integrated Control Number (ICN)	The ICN is the primary key and unique identifier associated with individuals in MVI.
Lineage	Lineage is information (usually metadata) that provides a record of changes made to a data element, including a description of the changes and when the changes were committed. In an enterprise data solution, lineage information supports rollback functionality, queuing, and the correct ordering of data updates.

Key Term	Definition
Logical Data Model	A logical data model is an entity-relationship data model that includes data attributes that represent the inherent properties of the data, such as names, definitions, structure, and integrity rules. The data attributes are independent of software, hardware, volumetrics, frequency of use, or performance considerations.
Master Veterans Index (MVI)	The MVI is the VA ADS for the identity information of approximately 17 million Veterans and beneficiaries. Identity traits are cross-referenced with DoD records.
Not Only SQL (NoSQL)	NoSQL is a type of DBMS that structures data in a non-tabular/non-relational format. NoSQL database types include key-value, column-family, document, and network. Some NoSQL databases can also store unstructured data.
Object-Oriented Database	An object-oriented database is a type of DBMS that organizes data into objects (in the sense of object-oriented programming). These databases are designed to integrate with object-oriented programming languages. For example, the advanced DBMS, InterSystems Caché (used for VHA's VistA systems), integrates with the MUMPS programming language to store objects written in MUMPS so that objects can be reused or reassembled. Caché is a high-performance object database that structures data in hierarchical multidimensional arrays. It does not use schema.
Provenance	Provenance is data (usually metadata) that indicates the origin of a data element/record or changes to that element/record. Provenance supports the capability to establish, record, and trace a clear "chain of custody" for data.
Relational Database	A relational database is a type of DBMS in which data is organized in a tabular fashion, such as in tables. Each row in a table is an individual instance of a type of entity, and each column is an attribute of the entity. For example, an Oracle database at VA may contain a table in which each row represents an individual Veteran record, and the columns (fields) contain attributes about the Veteran, such as date of birth, gender, branch of service, etc.



Key Term	Definition
Schema	The schema is the logical “blueprint” for how data is structured in a relational database. The schema determines the format of data in the data store and of data messages passed between systems, services, and applications within an enterprise. Most contemporary DBMSs support the use of multiple schema, making it possible to organize and view a single dataset in different ways for different purposes.
Structured Query Language (SQL)	SQL is a common language for accessing and manipulating data that is stored in relational databases.

## APPENDIX D. ACRONYMS

The following table provides a list of acronyms that are applicable to and used within this document.

Acronym	Description
ACID	Atomicity, Consistency, Isolation, Durability
ADR	Administrative Data Repository
ADS	Authoritative Data Source
API	Application Programming Interface
BII	Business Information Architecture
BIRLS	Beneficiary Identification Record Locator System
BOSS	Burial Operations Support System
CCD	Common Customer Data
CDC	Change Data Capture
CDI	Customer Data Integration
CDM	Conceptual Data Model
CRM	Customer Relationship Management
CRMe	Customer Relationship Management Enterprise
CPRS	VHA Computerized Patient Record System
COOP	Continuity of Operations Plan
DAR	Data Architecture Repository
DAS	Data Access Service
DBMS	Database Management System
DGC	Data Governance Council
DISA	Defense Information Systems Agency
DoD	United States Department of Defense
eCRUD	Enterprise Create, Read, Update, Delete
EDA	Enterprise Data Access
EDP	Enterprise Design Pattern
EIM	Enterprise Information Management
ELDM	Enterprise Logical Data Model
eMI	Enterprise Messaging Infrastructure
ESS	Enterprise Shared Services
ETSP	Enterprise Technology Strategic Plan
HC IdM	VHA Data Quality Healthcare Identity Management
IAM	Identity and Access Management

Acronym	Description
ICN	Integration Control Number
Ids ITT	VA Identity Services Integrated Technical Team
IT	Information Technology
LOB	Line of Business
MVI	Master Veterans Index
NCA	National Cemetery Administration
NIST	National Institute of Standards and Technology
NoSQL	Not Only SQL (database)
PGD	Patient-Generated Data
PHI	Protected Health Information
PII	Personally Identifiable Information
REST	Representational State Transfer
SDD	System Design Document
SOA	Service-Oriented Architecture
SQL	Structured Query Language
STIG	Security Technical Implementation Guide
UML	Unified Modeling Language
VA	Department of Veterans Affairs
VADI	VA Data Inventory
VA EA	VA Enterprise Architecture
VAMC	VA Medical Center
VASI	VA Systems Inventory
VBA	Veterans Benefits Administration
VEAR	VA Enterprise Architecture Repository
VHA	Veterans Health Administration
VIA	VistA Integration Adapter
VISN	Veterans Integrated Service Network
VistA	Veterans Information Systems and Technology Architecture

## APPENDIX E. REFERENCES, STANDARDS, AND POLICIES

This EDP is aligned to the following VA OI&T references and standards applicable to all new applications being developed in the VA, and are aligned to the VA Enterprise Technical Architecture (ETA):

#	Issuing Agency	Policy, Directive, or Procedure	Purpose
1	VA OIS	VA 6500 Handbook	The overall security framework for VA is defined, including data storage, retrieval, and exchange.
2	VA ASD	OneVA Enterprise Technology Strategic Plan (ETSP)	The EDA EDP will help programs develop applications in alignment with IT Vision Attributes 6, 8, 9, 10, and 12 (scale-out of enterprise data stores).
3	VA	Data Access Service (DAS) System Design Document (SDD)	A description of the planned construction of the VA DAS is provided. It translates identified business needs, business requirements, functional requirements, and non-functional requirements into a document from which developers can create the system.
4	VA	VA Directive 6518: Enterprise Information Management	Official policy for the implementation of ADSs in VA is provided. To comply with this directive, EDA provides access and synchronization with enterprise ADSs.
5	VA	VA Memorandum: VA Identity Management Policy (VAIQ 7011145)	The Master Veterans Index (MVI) is established as the authoritative source for identity traits of Veterans and VA stakeholders. It mandates a unique identifier for all Veterans and requires that all VA applications integrate with MVI.
6	VA	Data Access Service (DAS) eCRUD Service Interface Control Document	The functions, capabilities, and requirements of the eCRUD service are described.
7	VA	VA Directive 6051: Department of Veterans Affairs (VA) Enterprise Architecture (EA)	Department-wide policy for the establishment and implementation of an integrated One-VA EA is prescribed to be used for the development and management of all information assets. Operating principles are established for the VA EA. Specific responsibilities for its development and implementation are assigned.

#	Issuing Agency	Policy, Directive, or Procedure	Purpose
8	OMB	OMB M-13-13, Open Data Policy – Managing Information as an Asset	Describes the requirement for agencies to collect or create information in a way that supports downstream information processing and dissemination activities.



## APPENDIX F. DATABASE MANAGEMENT SYSTEMS USED AT VA

All projects will leverage the approved tools and standards located in the One-VA TRM<sup>17</sup> to comply with the architectural guidance provided in this document. Table 5 lists the approved tools for this EDP. This document establishes the technical standards that inform the decisions about mobile technologies that are included in the One-VA TRM. The list of DBMSs in the table below was derived from assessments conducted by the ASD team responsible for development and maintenance of the One-VA TRM. Future updates of this EDP will reflect forecasted standards and subsequent changes in approved products in the One-VA TRM.

Domain	Area	Category	Technology
Information Management Technologies	Data Management	Relational DBMS / Desktop DBMS	Microsoft Access
Information Management Technologies	Data Management	Non-Relational Database / In-memory Database	Redis
Information Management Technologies	Data Management	Non-Relational Database	FIS-GTM
Information Management Technologies	Data Management	Non-Relational Database	Integrated Data Store II
Information Management Technologies	Data Management	Non-Relational Database	MarkLogic Server
Information Management Technologies	Data Management	Non-Relational Database	MongoDB
Information Management Technologies	Data Management	Object-Oriented DBMS	Cache
Information Management Technologies	Data Management	Object-Oriented DBMS / Relational DBMS	PostgreSQL
Information Management Technologies	Data Management	Relational DBMS	DB2
Information Management Technologies	Data Management	Relational DBMS / Desktop DBMS	Filemaker Pro
Information Management Technologies	Data Management	Relational DBMS	InterBase
Information Management Technologies	Data Management	Embedded DBMS / Relational DBMS	Java DB
Information Management Technologies	Data Management	Non-relational Database	Model 204 (M204)

<sup>17</sup> <http://www.va.gov/trm/ToolListSummaryPage.aspx>

Domain	Area	Category	Technology
Information Management Technologies	Data Management	Relational DBMS	Microsoft SQL Server
Information Management Technologies	Data Management	Relational DBMS	MySQL
Information Management Technologies	Data Management	Object-Oriented DBMS / Relational DBMS	Oracle DB
Information Management Technologies	Data Management	Relational DBMS	SAP Sybase SQL Anywhere
Information Management Technologies	Data Management		CA IDMS
Information Management Technologies	Data Management	Relational DBMS / Embedded DBMS	SQLite
Information Management Technologies	Data Management	Columnar DBMS / Non-relational Database	Apache Cassandra

**Disclaimer:**

This document serves both internal and external customers. Links displayed throughout this document may not be viewable to all users outside the VA domain. This document may also include links to websites outside VA control and jurisdiction. VA is not responsible for the privacy practices or the content of non-VA websites. We encourage you to review the privacy policy or terms and conditions of those sites to fully understand what information is collected and how it is used.