Perceptive Reach

Integrated Reach Database System

(IRDS)

System Design Document



Department of Veterans Affairs

December 2014

Version 1.0

Revision History

| Date | Version | Description | Author |
| --- | --- | --- | --- |
| 12/11/2014 | 1.0 | Final Review | Monica Mohler/Paul Bradley |
| 12/10/2014 | .04 | Peer Review | Radina Ivanova |
| 11/26/2014 | .03 | Added additional content on application context and on continuous integration. | Robert Snelling, Andal FeQuiere, Paul Bradley |
| 11/13/2014 | .02 | Added additional conceptual design details for the database, analytic platform, risk model, dashboard, business benefits, and user stories. | Bill Balshem, Kristopher Hoover, Robert Snelling, Paul Bradley |
| 10/24/2014 | .01 | Added Database, Analytic, Dashboard, Architecture, Networking, Business Benefits. | Bill Balshem, Kristopher Hoover, Robert Snelling, Paul Bradley |
| 10/07/2014 | .00 | First Draft | Paul Bradley |

Artifact Rationale

The System Design Document (SDD) is a dual-use document that provides the conceptual design as well as the as-built design. This document will be updated as the IRDS solution is developed. This document is based on the Project Management Accountability System (PMAS) ProPath Template.

Table of Contents

[1. About this document 4](#_Toc406182591)

[2. Introduction 4](#_Toc406182592)

[2.1. Purpose of the SDD 4](#_Toc406182593)

[2.2. Identification 4](#_Toc406182594)

[2.3. Scope 4](#_Toc406182595)

[2.4. User Characteristics 5](#_Toc406182596)

[2.5. Relationship to Other Documents and Plans 5](#_Toc406182597)

[2.6. Acronyms and Abbreviations 6](#_Toc406182598)

[3. Background 8](#_Toc406182599)

[3.1. Overview of the System 8](#_Toc406182600)

[3.1.1. SDR Database Background 10](#_Toc406182601)

[3.2. Overview of the Business Process 10](#_Toc406182602)

[3.3. Business Benefits 14](#_Toc406182603)

[3.4. Assumptions and Constraints 15](#_Toc406182604)

[3.4.1. Design Assumptions 16](#_Toc406182605)

[3.4.2. Design Constraints 16](#_Toc406182606)

[3.4.3. Design Trade-offs 16](#_Toc406182607)

[3.5. Overview of the Significant Requirements 17](#_Toc406182608)

[3.5.1. Overview of Significant Functional Requirements 17](#_Toc406182609)

[4. Conceptual Design 19](#_Toc406182610)

[4.1. Conceptual Application Design 19](#_Toc406182611)

[4.1.1. Application Context 19](#_Toc406182612)

[4.1.2. High-Level Application Design 21](#_Toc406182613)

[4.2. Conceptual Data Design 22](#_Toc406182614)

[4.2.1. Project Conceptual Data Model 22](#_Toc406182615)

[4.2.2. Database Information 24](#_Toc406182616)

[4.2.3. Data Import Design 24](#_Toc406182617)

[4.2.3.1. SQL Server Integration Services (SSIS) 24](#_Toc406182618)

[4.2.3.2. Remote Procedure Calls (RPC) 24](#_Toc406182619)

[4.3. Conceptual Infrastructure Design 25](#_Toc406182620)

[4.3.1. System Criticality and High Availability 25](#_Toc406182621)

[4.3.2. Special Technology 25](#_Toc406182622)

[4.4. System Architecture 26](#_Toc406182623)

[4.5. Hardware Architecture 26](#_Toc406182624)

[4.6. Software Architecture 27](#_Toc406182625)

[4.7. Continuous Integration / Continuous Delivery 27](#_Toc406182626)

[4.8. Network Architecture 27](#_Toc406182627)

[4.9. Service Oriented Architecture / ESS 28](#_Toc406182628)

[4.10. Enterprise Architecture 29](#_Toc406182629)

[5. Detailed Design 29](#_Toc406182630)

[5.1. Hardware Detailed Design 29](#_Toc406182631)

[5.2. Software Detailed Design 29](#_Toc406182632)

[5.2.1. Dashboard 31](#_Toc406182633)

[5.2.1.1. RESTful Data Service 31](#_Toc406182634)

[5.2.2. Reach Database 32](#_Toc406182635)

[5.2.3. Analytics Platform 33](#_Toc406182636)

[5.2.4. Direct Messaging 34](#_Toc406182637)

[5.3. Specific Requirements 34](#_Toc406182638)

[5.4. Continuous Integration / Continuous Delivery 34](#_Toc406182639)

[5.5. Data Design 35](#_Toc406182640)

[5.5.1. Physical Data Model 36](#_Toc406182641)

[5.5.2. Data Dictionary 38](#_Toc406182642)

[5.5.3. Veteran De-Duping Process 38](#_Toc406182643)

[5.5.4. Non-DBMS Files 39](#_Toc406182644)

[5.5.5. Data View 39](#_Toc406182645)

[5.6. Service Oriented Architecture / ESS Detailed Design 39](#_Toc406182646)

[5.7. Data Analytics Platform 40](#_Toc406182647)

[5.8. Risk Model 41](#_Toc406182648)

[5.9. External System Interface Design 44](#_Toc406182649)

[6. Human-Machine Interface 44](#_Toc406182650)

[6.1. User Roles 44](#_Toc406182651)

[6.2. Interface Design Rules 45](#_Toc406182652)

[6.3. Inputs 46](#_Toc406182653)

[6.4. Outputs 46](#_Toc406182654)

[6.5. Navigation Hierarchy 46](#_Toc406182655)

[7. Security and Privacy 46](#_Toc406182656)

[7.1. Security 46](#_Toc406182657)

[7.2. Privacy 46](#_Toc406182658)

# About this document

This document is a “work in progress”. Updates are provided to this document as the design evolves with each sprint. The early versions of the document will focus on the conceptual design, key business requirements, and the primary user stories.

# Introduction

VA is seeking to expand suicide prevention to include upstream approaches, designed to reduce initiation or escalation of a risk factor. Upstream suicide interventions target individuals or groups who exhibit biological, psychological, or social risk factors that are more prominent among high-risk groups than among the larger population. Understanding the unique needs of our nation’s Veterans and the military culture as it relates to stigma and mental health is important for early intervention. The goal of the Integrated Reach Database System (IRDS) innovation is to promote the general health of the Veteran population and effectively intervene in issues before they escalate in crisis.

The IRDS solution innovates the current process of risk data collection, analysis, and use in effective intervention strategy. The solution will harness the power of large and diverse data stores to aggregate, analyze and identify risk onset as well as reveal previously unidentified at-risk individuals and populations as a holistic and integrated approach.

The IRDS innovation will serve to bolster the three major components of VHA’s Strategic Plan for Suicide Prevention: surveillance, risk and protective factors, and prevention interventions. The IRDS innovation will target antecedent events specific to Veteran populations prior to the onset of risk to mitigate the development of risk.

## Purpose of the SDD

The purpose of this document is to describe in sufficient detail how the proposed system will be constructed. The SDD translates requirement specifications into a document from which the developers can create the actual system. It identifies the top-level system architecture, and identifies hardware, software, communication, and interface components. For the purpose of this document, “system” will be used to reference the overall IRDS solution (Dashboard, Database, Direct Message, and Data Analytics) and “application” will be used to reference the portions of IRDS that will interface with an end user (Dashboard and Direct Message).

## Identification

## Scope

The IRDS SDD describes the architecture, functional components, and interfaces of the IRDS including the:

* Reach Database – a SQL database storing data used for analytic input
* Data Analytics Platform – an integrated collection of tools
* Risk Model – the predictive model(s) used to identify high-risk Veterans
* Dashboard – a multi-view information portal displaying results from the analytic platform and risk model
* Direct Messaging – a secure messaging solution to notify outreach and intervention resources/clinicians about high-risk Veterans

This document should be read in conjunction with the IRDS Interface Design Specification.

## User Characteristics

There are five user interaction scenarios envisioned, upstream at-risk notification, surveillance, research, reporting, and system sustainment, including:

1. Upstream At-Risk Notification – The primary users in this usage model are the VA outreach and intervention teams. The IRDS shall provide secure notification via a Direct Message of at-risk populations and at-risk individuals to these teams. The application will consolidate various data sources, risk factors, and statistical models to identify at risk individuals and populations.
2. Surveillance – The primary users in this model shall include VA leadership, VA Center of Excellence for Suicide Prevention staff, VA Mental Health leaders, and VA Suicide Prevention Coordinators. The surveillance dashboard will be available through a standard web browser that will be updated in near real-time (minimum weekly) with results produced from the continuous monitoring and processing of linked data sources.
3. Research – The users in this usage model are researchers and statisticians looking to leverage the tools and data available through Reach data analytics platform. The solution will provide a framework for these users to utilize the interfaces provided by the assembled tools to perform required research functions.
4. Reporting – This model shall include both direct and indirect users. The direct users are the individuals required to assemble reports. The indirect users are the consumers or target audience of the reports. The direct users will utilize the interfaces provided by the assembled tools to assemble reports. The report generation process shall be automated.
5. Sustainment - The Contractor shall provide the capability for users to edit and add to the IRDS Risk Stratification Model, permit creation to new models and mapping to interfaces.

## Relationship to Other Documents and Plans

The following IRDS documents may be referenced in tandem with the information recorded here:

* [Project Management Plan (PMP)](https://internal.vacloud.us/wiki/pages/81X0Z7X5/Perceptive_Reach_Deliverables.html)

<https://internal.vacloud.us/wiki/pages/81X0Z7X5/Perceptive_Reach_Deliverables.html>

* [Interface Design Specification](https://internal.vacloud.us/wiki/pages/81X0Z7X5/Perceptive_Reach_Deliverables.html)

<https://internal.vacloud.us/wiki/pages/81X0Z7X5/Perceptive_Reach_Deliverables.html>

* [Requirements Specification Document (RSD)](https://internal.vacloud.us/wiki/pages/81X0Z7X5/Perceptive_Reach_Deliverables.html)

<https://internal.vacloud.us/wiki/pages/81X0Z7X5/Perceptive_Reach_Deliverables.html>

* [User Research Report](https://internal.vacloud.us/wiki/pages/81X0Z7X5/Perceptive_Reach_Deliverables.html)

<https://internal.vacloud.us/wiki/pages/81X0Z7X5/Perceptive_Reach_Deliverables.html>

* [Requirements Traceability Matrix (RTM)](https://internal.vacloud.us/wiki/pages/81X0Z7X5/Perceptive_Reach_Deliverables.html)

<https://internal.vacloud.us/wiki/pages/81X0Z7X5/Perceptive_Reach_Deliverables.html>

* [Database Design Specification](https://internal.vacloud.us/wiki/pages/81X0Z7X5/Perceptive_Reach_Deliverables.html)

<https://internal.vacloud.us/wiki/pages/81X0Z7X5/Perceptive_Reach_Deliverables.html>

* [Data Analytics Sandbox Specification](https://internal.vacloud.us/wiki/pages/81X0Z7X5/Perceptive_Reach_Deliverables.html)

<https://internal.vacloud.us/wiki/pages/81X0Z7X5/Perceptive_Reach_Deliverables.html>

* [Dashboard Design Document](https://internal.vacloud.us/wiki/pages/81X0Z7X5/Perceptive_Reach_Deliverables.html)

<https://internal.vacloud.us/wiki/pages/81X0Z7X5/Perceptive_Reach_Deliverables.html>

## Acronyms and Abbreviations

Table 1: Acronyms and Abbreviations

| Acronym | Term |
| --- | --- |
| Army STARRS | Army Study to Assess Risk and Resilience in Service members |
| API | Application Program Interface |
| BIRLS | Beneficiary Identification Records Locator System |
| CD | Compact Disk |
| CDC | Center for Disease Control |
| CDW | Corporate Data Warehouse |
| CI | Continuous Integration |
| DB | Database |
| DoD | Department of Defense |
| EA | Enterprise Architecture |
| ETL | Extract, Transform, Load |
| FMQL | FileMan Query Language |
| GB | Gigabyte |
| HMAC | Hash-based Message Authentication Code |
| HTTPS | Hypertext Transfer Protocol Secure |
| ICD | International Classification of Diseases |
| IM/IT | Information Management/Information Technology |
| IRDS | Integrated Reach Database System |
| IT | Information and Technology |
| JDBC | Java Database Connectivity |
| JSON | JavaScript Object Notation |
| NDI | National Death Index |
| ODBC | Open Database Connectivity |
| OIT | Office of Information and Technology |
| OMHS | Office of Mental Health Services |
| PMP | Project Management Plan |
| REST | RESTful |
| RPC | Remote Procedure Call |
| RSD | Requirements Specification Document |
| RTM | Requirements Traceability Matrix |
| SAS | Statistical Analysis System |
| SDCD | State Death Certificate Data |
| SDR | Suicide Data Repository |
| SFTP | Secure File Transfer Protocol |
| SMITREC | Serious Mental Illness Treatment Resource and Evaluation Center |
| SPAN | Suicide Prevention Applications Network |
| SQL | Structured Query Language |
| SSIS | SQL Server Integration Services |
| SSN | Social Security Number |
| TB | Terabyte |
| TBD | To Be Determined |
| TRM | Technical Reference Manual |
| UBHC | University Behavioral Health Care |
| UI | User Interface |
| VA | Department of Veterans Affairs |
| VAMC | VA Medical Center |
| VCL | Veterans Crisis Line |
| VHA | Veterans Health Administration |
| VISN | Veterans Integrated Service Networks |
| VistA | Veterans Health Information Systems and Technology Architecture |
| VLER | Virtual Lifetime Electronic Record |
| VSO | Veterans Service Organizations |
| VSSC | VHA Support Service Center |

# Background

## Overview of the System

The IRDS development and field pilot combines technology outreach and clinical support to realize a clinically based data-driven early intervention and treatment solution aimed at suicide prevention. The application will include capability for analyzing multiple and integrated data sets with cutting-edge data analytic techniques and visualizations to identify at-risk individuals and populations and provide proactive and secure notifications of these results to Veteran support services.

A central component of the IRDS project is VA’s Suicide Data Repository (SDR.). The SDR is a centralized SQL Server database consolidating multiple sources of data containing suicide and mortality data of Veterans. The IRDS project proposes to expand the capabilities of the SDR to include new interfaces to clinical data sources, integrated data analytics/predictive modelling capabilities, a surveillance dashboard, and secure messaging – thus enhancing SDR capabilities to include upstream suicide intervention.

As shown in **Figure 1**, IRDS will be an integrated system that builds upon the existing SDR with the following capabilities:

* **Reach Database.** A SQL database used to aggregate new data sources and relevant SDR data.
* **Data Analytics Platform and Dashboard.** An integrated collection of analytics and visualization tools, including a surveillance dashboard aimed at identifying at-risk individuals and populations
* **Direct Messaging.** A method to construct and transmit a secure message to authorized outreach and intervention service providers.
* **Outreach and Intervention.** A pilot workflow that includes the process by which outreach and intervention resources are notified and act upon the data provided.



Figure 1: Conceptual Program Design

The underpinning technology and data analytics platform will provide methods by which at-risk populations and individuals can be identified. Specifically, we propose a programmable and configurable solution that can be tailored and enhanced over time as more data sources become available and as clinical research identifies new risk factors. As depicted in the bottom-center of  **Figure 2** a significant component in this effort is the identification of Veteran-specific risk factors, a precursor to the design of an automated reporting model.

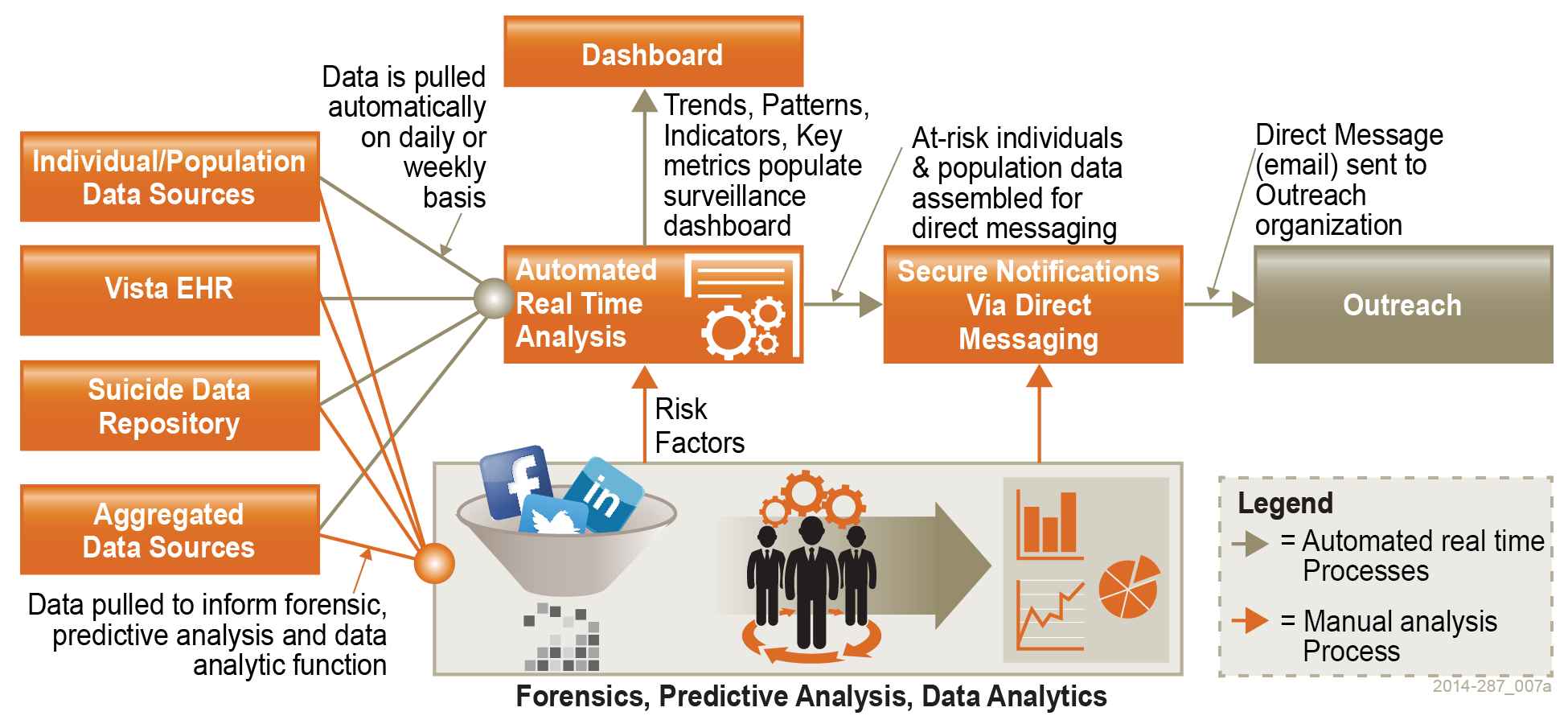


Figure 2: Conceptual System Design

This effort includes data analysis and predictive modelling, done in collaboration with VA stakeholders and clinical subject matter experts, while at the same time leveraging research data from sources such as DoD (e.g. Army Study to Assess Risk and Resilience in Service members [Army STARRS] and the Military Suicide Research Consortium) and VA (e.g. Center of Excellence for Suicide Prevention). Harnessing the automated reporting model, information will be presented in a customizable national surveillance dashboard and transmitted to authorize officials and Veteran support services organizations via Direct Messaging.

### SDR Database Background

As previously mentioned, the SDR is a centralized SQL Server database, which currently consolidates multiple sources of data containing suicide and mortality data of Veterans. The SDR will be one of the data sources for the IRDS innovation. Data from the sources below are periodically imported into the SDR. Sources include:

* National Death Index (NDI) based on DoD and VA NDI search criteria
* State Death Certificate Data (SDCD)
* Suicide Prevention Applications Network (SPAN)
* Veterans Crisis Line (VCL)

The SDR represents a significant step towards the deployment of a comprehensive suicide surveillance system as it enables a systematic collection of data on completed suicides, attempted suicides, and suicide ideations. The analytical value contained within the rich data sets of the SDR is largely untapped but primed for extraction via digital tagging, discovery, and analysis. The SDR provides a means for VA to quantify and monitor the scope of the suicide problem by analyzing SDR data using analytics to identify characteristics associated with higher or lower risk of suicide, and track changes in the suicide rate over time to evaluate the effectiveness of suicide prevention interventions.

Importantly, the SDR provides a view into the unique characteristics of Veterans pertaining to mortality, suicide ideations, suicide attempts, crisis line communications, and mental healthcare. The data contained within the SDR combined with additional real time clinical data and other data sources, such as public records, presents an opportunity to more accurately identify at-risk Veterans using data unique to Veterans.

## Overview of the Business Process

[Section 2.4](#_User_Characteristics) of this document outlines the major use cases associated with the IRDS application. Of these use cases, the ones that have the most significant business process component are Surveillance and Upstream At-Risk Notification. The business process diagrams below represents a high level overview of how the IRDS application will integrate within these use cases, however it should be noted that the details of each business process may vary on a case by case and facility by facility basis. As more business processes are identified during the development process, this document will be updated.

1. Data sources are imported into the IRDS system via SQL Server Integration Services (SSIS) import solutions. Each data source will have its own SSIS package (.dtsx). In the case of importing data from a VistA installation, one or more RPC calls will be executed from within an SSIS import. The import solution will transform the data and load into the appropriate tables in the Reach database.
2. An R program is run periodically (every year or so) on the production server to update the Risk model. The results of the run are stored in a table in the Reach database.
3. On a regular basis (daily, weekly) a SQL Server process runs that does surveillance against a list of Veterans tracked in the Reach database tables against the risk model. The process identifies who are at high risk of attempting suicide.
4. The results of the surveillance run are emailed to the appropriate contacts for those Veterans via a secure message that leverages the VA Virtual Lifetime Electronic Record (VLER)architecture.
5. A system user will access the dashboard via a compatible web browser. The dashboard is a web application that provides a visual presentation in the web browser of data that has been accessed via a REST API, which queries the Reach database for both specific and aggregate data regarding high risk Veterans at their management level (region, state, VISN, VAMC). The query results are passed to the client browser and populated in the web page.



Figure 3: Dashboard Surveillance



Figure 4: Direct Message, New Veteran Identified



Figure 5: Direct Message, Veteran Data Update

Table 2: Business Processes

| Business Process ID | Business Process Name | Type | Owner | Description |
| --- | --- | --- | --- | --- |
| 1 | Dashboard Surveillance | Modernized | VA Mental Health Staff | High-level description of how VA staff identifies high-risk Veterans and adds them to the local “High Risk List,” the primary tool for tracking and monitoring high risk Veterans within a facility’s service area. Staff will also use the dashboard for ad hoc research and lookups not associated with a defined business process. |
| 2 | Direct Message, New Veteran Identified | New | VA Mental Health Staff | Description of automated message generation and delivery when the application has identified a Veteran not previously displayed in the system. |
| 3 | Direct Message, Veteran Data Update | New | VA Mental Health Staff | Description of automated message generation and delivery when the application identifies a Veteran who has been previously identified, but has a significant negative event or trigger in the database which may induce an increased risk for suicidal behavior. |

## Business Benefits

Surveillance, identification of risk and protective factors, and interventions are three components of a holistic suicide prevention program. IRDS aims to implement and test this approach using data, technology and clinical expertise to establish a systematic approach, automating the data collection, data analysis/predictive modelling, identification of risk factors and Veterans at risk, notification, reporting, and continuous monitoring processes.

The IRDS concept will address two major elements of suicide intervention, information and time.

* Can information on Veterans both under and outside VHA care be aggregated to produce a useful prediction of suicide risk?
* Can effective interventions be developed and deployed in time to avoid problems from escalating into crises?

We contend that the proposed model will realize a real-time surveillance and intervention solution that will answer both questions in the affirmative.

|  |
| --- |
|  |
| **Source:** CRS analysis of major components of U.S. Department of Health and Human Services (HHS) Office of the Surgeon General and National Action Alliance for Suicide Prevention, 2012 National Strategy for Suicide Prevention: Goals and Objectives for Action, Washington DC: HHS, September 2012 |

Figure 6: Surveillance Process Model

|  |
| --- |
| Impact |
| * The IRDS innovation will dramatically change the manner in which VA plans, funds, manages, and assesses suicide intervention and prevention programs. * IRDS will provide a cost effective framework from which health data can be studied, hypotheses tested, and where proven analytic methods can be automated, including the automation of report and notification messages. * As risk factors change over time, and as new risk factors are identified, new analytical models and new data sources can be incorporated into IRDS and the method of identification, notification, and intervention can be re-applied. * Through near real time analysis, and a surveillance dashboard, IRDS will allow VA to respond to regional and temporal events and trends with more agility and precision, while also providing a means to monitor and measure the results from specific initiatives. * Likewise, the precision provided in the system will reduce program expenditures as more focused funding can be applied, and ineffective programs can be assessed and terminated. |
| Benefits |
| * The most significant benefit of the IRDS innovation will be the reduction in attempted and completed suicides through early identification, effective intervention, and early treatment. * Due to the early identification and treatment for a broad range of clinical, socioeconomic, and environmental conditions, outreach and intervention programs leveraging IRDS will promote wellness and are therefore likely to decrease the probability of more serious health conditions in the future. * The IRDS solution is aimed at identifying individuals and populations with characteristics that may, if left untreated, increase the probability of future suicidal crises. That is, we are looking for early warning signs, and the application of preventative care that will reduce suffering, suicides and treatment expenses. * The organization and visualization of near real time information will simplify VA business processes, minimizing or eliminating the costs of producing static reports, and eliminating the costs of actions taken on stale data. * Program funding and resources can be optimized and tailored to specific regional needs, preventing waste. * The IRDS innovation will highlight the need for policies and governance surrounding the use of public and non-public data to manage both population and individual health outcomes. |
| Scalability |
| * The IRDS system will provide a framework of data aggregation, data analysis/predictive modelling, reporting, notifications, and visualizations. The integrated system and each of the individual components will be defined and architected using standards and design paradigms that enable interoperability and scalability. * This framework provides a platform on which new analytic techniques, tools, and theories can be tested and studied, eliminating the need to construct new research platforms from scratch – thus saving time and money on future programs. * This framework and reference design will be the basis from which an enterprise level solution can be designed, implemented, and deployed within VA, consistent with OneVA Enterprise Architecture principles and requirements. |

Figure 7: System Benefits

## Assumptions and Constraints

The details of the system design are subject to change as requirements are being gathered in parallel with development.

This SDD will be regularly updated and identified with version numbers to describe the expanded system design.

Unless otherwise stated, software will be open source and compliant with VA's Technical Reference Manual (TRM).

### Design Assumptions

* The overall design and development process is based on iterative development and rapid prototyping incorporating key stakeholder input allowing for design factoring and enhancement.

### Design Constraints

* System designers have attempted to utilize open source tools wherever possible, including the design of the user interface / front end presentation layer of the system, testing tools, and statistical / analytics tools.
* System designers used VA tools approved for use in the VA Technical Reference Model (TRM) or have requested a waiver for any tools not included in the TRM

### Design Trade-offs

The system responsibility is to provide a feature-rich and responsive Dashboard UI. The Reach Database provides the consolidated data services to drive the UI. It is expected that the number of users will be relatively low but the responsiveness of the Dashboard to query and display near real-time data is essential. Thus, the IRDS Dashboard is architected to scale by ensuring that performance does not degrade as the number of users increases and as additional data becomes available. This architectural focus can be demonstrated in the following design trade-offs:

* Maximize the responsiveness of the UI by utilizing client side controllers, and only making network calls to retrieve data. This is achieved through the use of a Single Page Application
* Favor an asynchronous API interface to maximize performance. If an API call is synchronous, it means that code execution will block (or wait) for the API call to return before continuing causing possible performance issues. However, asynchronous calls do not block (or wait) for the API call to return from the server allowing execution to continue in the application. In light of this approach, the Dashboard will cache data when sensible in order to continue performance gains. This approach is further explained in Section 5.2.5 (RESTful Data Services).
* Favor the usage of micro-services, rather than traditional web server clusters, to maximize discrete scalability options and to ensure that services remain loosely coupled.
* The IRDS database is composed on data aggregated and refined through an analytics platform and does not include all available data from external data sources utilized by the analytics platform.

## Overview of the Significant Requirements

### Overview of Significant Functional Requirements

The table below includes an overview of the major user requirements associated with the proposed solution. A full listing of the project’s Product Backlog is maintained in Jira at <https://opensourceehr.atlassian.net/secure/Dashboard.jspa>. In addition, a more detailed snapshot of the project’s major functional requirements will be included in the project’s RTM. The PR tags in the table below, for example, “PR-158,” derive from the tracking system in Jira.

Table 3: Functional Requirements

|  |  |
| --- | --- |
| Jira Key | **Description** |
| PR-158 | As an Outreach Provider, I want to be sent secure notification via a Direct Message of at-risk Veterans and populations so I can provide outreach services to these groups. |
| PR-160 | As a member of VA leadership, VA Center of Excellence for Suicide Prevention staff, VA Mental Health leaders, or VA Suicide Prevention Coordinator, I want to view a surveillance dashboard with results produced from the continuous monitoring and processing of linked data sources so I can monitor and understand Suicide Outreach outcomes. |
| PR-161 | As a Researcher, I want to access the access the tools and data in the application so I can perform research-related tasks and projects. |
| PR-162 | As a Researcher, I want to generate reports using the data and automated tools in the application so I can use reports as management and communication tools. |
| PR-163 | As a User, I want to edit, add to, and create new IRDS Risk Stratification Models and mapping to interfaces so the application can be updated over time. |
| PR-346 | As an Outreach Provider, I want to view a Direct Message that highlights Veterans at high risk for suicide so I can provide outreach services to them. |
| PR-349 | As a Dashboard User, I want to log in to the Perceptive Reach application. |
| PR-351 | As a Dashboard User, I want to see data from my "home" facility when I log in. |
| PR-352 | As a Dashboard User, I want to view newly identified at-risk Veterans during a specific time frame. |
| PR-353 | As a Dashboard User, I want to view information about groups of at-risk Veterans filtered by geographic area. |
| PR-354 | As a Dashboard User, I want to view information about groups of at-risk Veterans filtered by other data attributes (see details in Jira ticket). |
| PR-355 | As a Dashboard User, I want to view information related to the change in suicide rates over time. |
| PR-356 | As a Dashboard User, I want to pick and choose which screen elements I see on the dashboard, so I can first see only the data that is important to me. |
| PR-357 | As a Dashboard User, I want to move screen elements I see on the dashboard so I can customize the look of the dashboard to suit my preferences. |
| PR-505 | As an Outreach Provider, I want to view a Direct Message when a Veteran experiences a high risk trigger or event, so I can provide outreach services to them. |
| PR-521 | As a Perceptive Reach user, I want to see a sortable / filterable list of high risk Veterans. |
| PR-522 | As a Perceptive Reach user I want to click a high risk Veteran list so I can see more detailed information about the Veteran I selected. |
| PR-158 | As an Outreach Provider, I want to be sent secure notification via a Direct Message of at-risk Veterans and populations so I can provide outreach services to these groups. |
| PR-160 | As a member of VA leadership, VA Center of Excellence for Suicide Prevention staff, VA Mental Health leaders, or VA Suicide Prevention Coordinator, I want to view a surveillance dashboard with results produced from the continuous monitoring and processing of linked data sources so I can monitor and understand Suicide Outreach outcomes. |
| PR-161 | As a Researcher I want to access the access the tools and data in the application so I can perform research-related tasks and projects. |
| PR-162 | As a Researcher, I want to generate reports using the data and automated tools in the application so I can use reports as management and communication tools. |
| PR-163 | As a User, I want to edit, add to, and create new IRDS Risk Stratification Models and mapping to interfaces so the application can be updated over time. |
| PR-346 | As an Outreach Provider, I want to view a Direct Message that highlights Veterans at high risk for suicide so I can provide outreach services to them. |
| PR-349 | As an SPC, I want to log in to the Perceptive Reach application. |
| PR-351 | As an SPC, I want to see data from my "home" facility when I log in. |
| PR-352 | As an SPC, I want to view newly identified at-risk Veterans during a specific time frame. |
| PR-353 | As an SPC, I want to view information about groups of at-risk Veterans filtered by geographic area. |
| PR-354 | As an SPC, I want to view information about groups of at-risk Veterans filtered by other data attributes (see details in Jira ticket). |
| PR-355 | As an SPC, I want to view information related to the change in suicide rates over time. |
| PR-356 | As an SPC, I want to pick and choose which screen elements I see on the dashboard, so I can first see only the data that is important to me. |
| PR-357 | As an SPC, I want to move screen elements I see on the dashboard so I can customize the look of the dashboard to suit my preferences. |
| PR-505 | As an Outreach Provider, I want to view a Direct Message when a Veteran experiences a high risk trigger or event, so I can provide outreach services to them. |
| PR-521 | As a Perceptive Reach user, I want to see a sortable / filterable list of high risk Veterans. |
| PR-522 | As a Perceptive Reach user, I want to click a high risk Veteran list so I can see more detailed information about the Veteran I selected. |

# Conceptual Design

## Conceptual Application Design

The application design we propose will be a programmable and configurable solution that can be tailored and enhanced over time as more data sources become available and as clinical research identifies new risk factors. As depicted in the bottom-center of **Figure** **8** a significant component in this effort is the identification of Veteran-specific risk factors, a precursor to the design of an automated reporting model.

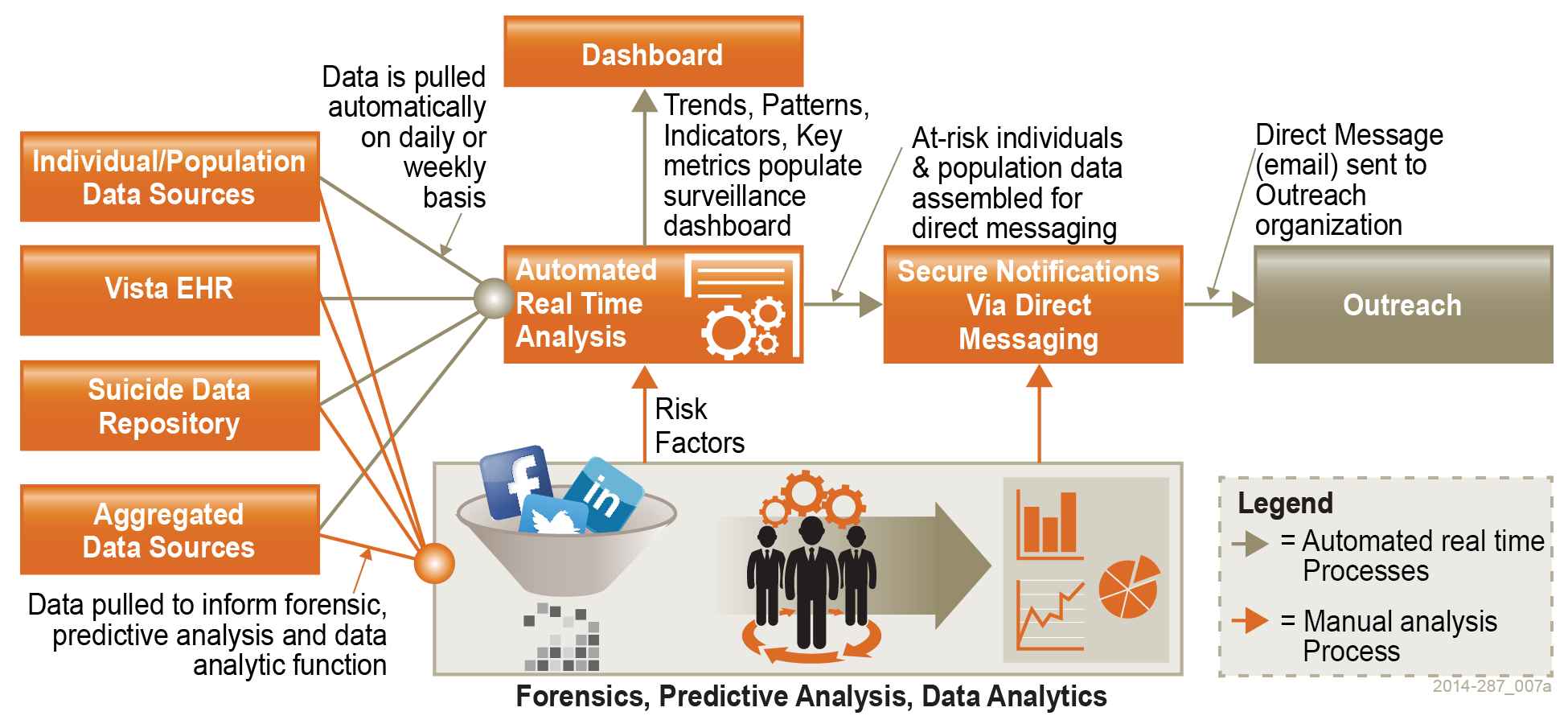


Figure 8: Conceptual System Design

### Application Context

While traditional interventions must remain in place, this proposed approach introduces an upstream intervention. Geographic “hotspots” will be identified for issues such as Veterans defaulting on mortgages, and craft community interventions using the VA, Veterans Service Organizations (VSOs) and other existing resources aimed specifically at the identified problem. As individuals self-identify, connections to existing, appropriate resources will be made and follow-up contact with the Veteran can determine whether service was received and the problem was mitigated. The IRDS solution is a novel and complimentary approach to more traditional forms of suicide prevention.

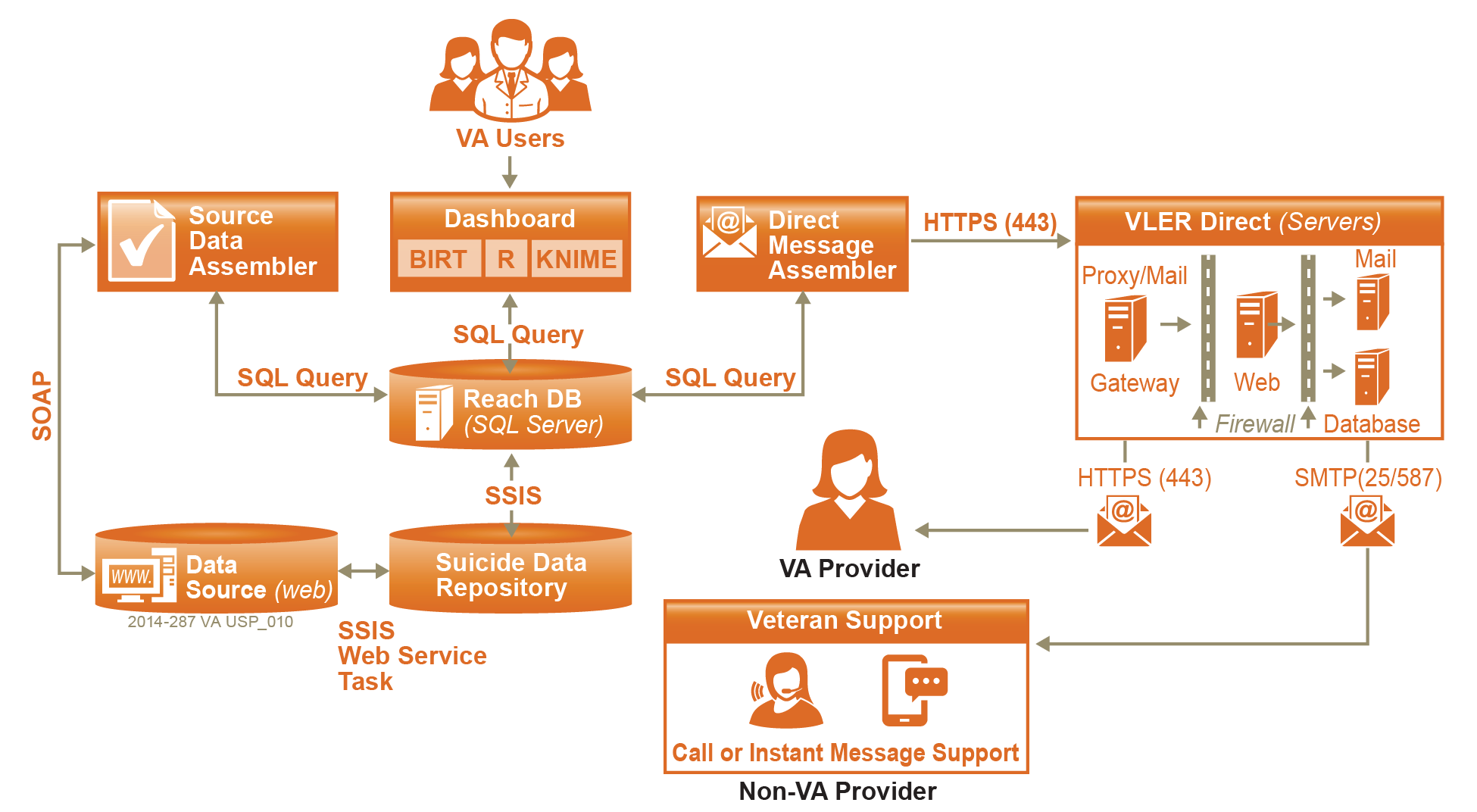


Figure 9: System Diagram

Table 4: (Grouping) Application Context Description Object

| ID | Name | Description | Interface Name | Interface System |
| --- | --- | --- | --- | --- |
|  | Dashboard | The IRDS dashboard will be an operations style surveillance tool providing near real-time views of regional and temporal data, trends, events, and key performance metrics. | SQL Query | Reach DB,  VA Users |
|  | Direct Message Assembler | To facilitate intervention through outreach programs, the IRDS system will create and transmit notification messages via Direct Messaging to VA designated and authorized intervention service providers. | SQL Query, HTTPS | Reach DB, VLER Direct |
|  | Data Analytics Platform | A combination of business intelligence (BI) tools and data analytics packages which will satisfy three key objectives: data integrity, flexibility, and simplicity. Examples: BIRT, R, & KNIME | SQL Query | Reach DB,  VA Users |
|  | Reach DB | The Reach Database will be developed to leverage the SDR and other data sources to create a robust data collection against which analytics can be performed. | SSIS, SQL Query | Dashboard,  SDR, VLER Direct, External Sources |

### High-Level Application Design

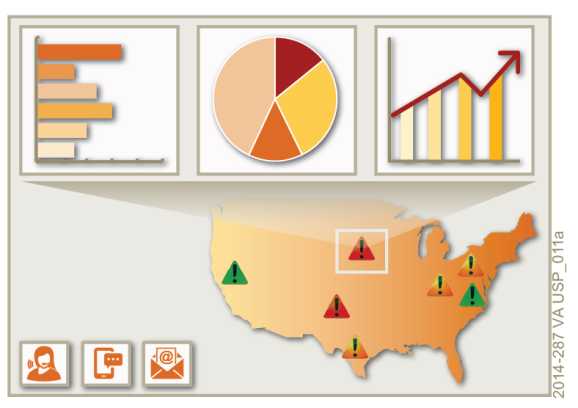
The IRDS System will involve visualization tools for the IRDS System. Data visualization tells the story of the analysis results using charts, tables, and other graphics. Data visualization is the step that makes the body of data analytics work accessible to a broad range of stakeholders.

Figure 10: Dashboard Mockup

The surveillance dashboard will be accessed through a VA approved web browser. The dashboard will be configurable, providing different user groups distinct views that meet their business needs. As shown in the figure the dashboard will be designed to support interactive viewing and formatting, and created using the open source tools. The design of the dashboard will involve the input of stakeholders and end users. The application will leverage a Node.js runtime environment, which will manage the server-side functions (http server, database connections, APIs, and web service interfaces). The visualization components will be browser based and built using Angular.js, an open source web application framework. This will provide for flexibility and ease of coding.



Figure 11: Conceptural Dashboard Application Design

## Conceptual Data Design

### Project Conceptual Data Model

The Reach Database contains the following tables:

Data Tables

The Reach database will store source imported data on two levels:

* Veteran Demographics
* Veteran Case Data

For each Veteran, a master record will be created in the ‘Veterans’ table and a unique ID will be assigned, ‘ReachID’. The table will also contain basic demographic information for that individual, such as Name, SSN, DOB, Gender, etc… The table also contains a risk ‘Score’ field, which the surveillance component of the system will populate.

All other data elements imported for Veterans will be stored in case level tables. There will be a one to many relationship from the Veteran table to any case level table. For instance, any data that may be available from the SDR on previous suicide attempts for that individual will stored in a suicide attempt table. For each suicide attempt that is documented for a Veteran, a ‘SuicideAttempt’ record will be created, including a ReachID column that will link the suicide Attempt record back to the Veteran.

Reference Tables

Examples are lists to be used for reporting and normalizing of the data such as a list VAMCs and a list of ICD Codes that contain a diagnosis description associated with each code.

High Risk Veterans View

The High Risk Veterans view will roll up data from the Veterans master list and case level data associated with each Veteran to show all vets that are identified as high risk with their risk score and the factors that contributed to that score. The dashboard will access this view to populate the dashboard screens.

Risk Factors

The Risk Factors determined by the IRDS Risk Model will be captured in the RiskFactors table(s). This table will be updated any time the Risk model is updated, by running the methodology encapsulated in an R program that exists on the IRDS server.

System Tables

Tables used by the dashboard such as User roles and Preferences (see Data Access).



Figure 12: IRDS Reach Database Logical Data Model

### Database Information

Table 5: Database Inventory

| Database Name | Description | Type | Steward |
| --- | --- | --- | --- |
| Reach | SQL Server database(s) that will   1. Import data from SDR and other internal/external sources 2. Store Analytics output to be used by IRDS dashboard and messaging | Create | TBD |
| SDR | See section 5.5, IRDS Data Sources table | Interface |  |

### Data Import Design

#### SQL Server Integration Services (SSIS)

SSIS will be the primary tool for importing external data sources into the IRDS Reach database

For a specific data import, an SSIS package will be developed to

1. Make a connection to the source (SQL table, text file, other)
2. Import the data into a staging area
3. Make the appropriate data transformations (cleaning, standardization)
4. Load the transformed data into the appropriate Reach data store tables

The execution of SSIS packages (.dtsx files) can be automated by scheduling them as a Windows process via SQL Server Agent.

#### Remote Procedure Calls (RPC)

The VA uses the Veterans Health Information Systems and Technology Architecture (Vista) system, for managing Veterans health data. Data will be imported into the IRDS system directly from VistA using RPC calls. VistA data is stored against a MUMPS back end, which uses text-based files for data storage. For each set of VistA data imported into IRDS:

1. Either a custom RPC will be written (in M) or a currently existing one will be leveraged
2. An automated java process will execute the RPC and return the query results in text format
3. Those results will be stored in a flat file on the IRDS server to be imported into the reach database via an SSIS package



Figure 13: IRDS Data Import Process Flow

## Conceptual Infrastructure Design

The architecture of this system supports cloud computing and the principles of OneVA EA. The IRDS system is intended to be implemented on a virtual or single cloud resource.

### System Criticality and High Availability

The IRDS is not a high availability system. System redundancy is not a requirement within this innovation program at this time. The availability of this system is expected to be 365/24 with the exceptions of the times when the underlying infrastructures are not available due to maintenance.

### Special Technology

As the IRDS Innovation is granted approval or waivers for any special technologies listed below, the table will be updated.

Table 6: Special Technology Requirements

| Special Technology | Description | Notional Location | TRM Status |
| --- | --- | --- | --- |
| Kinme | An open source data analytics, reporting and integration platform. | This would be deployed within the IRDS Solution | No |

## 

## System Architecture

The system developed under the Perceptive Reach (IRDS) project will be designed to run on a cloud-based environment consistent with the principles of OneVA EA. The following diagram (also shown in section 4.1) provides an overview of the System Architecture.

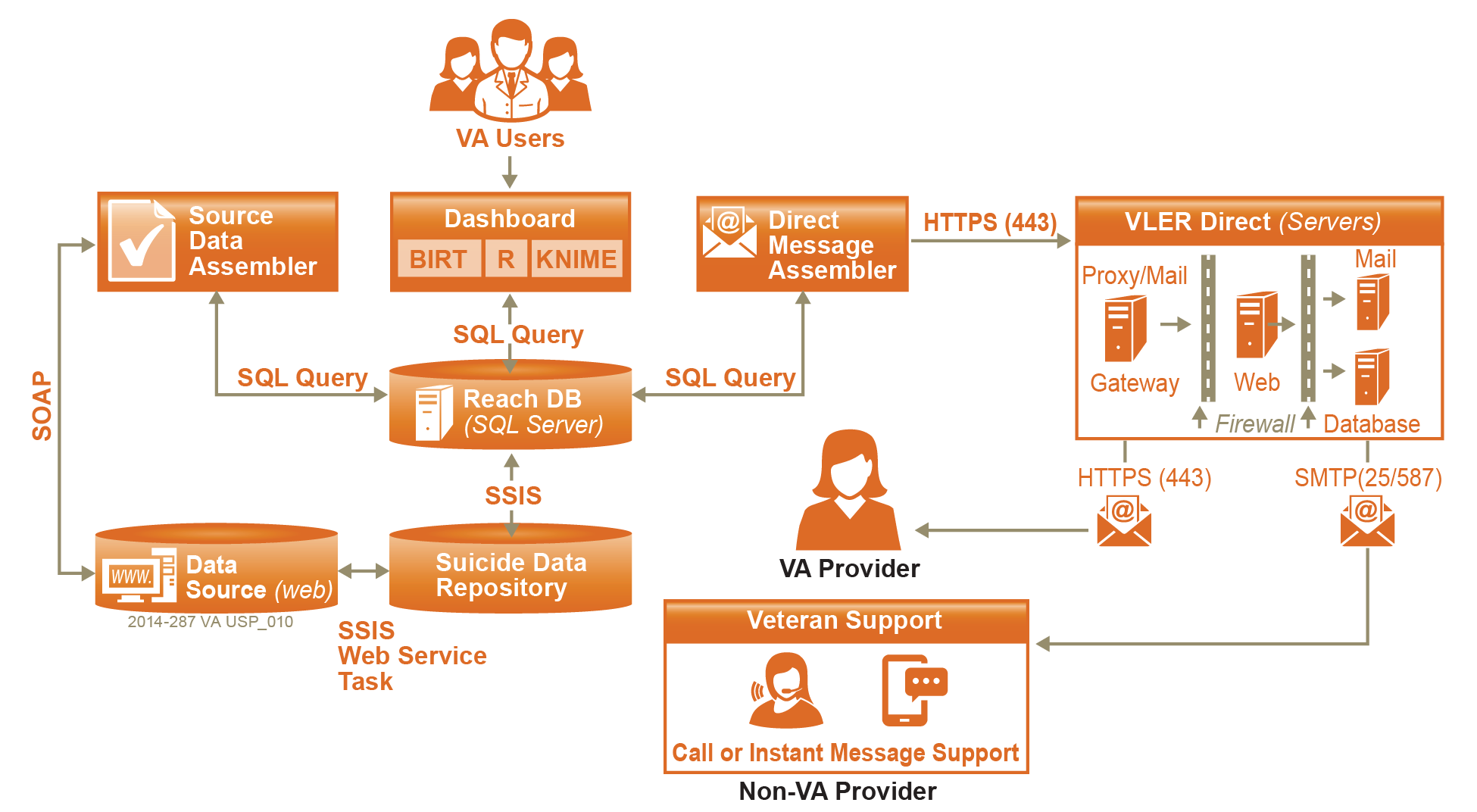


Figure 14: System Architecture

## Hardware Architecture

The solution will be deployed within the VA’s enterprise environment.

Table 7: Initial Requirements (Cloud Based – Prototype Server)

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Type** | **OS** | **Processor** | **Memory** | **Storage** | **Users** | **Applications** |
| Cloud | MS Windows Server 2012 64-bit | Intel Xeon CPU E5-2670, 2.6GHZ | 16 GB | 60 GB | 16 | MS SQL Server 2012 Enterprise Edition |

Table 8: Planned Requirements (Cloud Based – Development/Test Server)

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Type** | **OS** | **Processor** | **Memory** | **Storage** | **Users** | **Applications** |
| Open | MS Windows Server 2012 64-bit | Intel Xeon CPU E5-2670, 2.6GHZ or better | 32 GB | 500 GB | 16 | MS SQL Server 2012 Enterprise Edition |

Future Requirements:

* The future requirements will be provided as requirements are finalized.

For further details on the Hardware Architecture, please refer to [Section 5.1](#_Hardware_Detailed_Design).

## Software Architecture

The IRDS innovation will develop and demonstrate a new SQL database that aggregates both VA and non-VA data sources to be used to facilitate identification of at-risk individuals and populations, an integrated data analytics solution that includes open source data analysis and visualization tools, and an open standard based secure messaging solution to inform authorized individuals of analysis results. For details on the Software Architecture, please reference **Section 5.2 Software Detailed Design.**

## Continuous Integration / Continuous Delivery

The IRDS solution will be maintained through a Continuous Integration / Continuous Delivery process (CI). This will isolate changes as they are added to the larger code base and are immediately tested and reported on. The goal of CI is to provide rapid feedback so that if a defect is introduced into the code base, it can be identified and corrected as soon as possible.

## Network Architecture

The following illustration shows the notional network configuration leveraged by the IRDS system. The system will utilize VA LAN and WAN networking resources to transfer data from various data repositories to the IRDS database, and to support direct message emails sent to VAMC resources and to external resources such as partners at Rutgers University Behavioral Health Care (UBHC).

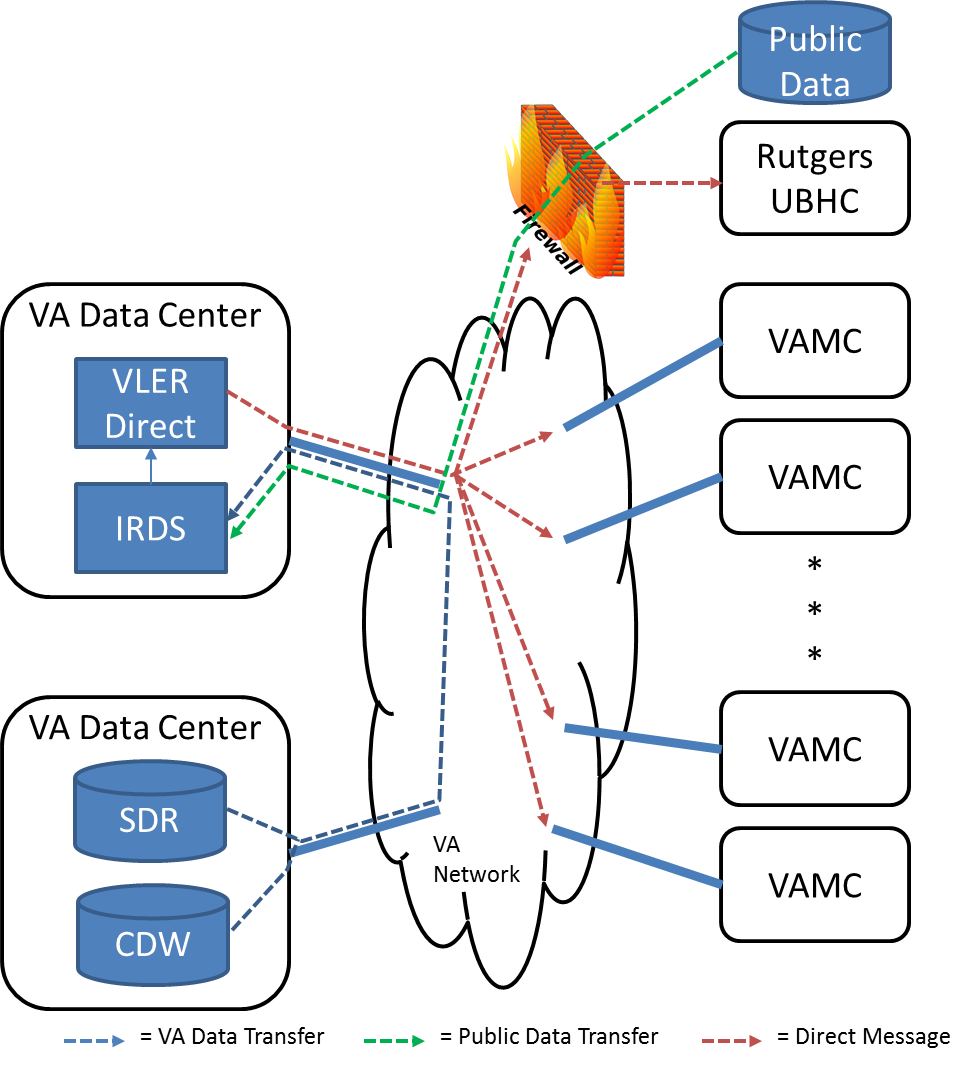


Figure 15: Notional Network Architecture

## Service Oriented Architecture / ESS

While the IRDS may be extended to provide Enterprise Shared Services in the future, the innovation objectives do not include the delivery of web services functionality. However, it is expected for IRDS to consume the VLER Direct ESS for Direct Message.

## Enterprise Architecture

IRDS conforms to the principles of OneVA Enterprise Architecture, utilizing technologies approved on the VA Technical Reference Model (TRM), open and standardized interfaces. Through standard design patterns and the use of virtualization and cloud technology, the IRDS architecture supports portability, modularity, and scalability consistent with VA directives.

# Detailed Design

The following section details each aspect of the IRDS solution. This will be expanded to include each component of the solution as each are defined via the requirements process

## Hardware Detailed Design

Table 9: Hardware Detailed Design

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Type** | **OS** | **Processor** | **Memory** | **Storage** | **Users** | **Applications** |
| Open | MS Windows Server 2012 64-bit | Intel Xeon CPU E5-2670, 2.6GHZ or better | 32 GB | 500 GB | 16 | MS SQL Server 2012 Enterprise Edition |
|  |  |  |  |  |  |  |

## Software Detailed Design

**Figure 16: Architecture Overview**

### Dashboard



Figure 17: Dashboard

The role of the dashboard is to provide visualization tools that display the status of metrics and key performance indicators (KPIs) for IRDS. The dashboard will consolidate and arrange preselected metrics and analytic results within a browser based web application. Node.js will provide the base architecture for the web server providing the primary services of the dashboard through HTML, JavaScript, and CSS to drive the behavior of the application.

* Node.js (JavaScript runtime engine)
* Express.js (Node.js module used to handle routing of HTTP calls)
* HTML, JavaScript, and CSS
* Angular.js (JavaScript framework for extending HTML)
* Bootstrap (powerful front-end framework for faster and easier web development)

#### RESTful Data Service

The Dashboard application will implement RESTful Data APIs to be access internally from the client code and provided by the server code. These services will provide basic Read, Write, Update, and Delete options for data stored in the Perceptive Reach database. All data related to Veterans will be accessed via Read Only APIs, however data related to the actual dashboard application such as user settings, application settings, and configuration parameters may be accessed via all of the available API actions. These RESTful APIs will be hosted in the server process based on Node.js using Express.js as the HTTP framework.

### Reach Database

Figure 18: Reach Database

* Platform: SQL Server 2012
* Data Importing and ETL Tools: SQL Server Integration Services, FMQL or RPC (for interfacing directly with VISTA)

### Analytics Platform

Figure 19: Analytics Platform

The enhanced risk model will be coded in the open source statistical language R. Once an enhanced model is finalized, the model coefficients will either stay in R or be transitioned into SQL code. If the model remains in R, this may make running the model daily potentially a more difficult process due to lengthy data processing times. The alternative to this possible issue might be to store the data processing and model coefficients in SQL. If this approach were used, R would still be used at a specified frequency to refresh the parameter estimates of the risk model.

* R 3.1.2 (a language and environment for statistical computing and graphics)
* RStudio (a powerful and productive user interface for R)

For more detailed information on the Analytics Platform, please see [Section 5.7](#_Data_Analytics_Platform).

### Direct Messaging

Figure 20: Direct Messaging

The Direct Messaging component will integrate into the existing VLER Direct Messaging architecture. The Direct Messaging will create unique messages based on new and existing data annotated in the IRDS. The application will run on a Node.js application server and communicated to the VLER Direct Messaging over the RESTful web services and hash-based message authentication code (HMAC) for secure authorization, sending, and receiving.

* Node.js (JavaScript runtime engine)
* RESTful web services (HTTP, XML, JSON)
* HMAC (hash-based message authentication code)

## Specific Requirements

For further details on the specific requirements, please refer to the IRDS Requirements Specification Document.

## Continuous Integration / Continuous Delivery

The IRDS solution will be maintained through a Continuous Integration / Continuous Delivery process. This is an automated process, initiated when source code is submitted to the GitHub source code repository. Jenkins detects the submission and initiates a build and test process utilizing build tools such as Gulp and testing tools such as Selenium, Cucumber, and Maven.

As shown in the figure below, the development and testing (Dev-Test) environment and the pre-production environment are very similar. The Dev-Test environment is located in the VA Cloud and will not connect to production systems or utilize any PII/PHI. All testing will be done with a test data test. The Pre-Production environment is located behind the VA firewall and will connect to production systems such as the SDR and CDW databases and utilize PII/PHI for development, testing and operations.

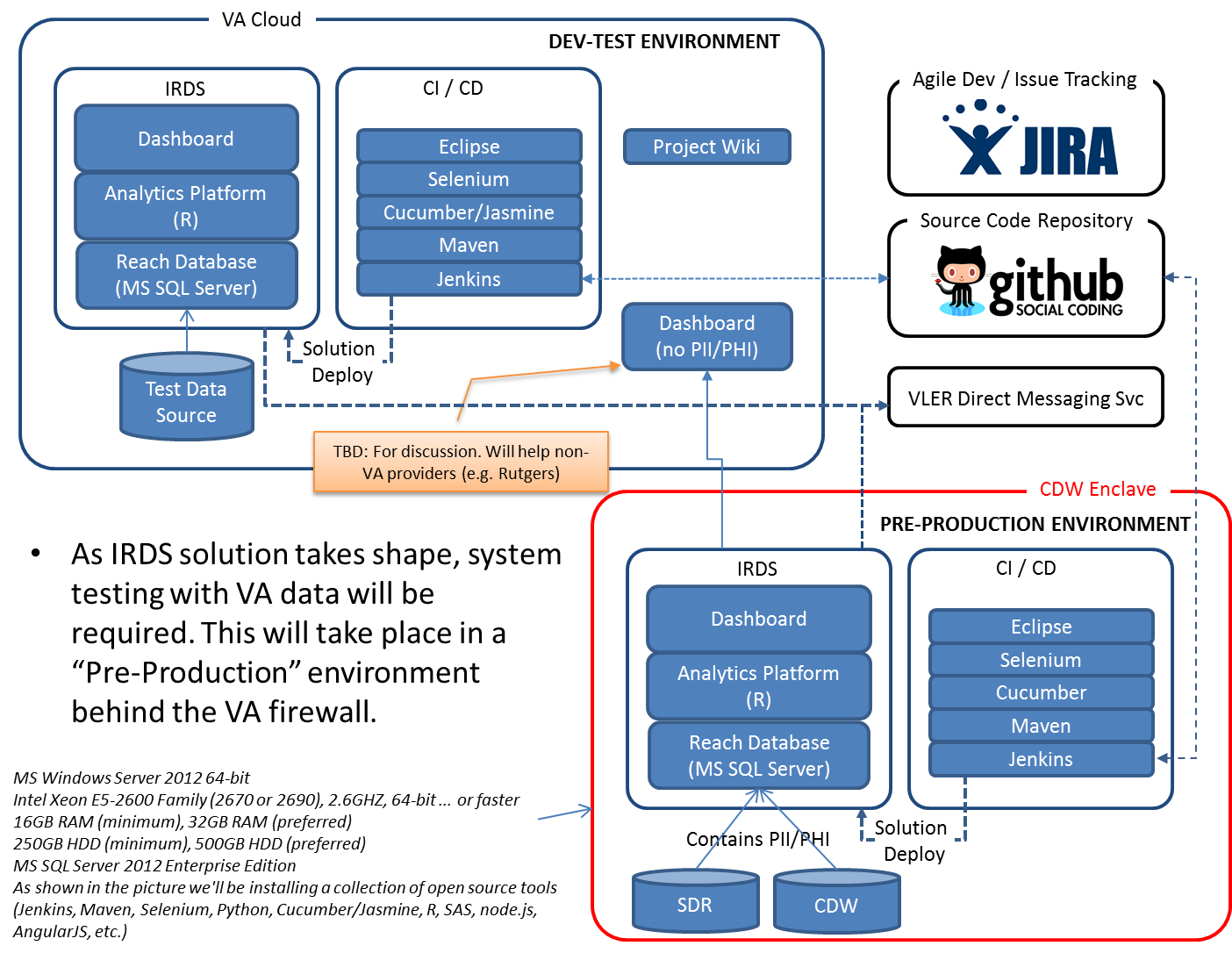


Figure 21: Continuous Integration

## Data Design

The Reach Database component of IRDS will use SQL implemented with Microsoft SQL Server Enterprise Edition 2012.The logical and physical data models are not yet defined.

At this time assumptions are:

1. SDR Data will be pulled directly from the SDR server/databases via a SQL connection and querying of the data
2. VBA data will pulled from SQL tables located at Corporate Data Warehouse (CDW) via a SQL connection and querying of the data
3. VISTA data will be accessed and imported through either currently exiting and/or custom RPC calls, or FMQL.
4. It is possible that some data sources may be interfaced by other means such as Web Service calls or file formats such as FLAT files.

Requirements are currently being gathered to determine what data will be imported into the IRDS system and how the data model will be designed to store the data. It has been decided that the are 2 components to the data

1. Tables to store source data that will be imported into the Reach database
2. Tables to store analysis output from R programs which have code to encapsulate a risk model and a surveillance model

Each of the data components above will be contained in their own schema. The source data tables will be stored in the default .dbo schema, and the analysis output tables will be stored in the ‘Analytics’ schema. An additional schema may be added for system tables and temporary tables accessed by processes such as SSIS packages.

Table 10: IRDS Data Sources

| ID | Name | Description | Internal to VA |
| --- | --- | --- | --- |
| 1 | VA Suicide Data Repository (SDR) | Periodically imports data from 4 VA data sources into SQL format:   * 1. National Death Index (NDI) Mortality Search Results   2. State Death Certificate Data   3. Suicide Prevention Applications Network (SPAN)  1. 4. Veterans Crisis Line | X |
| 2 | Corporate Data Warehouse (CDW) | Warehouses multiple VBA and VHA data sources in SQL Format. | X |
| 3 | Veterans Health Information Systems and Technology Architecture VistA | Electronic health record system for VA patients. | X |
| 4 | LexisNexis (Potential) | Provider of multiple data sources such as legal, risk management, corporate, government, law enforcement, accounting, and academic. (POTENTIAL DATA SOURCE) |  |
| 5 | Public Access to Court Electronic Records (PACER) (Potential) | Electronic public access service that allows users to obtain case and docket information online from federal appellate, district, and bankruptcy courts. (POTENTIAL DATA SOURCE) |  |

### Physical Data Model

At this time, the following data elements have been identified to be imported into the Reach database. Their exact location within the VA data sources is being determined and access/documentation to those sources is in the process of being requested.

When the database table objects have been defined and created, a SQL Server database diagram will be inserted into this section.

General Veteran Demographics

* First Name
* Last Name
* Middle Name/Initial
* SSN
* DOB
* Race
* Gender
* Contact Information (address, phone)
* Emergency Contacts (name, relationship)

VHA Data

All data elements included in the current VA suicide risk model. For a complete list, please refer to the Risk Model Data Dictionary.

VBA Data

* Presence of co-morbid psychiatric disorders
* Alcohol abuse/dependence (303. ICD-9)
* Substance abuse/dependence (304. ICD-9)
* Deployment history, and location of deployments
* History of TBI
* Marital status
* Financial status
* Homeless
* Chronic/terminal illness (non-pain)
* Foreclosure/bankruptcy
* % of service connected disability
* Employment status
* First notice of death
* Enrollment in VR&E
* Recency of divorce
* Legal history (domestic violence, arrests, assaults, other violent/drug offenses)
* History of violence (with or without legal charges)
* Does Veteran have beneficiaries
* History of motor vehicle accidents
* Beneficiary travel reimbursement qualification
* Medical diagnoses to include chronic pain conditions
* Revocation of driver’s license
* Homeownership

### Data Dictionary

The data dictionary will be completed as access and documentation to the data sources become available and requirements are further defined.

### Veteran De-Duping Process

The Reach data model will contain one master list of all Veterans imported into the IRDS system. As each Veteran is imported into the database through one of the data sources, a record will be created for that Veteran in the Veteran table and a unique ID will be assigned. It is possible that an Individual might be imported into the system through multiple sources. When this occurs, the multiple Veteran records created for the individual will be merged and all case level data will be linked to that one merged Veteran record. To achieve this, a de-duping process will be run after each data import is run.

This de-duping process will be contained in an SSIS package, which does the following:

1. Reviews the Veteran table for possible duplicates
2. Duplicate groups are determined (2 or more records that could be duplicates)
3. For each duplicate group a text file is created containing the pertinent demographic information(Name, SSN, DOB, Gender) for all records in that group
4. A python program processes these files and evaluates all the records in a group via a record de-duping algorithm
5. The python program returns the results to the SSIS process via another set of text files
6. The results from these files are placed in a temporary SQL table and the Veteran table is reconciled accordingly, by merging any records that were determined to be duplicates

 Figure 22: Veteran De-duping Overview

### Non-DBMS Files

It is assumed that some VHA data will be imported from the VISTA system, which uses A MUMPS data store and will be accessed through either RPC calls and/or FMQL.

### Data View

Requirements are currently being gathered for which data sources and data elements from those sources will be imported into the reach database and what the relationship will be between those elements when they are stored in the database.

## Service Oriented Architecture / ESS Detailed Design

The IRDS provides a dashboard application accessed through a web browser, and secure messaging delivered through email client or web portal. The Data Analytics Platform will be accessed through direct access to the server on which the IRDS resides. IRDS will consume the VLER Direct ESS to deliver Direct Messaging. The overall architecture is still being developed for this section as requirements are finalized.

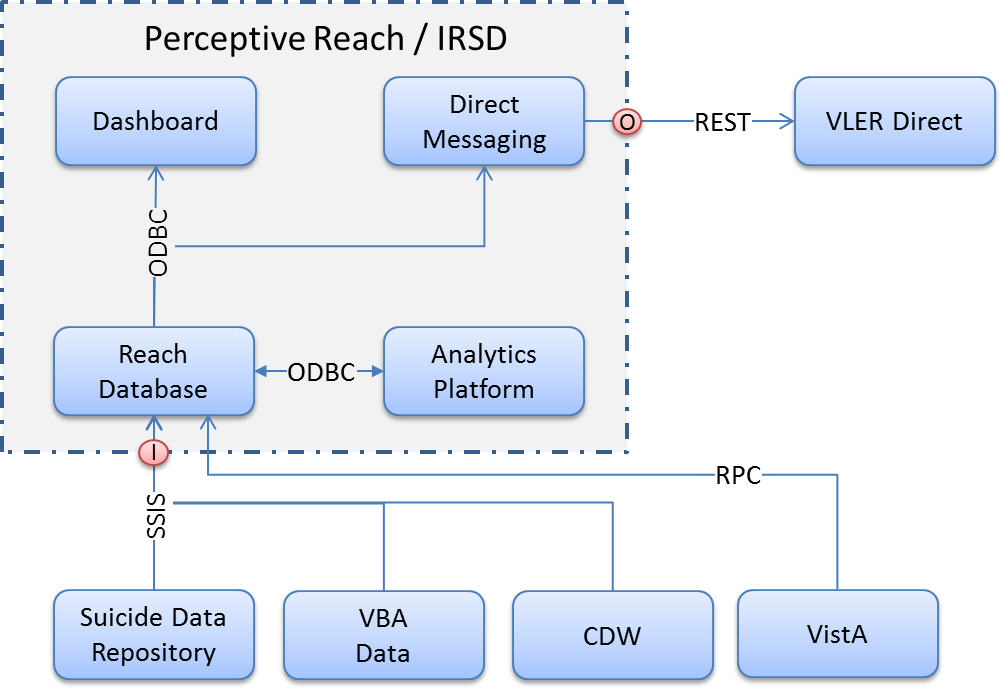


Figure 23: Architecture Overview

##### 

## Data Analytics Platform

To assess the functionality of the current Risk Model and its potential modifications/improvements, the Perceptive Reach team requires that the appropriate analytical tool is made available in the Perceptive Reach environment. There are a number of criteria that an analytical tool must meet for it to be appropriate for use:

* Open source tool able to integrate with mainline databases (i.e. JDBC, ODBC)
* Ability to calculate statistical tests (i.e., t-test, chi sq test, etc) and basic modeling (i.e., linear, logistic, etc)
* Execute advanced analytical/statistical models
* Ability to import, merge, manipulate, and export data
* Visualize data natively within tool interface
* Approved for use on the VA TRM
* Perceptive Reach team experience with tool

A number of tools were assessed to determine whether they satisfy each of the above mentioned criteria: SAS, SPSS, Stata, R, Python, Weka, Gephi, Eclipse, Business Intelligence and Reporting Tools (BIRT), Jasper Reports, Konstanz Information Miner (KNIME). The only tool that, met all of the criteria was R.

R is the leading open source statistical analysis tool used across many disciplines. The strength of R lies in its large and devoted academic and industry user base. This has allowed the language/software to contain a wide variety of in-built capabilities and remain an innovative tool.

## Risk Model

The main goal of the modeling effort is to continue to utilize and improve the existing suicide completion risk model developed by VA so that (1) near-time data on Veterans can be used to predict the risk of suicide completion for a specific Veteran and (2) the VA Suicide Prevention Coordinators and outreach staff are notified of Veterans with elevated risk levels for suicide and can take appropriate preventative actions. The current risk model uses a logistic regression framework and approximately 380 inputs from VHA and NDI data. Using this model as a starting point, potential improvements will be considered and implemented if they are shown to improve the out-of-sample predictive power of the existing model. Potential improvements may include but will not be limited to new data sources, alternative input variable definitions, testing of variable interactions, and alternative model structure.

New data sources may be utilized and variables constructed and tested to determine if they can be used as enhancements to the current risk model. Alternative methods to logistic regression may be considered to determine if such can improve the predictive power of the current risk model.

The strength and robustness of a predictive model is dependent on the data inputs that are used to develop the model. Thus, identifying a robust and reliable list of potential data inputs is essential. By reviewing academic literature and holding discussions with clinicians and other subject matter specialists on both risk and protective factors for suicide ideation and completion, a number of variable categories have arisen as critical inputs into the risk model: demographics, clinical, and socioeconomic. It is expected that the list of specific data inputs will increase as academic literature review and discussions with subject matter specialists continue.

After identifying the data points potentially useful for model development, data sources will need to be identified that capture the information on relevant variables. The current risk model uses data from VHA and NDI. It is expected that some of the additional data inputs will be derived from the same VHA and NDI data sources; however, additional data inputs will be considered from new data sources. One of the data sources that may be useful for model development is VBA. From the initial discussions with subject matter specialists on the VBA data, there are multiple datasets at VBA that contain information on Veterans including military service, financials, medical, and demographics.

Model input data will be housed within a SQL Server database behind a VA firewall. To be able to access the data, R statistical software will connect directly to the database to access the data for analysis.

The ultimate output from the enhanced risk model will be a Veteran-specific risk score based on the predicted likelihood of suicide risk. The risk score may be a numeric score bound between two limits, an unbounded numeric score, an ordinal categorical score (for example: High, Medium, Low), or an unordered categorical variable (for example: PTSD, Divorce, Family Death, etc.). The final decision on risk score methodology will be made after the risk model is finalized and dashboard/notifications end users are consulted.

The outputs from the risk model will be stored in the SQL Server instance that also houses the raw data inputs into the model. Since the exact output of the statistical model is not yet defined, there is no definition as to the exact storage protocol of model outputs, only that they will be stored in the SQL Server. The logic behind storing the model outputs in the SQL Server is to allow the dashboard and messaging applications to pull these data from a single source rather than multiple sources.



Figure 24: Risk Model Data Sources

**Suicide Completion Risk Model Inputs**

It is expected that at the minimum, the suicide completion risk model will utilizeVHA and SDR/NDI data. However, if feasible, this data will be supplemented by additional information from VBA data sources, and potentially third party data sources (e.g., local unemployment rates, country-level general suicide incidence from CDC, DoD data):

VHA Data – The VHA data contains Veteran health information including but not limited to inpatient and outpatient care administered, diagnoses, and prescriptions for Veterans who have used VHA services. The VHA data will be stored in the Reach Database behind the VA firewall. This data will be used both for developing the suicide completion risk model as well as for iterative, near-time development of suicide likelihood scores for Veterans.

SDR/NDI Data – The SDR and/or NDI data is used to identify Veterans who have committed suicide; thus, this data is crucial for constructing the dependent variable in the models. This data will be imported into R and used to train and test the risk model.

VBA Data – The VBA data contains, but is not limited to: benefit, military, financial, pension, and disability information on Veterans who are using VBA services. Similar to the VHA data, the VBA data will be stored in the Reach Database behind the VA firewall. If feasible, this data will be used both for developing the suicide completion risk model as well as for iterative, near-time development of suicide likelihood scores for Veterans.

Third Party Data – There are third party data sources being considered to obtain information on Veterans that is not contained within VHA or VBA data. Some of these sources are county unemployment rates, county suicide rates, LexisNexis, court records, and social media data.



Figure 25: Data Sources

**Suicide Completion Risk Model Outputs**

Predicted Suicide Completion Risk – The developed model in R will produce, for each Veteran in the training and testing cohort, the predicted probability of completing suicide. The predicted values will be computed based on fitted model coefficients. It is not necessary for these predicted probabilities to be imported into the SQL server (see section below on Model Coefficients).

Model Coefficients – The suicide completion risk model coefficients (for all included variables) derived in R will be imported into the SQL server. Using these coefficients, for every record (Veteran) with an update to variable values, the coefficients will be applied to Veteran data to re-calculate the risk of suicide completion. This updating will occur on the SQL server (and outside of R).

**Components**

R – The risk model will be trained and tested using R. When training the model, R will be importing data from tables in the Reach Database. The tables will contain VHA, VBA, SDR/NDI data and third party data.

SQL Server – The data used for the risk model in the Reach Database will be stored in a SQL Server instance. R will pull data from SQL Server in order to train and test the risk model. The suicide completion risk will be calculated/updated directly on the server using Veteran data and model coefficients developed in R.

**Dependencies within Components**

R to SQL Server – In order to be able to develop a risk model and save model coefficients, a connection between R and the SQL Server must exist to allow for the transferring of data between the two components. R has a package called RODBC, which allows for this functionality. Utilizing the RODBC package, R and SQL Server will be able to transfer data between the two components allowing for modeling development and suicide completion risk scoring calculations.

## External System Interface Design

For further details on the System Interface Design, please refer to IRDS Interface Design Specification.

# Human-Machine Interface

## User Roles

Users of the Integrated Reach Database system consist of intervention service providers, other VA mental health providers, and leadership resources. In addition, approved researchers may also be granted access to the system to analyze data and run reports. For a full description of the user classes, refer to the User Research Report.

Table 11: User Roles

|  |  |
| --- | --- |
| User Class | Description |
| VA Outreach and Intervention Team Members | User class includes Suicide Prevention Coordinators (SPC), clinical professionals and other Mental Health staff. Some but not all will have limited experience using IT tools. Experience in Social Work and Clinical Psychology is typical. |
| VA leadership, VA Center of Excellence for Suicide Prevention staff, VA Mental Health leaders | User class includes VA staff in leadership positions with backgrounds in Medicine, Clinical Psychology, Nursing, Public Health, Social Work and various social sciences. Expertise with information technology tools varies broadly depending on background and role at VA. User class may also include Outreach and Intervention Team Members described above. |
| Researchers and Statisticians | User class includes individuals with backgrounds in Statistics, Epidemiology, Public Health, Medicine, Psychology, and various social sciences. Technical expertise for many users will be high, especially in using tools for statistical analysis and related research methods. |
| Reporting Users | User class includes individuals who generate reports related to Veteran suicide. These users may include individuals described in the user classes above plus additional VA managers, analysts, and administrators who are asked to generate reports. Expertise with various technology tools for many users will likely be high, especially in areas related to reporting tools, databases, and statistics. |
| Reporting Consumers | User class includes individuals in the target audience for reports, including internal VA clinical and administrative staff in addition to groups external to VA including other government agencies, US Congress, news media, Veterans Service Organizations, and the general public. Expertise understanding and interpreting statistical data will vary across target audiences. |
| Sustainment Staff | User class includes technical VA staff with backgrounds in Information Technology, Computer Science, Statistics, and related fields with the ability to create new statistical models in the application and map the application to new interfaces. |

## Interface Design Rules

This section of the document is TBD and will be provided as user requirements are finalized.

## Inputs

This section of the document is TBD and will be provided as user requirements are finalized.

## Outputs

This section of the document is TBD and will be provided as user requirements are finalized.

## Navigation Hierarchy

TBD – Provide Screen Shot of Application Navigation

# Security and Privacy

This section will be completed upon identification of the IRDS host environment and the requirements for the host location.

## Security

This section of the document is TBD and will be provided the host environment is decided.

## Privacy

*This section of the document is TBD and will be provided t*