**System Design Document**

Genomic Information System for Integrated Science 2 (Genisis2) Technical Services

Build 7



December 2017

Document Version 3.0

### Department of Veterans Affairs

### Document Revision History

|  |  |  |  |
| --- | --- | --- | --- |
| **Date** | **Revision** | **Description** | **Author** |
| 03/03/2016 | 0.1 | Initial version | Engility |
| 04/07/2016 | 0.2 | Updated draft | Engility |
| 06/16/2016 | 0.3 | Updated draft | Engility |
| 08/16/2016 | 0.4 | Updated version | Booz Allen Hamilton |
| 08/29/2016 | 0.5 | Adding design detail | Booz Allen Hamilton |
| 09/11/2016 | 0.6 | Added design detail | Booz Allen Hamilton |
| 04/14/2017 | 0.7 | Made modifications to design detail | Booz Allen Hamilton |
| 05/12/2017 | 0.8 | Updated Version after STAT Review | Booz Allen Hamilton |
| 06/16/2017 | 0.9 | Updated for Build 3 | Booz Allen Hamilton |
| 09/15/2017 | 1.0 | Updated for Build 4 | Booz Allen Hamilton |
| 11/03/2017 | 2.0 | Updated for Build 5 Terminology Service additions | Booz Allen Hamilton |
| 12/15/2017 | 3.0 | Updated for Build 7 Terminology Service Additions | Booz Allen Hamilton |

### 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 product is built to reflect the as-built product.

### When to Complete Each Section of the SDD

|  |  |  |
| --- | --- | --- |
| **Section** | **Completed on or Before PMAS Phase** | **Rationale** |
| 1 – Introduction | MS 0 Review; updated thereafter | Conceptual design should inform evaluation of investments |
| 2 – Background | MS 0 Review; updated thereafter | Conceptual design should inform evaluation of investments |
| 3 – Conceptual Design | MS 0 Review; updated thereafter | Conceptual design should inform evaluation of investments |
| 4 – System Architecture | MS 0 Review; updated thereafter | Conceptual design should inform evaluation of investments |
| 5 – Data Design | MS 1 Review; updated thereafter | Design details should be elaborated upon during PMAS Planning phase and prior to development |
| 6 – Detailed Design | MS 1 Review; updated thereafter | Design details should be elaborated upon during PMAS Planning phase and prior to development |
| 7 – External System Interface Design | MS 1 Review; updated thereafter | Design details should be elaborated upon during PMAS Planning phase and prior to development |
| 8 – Human Machine Interfaces | MS 1 Review; updated thereafter | Design details should be elaborated upon during PMAS Planning phase and prior to development |
| Attachments | MS 1 Review; updated thereafter | Design details should be elaborated upon during PMAS Planning phase and prior to development |

A product’s system design should be defined conceptually prior to the allocation of personnel and resources that occur at project initiation. This gives the enterprise an opportunity to evaluate IT investments before project teams are stood up and funding is allocated. Sections 1- 4, which discuss the high-level design, should be completed prior to MS 0. All sections should be completed and updated before MS 1. Projects will need to address all SDD approval constraints prior to the MS 2 review. In addition, the SDD should reflect the as-built product going into the MS 2 review.

|  |  |  |
| --- | --- | --- |
| **Activity** | **New Capability (1)** | **Feature Enhancement (2)** |
| Field Deployment (A) | Yes | Yes |
| Cloud/Web Deployment (B) | Yes | Yes |
| Mobile Application (C) | No | No |

### Table of Contents

1. [Introduction 1](#_bookmark0)
   1. [Scope 1](#_bookmark1)
   2. [User Profiles 4](#_bookmark2)
2. [Background 5](#_bookmark3)
   1. [Overview of the System 5](#_bookmark4)
   2. [Overview of the Business Process 5](#_bookmark5)
   3. [Overview of the Significant Requirements 6](#_bookmark6)
3. [Conceptual Design 8](#_bookmark7)
   1. [Conceptual Application Design Principles 8](#_bookmark8)
   2. [Conceptual Application Design 10](#_bookmark9)
   3. [Conceptual Data Design 21](#_bookmark10)
   4. [Conceptual Infrastructure Design 32](#_bookmark11)
4. [System Architecture 39](#_bookmark12)
   1. [Hardware Architecture 40](#_bookmark13)
   2. [Software Architecture 41](#_bookmark14)
   3. [Network Architecture 46](#_bookmark15)
   4. [Service Oriented Architecture / ESS 47](#_bookmark16)
   5. [Enterprise Architecture 47](#_bookmark17)
5. [Data Design 48](#_bookmark18)
   1. [Human Machine Interface 50](#_bookmark19)
   2. [DBMS Files 55](#_bookmark20)
   3. [Non-DBMS Files 56](#_bookmark21)
   4. [Data View 56](#_bookmark22)
6. [Detailed Design 57](#_bookmark23)
   1. [Hardware Detailed Design 57](#_bookmark24)
   2. [Software Detailed Design 58](#_bookmark25)
   3. [Network Detailed Design 66](#_bookmark26)
   4. [Security and Privacy 66](#_bookmark27)
   5. [Service Oriented Architecture / ESS Detailed Design 67](#_bookmark28)
7. [External System Interface Design 67](#_bookmark29)
8. [Attachment A – Approval Signatures 68](#_bookmark30)
   1. [Identification of Technology and Standards 69](#_bookmark31)
   2. [Constraining Policies, Directives and Procedures 69](#_bookmark32)
   3. [Requirements Traceability Matrix 69](#_bookmark33)
   4. [Packaging and Installation 69](#_bookmark34)
   5. [Design Metrics 69](#_bookmark35)

# Introduction

The Genomic Information System for Integrated Science 2 (Genisis2) Technical Services application provides a central infrastructure to support the multiple Department of Veterans Affairs (VA) genomic medicine studies for facilitating recruitment and enrollment of Million Veteran Program (MVP) participants, automating most study-related logistics, including study enrollment, capturing clinical study data, consent, blood-sample tracking, and genomic-data storage. Genisis2 also provides the secure, analytical infrastructure necessary to conduct robust genomic and bioinformatics-related data management and data analysis. Genisis2 includes the planning and analysis of the VA-hosted computing cluster, as well as VA-approved cloud (or other) computing infrastructure, planning and analysis of the Internet 2 (or other WAN connectivity) for genomic data transmission, storage, and planning and analysis of the “big data” tools and systems.

The requested enhancements to the “As Is” functionality is needed to support the anticipated rapid expansion of the MVP.

Genisis2 is the name of the project. Builds are named Build 1, Build 2, Build 3 and so forth.

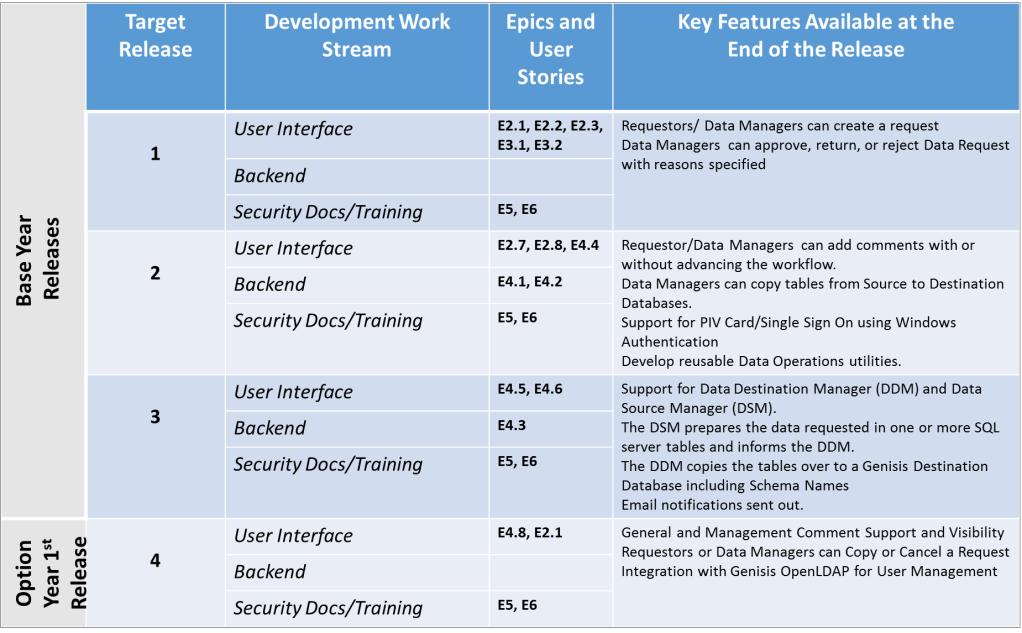
## Scope

The Genisis Requirements Specification Document (RSD) defines the project scope; however, the initial requirements documented in the RSD have been updated in the Genisis2 project Requirements Traceability Matrix (RTM) and will be maintained in the Rational tool.

The following sections outline the scope for Genisis2.

### Overall Scope of Genisis2 to Date

Figure 1 provides an overview of the overall scope of Genisis2 to date. (Builds 1 through 4),



#### Figure 1: Genisis2 Scope to Date

**Build 5 Terminology Service Scope:** A Researcher uses Terminology Service to locate the data elements they need by navigating a set of Concepts from many Ontologies like SNOMED-CT, HP, and NCIT. The goal of Terminology Service is to enable Researchers to start their searches with the clinical and bio-medical language they are familiar with, and arrive at concepts in these Ontologies. The concepts are attached to specific data elements in Genomic, Survey and Clinical data. Terminology Service provides a guided search capability to data elements of interest for research use.

***Build 5 is NOT a production release.*** However, features from Build 5 are implemented in Build 7 and in future Build 8. These three builds will provide the full Terminology Service functionality as planned for Genisis2.

##### Build 7 is a Researcher ONLY release. The only role recognized in Release 7 is the Researcher role. The combination of Builds 5 and 7 will be the third production release, and will be named Genisis2 Release 4 in the VA production environment.

### Release Scope

Table 1 provides the Scope of Genisis2 builds to date.

**Table 1: Scope of Each Release to Date**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Date** | **Revision** | **Description** | **Project Name** | **VA**  **Department** |
| 02/2017 | Build 1 | * Enable Researcher/ Data Manager to create a request for Clinical, Genomic or Survey Data. * Enable Data Managers to approve, return, or reject Data Request with reasons specified. | Genomic Information System for Integrated Science 2 (Genisis2) | VA ORD  (Office of Research and Development) |
| 05/2017 | Build 2 | * Enable Researcher / Data Manager to add comments with or without advancing the workflow. * Enable Data Managers to copy tables from Source to Destination Databases. * Develop reusable Data Operations utilities. | Genisis2 | VA ORD |
| 06/2017 | Build 3 | * Support for two kinds of Data Managers: Genisis Data Manager and VINCI Data Manager:   1. The VINCI Data Manager prepares the data requested in one or more SQL server tables and informs the Genisis Data Manager.   2. The Genisis Data Manager copies the tables over to a Genisis Destination Database. Email notifications are sent. | Genisis2 | VA ORD |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Date** | **Revision** | **Description** | **Project Name** | **VA**  **Department** |
| 09/2017 | Build 4 | * General Workflow Additions:   1. Copy Request Function – Ability to create a Draft Request from another a request that already exists.   2. Cancel a Request.   3. Ability to add Comments that are visible to everyone (General Comments) and only to management (Operational Comments). * System Administrator Functions:   1. Ability to add users to the system by synchronizing with the Genisis2 OpenLDAP directory and searching for a specific user.   2. System Administrator Dashboard functions: Summary of Number of Users by Role, Request Metrics – Raw Numbers of Requests in Draft, Approved, Fulfilled, etc., and drilling down to list of requests of any one category, down to the details of any one request. | Genisis2 | VA ORD |
| 11/2017 | Build 5 | * Microservice APIs – Support the TS Programmer to define 4 microservice APIs to support the Search, Concept, Bookmark and Mapping services. * Cohort Definition Functions – Support Researcher to (1) search for a concept and view the search results, (2) Concept Card, (3) bookmarked and un-bookmark concepts, and (4) label and un-label bookmarked concepts. * Terminology Creating - Support Terminologist with Create, Modify and Delete functionalities through CLI. | Genisis2 | VA ORD |
| 11/2017 | Build 7 | * Cohort Definition Functions: Support Researcher with useful Search Results and a simplified Concept Card for a Data Element to define cohort. * Multiple Roles Connection Functions: Support Researcher with Single Sign-On (SSO), iFrame, Data Elements Integration, and Baseline Survey Data Element List Display. | Genisis2 | VA ORD |

## 1.2. User Profiles

The primary intended user population will be Researchers who are experienced in performing analysis on computational clusters.

The Genisis2 application is designed for its current users to provide, collect, manage, and analyze genomic data for research operations. Table 2 identifies the users and their responsibilities.

##### Please note that in Build 7 the only role supported below is a Researcher Role. This is because Build 7 is for Researchers’ use only and not for Terminologists or System Administrators. In addition, since Data Request workflows are not supported in Build 7 (Only Terminology Services), the Genisis Manager and the VINCI Data Manager roles are not supported.

**Table 2: User Profiles**

|  |  |  |
| --- | --- | --- |
| **Name of Users** | **Description** | **Responsibilities** |
| Genisis2 System Administrator | Performs system administration functions for the Genisis2 application | * Assumes system administrator privileges to maintain Genisis2 application and network interfaces * Monitors and tracks user access privileges for all users * Imports relevant data from non-VA /third- party sources * Performs application updates * Inserts content changes, including documentation * Deletes obsolete content * Configures/reconfigures the computing environment * Provides on-demand updates and content changes * Manages job queue to create and configure environments * Ensures that Genisis2 is maintained as an enterprise-level operation servicing the needs of MVP |
| Genisis2 Manager | Approves data request, new users, and provisions data | * Performs system updates. * Ensures Genisis2 complies with regulatory and security changes. |
| VINCI Manager | Prepares the data being asked for by the Researcher and delivers the data | * Addresses questions and clarifications about the data being requested. * Prepares the data in one or more SQL server tables. * Informs the Genisis2 Manager that the data is ready for use. |

|  |  |  |
| --- | --- | --- |
| **Name of Users** | **Description** | **Responsibilities** |
| Researcher | Request data; performs analysis | * Uses Genisis2 Terminology Service to identify data elements they need using Ontologies/Concepts. * Creates, modifies, and cancels data requests. * Performs research on Genisis2 clinical, genomic, and survey data. |

# Background

## Overview of the System

A Genisis2 Researcher uses other systems to query on data collected from various VA locations and public data sources to help form evidence for research proposals. They then formulate relevant data evidence and use it to submit a proposal to the Genisis2 governance process and any applicable Internal Review Board (IRB). Genisis2 is a workflow system used by the Researcher after obtaining this approval(s) to create and submit a request for detailed clinical, genomic and survey data. Genisis2 also provides a central repository for integrating clinical and genomic data collected from various VA locations and public data sources perform data cleaning operations on this data and use it for research.

A detailed set of requirements are specified in the Requirements Traceability Matrix (RTM) in the form of Epic stories that constitute the desired functionality for Genisis2. These requirements are decomposed to actionable items with acceptance criteria to produce the target solution design. The design specifications outlined in this document are based on the approved requirements and acceptance criteria and will serve as the guiding framework for the design and development for the Genisis2 system. Design elements in this document will also be mapped to requirements within the RTM to demonstrate how each requirement will be implemented.

The major design elements included in this document range from the automation of the Researcher request for clinical data, approval of the request, system usage and monitoring by system administrators and regulatory and security compliance associated with these operations.

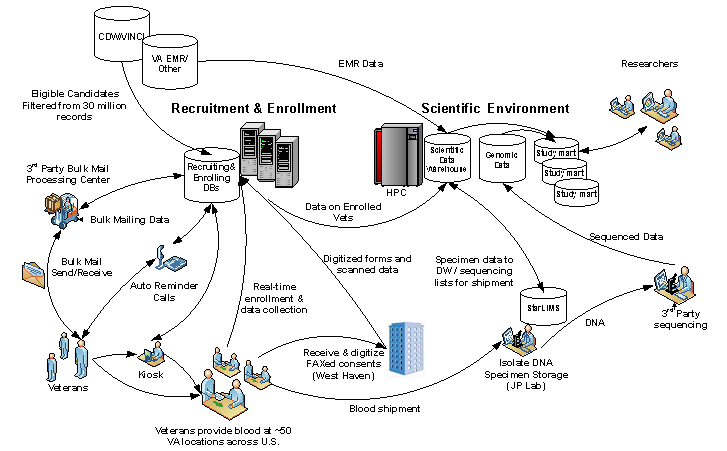
## Overview of the Business Process

Figure 2 provides a high-level view of the Genisis2 business processes. As depicted in the flow, Genisis2 is a recipient of external data from multiple sources, including clinical data from CDW/VINCI, consent forms and personal surveys from MVP sites, sequenced genotype data from MAVERIC, and other external sources, such as National Death Index (NDI) and Centers for Medicare and Medicaid (CMS).

The Honest Broker extracts Personally Identifiable Information (PII) from the data, and then stores the de-identified data in a secure database schema within the Genisis2 data warehouse.

De-identified data is then integrated with MVP genotype data, survey data, and consent data into a schema, and moved into study marts for Researchers. This study mart also provides

Researchers a virtual workspace to perform analysis and run computations, and are only given access to the data that has been approved for their studies.



**Figure 2: High-Level Genisis2 Business Process Flow**

## Overview of the Significant Requirements

The requirements identified in this section are not to replace the existing RSD, nor serve as the basis for the RTM. This section offers a high-level perspective to individuals who are not familiar with the requirements; and therefore, may need more background information in order to better understand the basis for the design.

These requirements are explained in more detail in the Genisis2 RSD, which will be uploaded to TSPR once finalized, with preliminary requirements in the Genisis2 BRD and Genisis2 RTM. Please note that the updated RTM will be developed in Rational Tools and added to the TSPR site upon completion.

The Genisis2 project will execute the requirements described in the following sections.

### Unified Genisis2 System Administrator/Researcher Dashboard

The Genisis2 requirements propose enhancements to the existing dashboard to provide a unified dashboard for Researchers and Genisis2 System Administrators. The enhanced dashboard will

help Researchers request study data and track the status of their study data request throughout its lifecycle, and allow System Administrators to monitor the Genisis2 application.

#### Genisis2 System Administrator Improved Functionality and Dashboard

Enhancements for Genisis2 System Administrators are addressed in the Genisis2 requirements to support the rapid ramp-up of demand for system resources, user access, and setup of new study marts. A new Genisis2 System Administrator dashboard complements the improved functionality and supports the following features:

* + - * + Supporting Personal Identification Verification (PIV) authentication for all users to access the Genisis2 application. (This might be achieved through Windows Authentication through the Browser when someone has used a PIV card to gain access to the VA environment.)
        + Providing role-based access for each Researcher account.

#### Genisis2 Manager Dashboard

Genisis2 Managers (Genisis2 Managers) will have their own set of permissions that are lesser than those of Genisis2 System Administrators. They will not be able to add users or modify some Genisis2 application parameters such as workflow definitions and approval chains. They will have all the permissions needed for reviewing and approving Researcher data requests, performing data management activities, copying data, and performing quality control on them.

#### VINCI Manager Dashboard

VINCI Managers will have their own set of permissions that are not as extensive as Genisis2 Managers. They will not be able to add users or modify some Genisis2 application parameters like workflow definitions, approval chains, etc. They will have the permissions needed for reviewing and approving data requests coming from Genisis2 Managers, preparing the data asked for, and providing them in SQL server tables.

#### Researcher Dashboard

Researchers will have a new dashboard to submit requests for clinical, genomic or survey data, track and submit change requests, track status updates about their study, access VINCI tools, and receive notifications when there are updates to their requests.

### 2.3.2. Security Enhancements

Security enhancements are required to meet Federal mandates, for example, the addition of PIV for controlling access to VA systems by all VA users. This is achieved through Windows Authentication through the Browser when someone has used a PIV card to get into the VA environment in the first place.

In addition, new functionality is added to improve provisioning of ACLs to manage Researcher access to study marts. The new security enhancements will meet the two-factor authentication required for Authority to Operate (ATO) certification.

Genisis2 security will be reviewed for approval to upgrade from its current Temporary Authority to Operate (TATO) to the ATO certification through the Assessment and Authorization (A&A)

process. A&A approval is a prerequisite for ATO certification. Genisis2 will undergo a series of scans and assessment tests to validate security of the software code, the database, and the firewall. The discovery of any “Critical” or “High” risks will be mitigated, and subsequent scans will validate that the risks have been remediated. In addition, Genisis2 will undergo a level of penetration testing as needed by the requirements for an ATO. For Genisis2, Service Level Agreement Modification (SLAM) #16-051 is now in place to engage the support of Enterprise Operations (EO). EO will facilitate the ATO process and collaborate with Genisis2 management to ensure that Genisis2 achieves its ATO certification.

# Conceptual Design

This section of the SDD provides details about the following topics:

* Conceptual Application Design Principles
* Conceptual Application Design
* Conceptual Data Design
* Conceptual Infrastructure Design

## Conceptual Application Design Principles

These are the general principles used in the Conceptual Application Design:

#### Genisis2 Application Design needs to be Extensible and Flexible

* + ***Flexibility* –** Genisis2 envisions a single workflow process handling Researchers’ requests and fulfilling them. As research efforts scale in number and variety of what they study, there may be many variations of these processes needed depending upon who the Researchers are and the purposes of their studies. Modular design that separates the workflow functions from the data handling functions will go a long way in achieving this needed flexibility.
  + ***Extensibility* –** Genisis2 Application design needs to be extensible so that it can work with other new applications, other new sources of data, other infrastructure back- ends, and capacity that may be used in the future for the MVP project. Modular design of the Genisis2 applications and loosely coupled components will help achieve this extensibility in the future. Genisis2 will support a RESTful Application Programming Interface (API). The Genisis2 User Interface will use this API for its implementation. Other VA systems that want to talk directly to Genisis2 can use the same approach. Documentation is provided and updated with each release of Genisis2.

#### Separate “Genisis2 Request Workflow” functions from Data Operations Functions

* + ***Genisis2 is a Workflow Application.*** Workflow steps involve:
    - The creation of requests by Researchers.
    - Modification or cancellation of these requests.
    - Verification and validation of these requests by Genisis2 Managers *(Data Destinations).*
    - Receipt of data from VINCI/CDW or other *Data Sources.*
    - Data Quality Control of this data.
    - Copying the data to the Researcher’s Study Marts
    - Informing the Researchers that their requests have been reviewed, approved/denied, or fulfilled (i.e., their data is ready for analysis.)

In that sense, the Genisis2 application needs capture the above steps in a workflow and is a *General Workflow Enabler*.

* + ***Data Operations Management Enabler –*** Genisis2 Managers perform ETL Functions and data transformation functions for quality***.*** The Genesis2 application will capture metadata about these operations but leave the actual cleansing of data to the Genisis2 Managers.
  + ***Keep the Workflow and Data Operations Business Logic Separate*** – The application design will keep the request handling workflow business logic separate from the data operations business logic. As research efforts scale, this is critical if we want to keep the application flexible and extensible. This will help us accommodate new kinds of data operations in the future, especially those we do not foresee today.
  + ***Associate and archive Data Management Logic with Study Mart Requests*** – The application design must let Genisis2 Managers do their work manually for a Researcher query or study request. However, the design must have a way to capture all the steps involved in that request along with Meta data about what the inputs of that data operation were, inputs, outputs, etc. The application should not care about what is used in ETL or what the query is, but allow archival of those utilities with descriptions of what they do, how they do it, and accommodate a library of such tools. For example, cleansing the Date of Birth field may involve a query/update script while removing PII using a script that talks to the Honest Broker system may be another utility. Also, allow the documenting of the sequence of data operations use. This will form a secure audit trail for analysis if need be later on.

#### Provide for a Semi-Automated Workflow with Manual Checkpoints

* + ***Integrated Workflow/Document Management tool.*** The application design will consider an ***Integrated Workflow tool*** to implement Genisis2 so that implementing the first type of workflow along with possible future ones are easy.
  + ***Manual Workflow with Automated Steps.*** As of today, not the entire request handling workflow steps or data operations can be fully automated end-to-end. The Genisis2 Manager executes manual checkpoints to say, checking on study approvals before approving a data request. The approval and the workflow steps will be automated with data operation scripts created and libraries of those archived over time. In the future, these operations can be automated so that the workflow proceeds smoothly and stops only when a manual checkpoint is absolutely needed.
  + ***Tracking and Reporting of Workflow Progress for Transparency and Process Improvement*** – The application design will enable the tracking and reporting of the progress of a request since it is captured in a workflow at any time. This way, Researchers can have transparency in knowing exactly where their request for a query mart or a study mart is at any time. You can compare expected time intervals between

steps with actual time taken for reporting and analysis and use it to improve processes.

## Conceptual Application Design

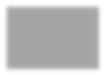
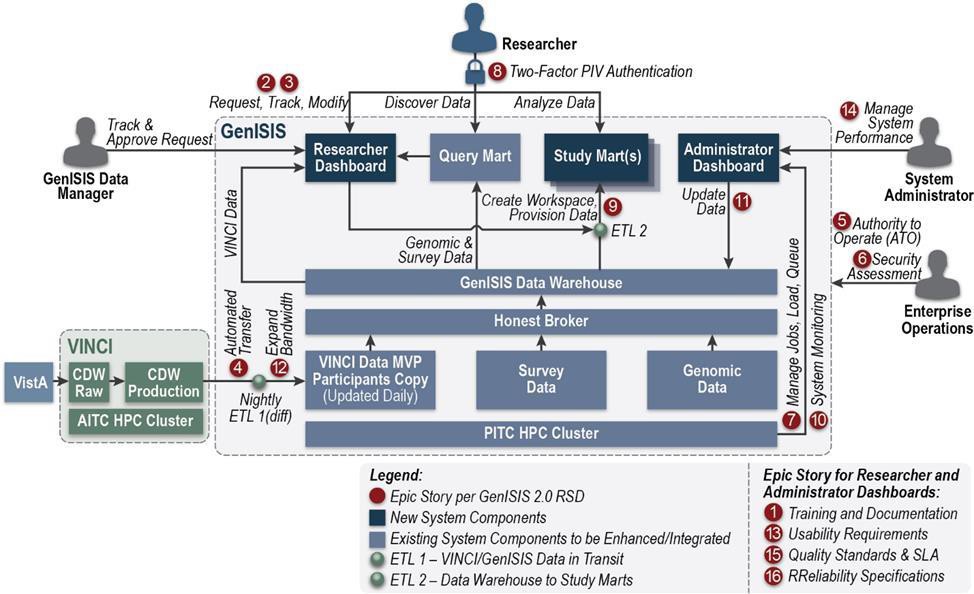
### Genisis2 Data Request Workflow

Figure 3 captures the overall flow of requests for Research Data and how the data itself gets prepared and sent to Researchers/Researchers.

Salient points in Figure 3 are as follows:

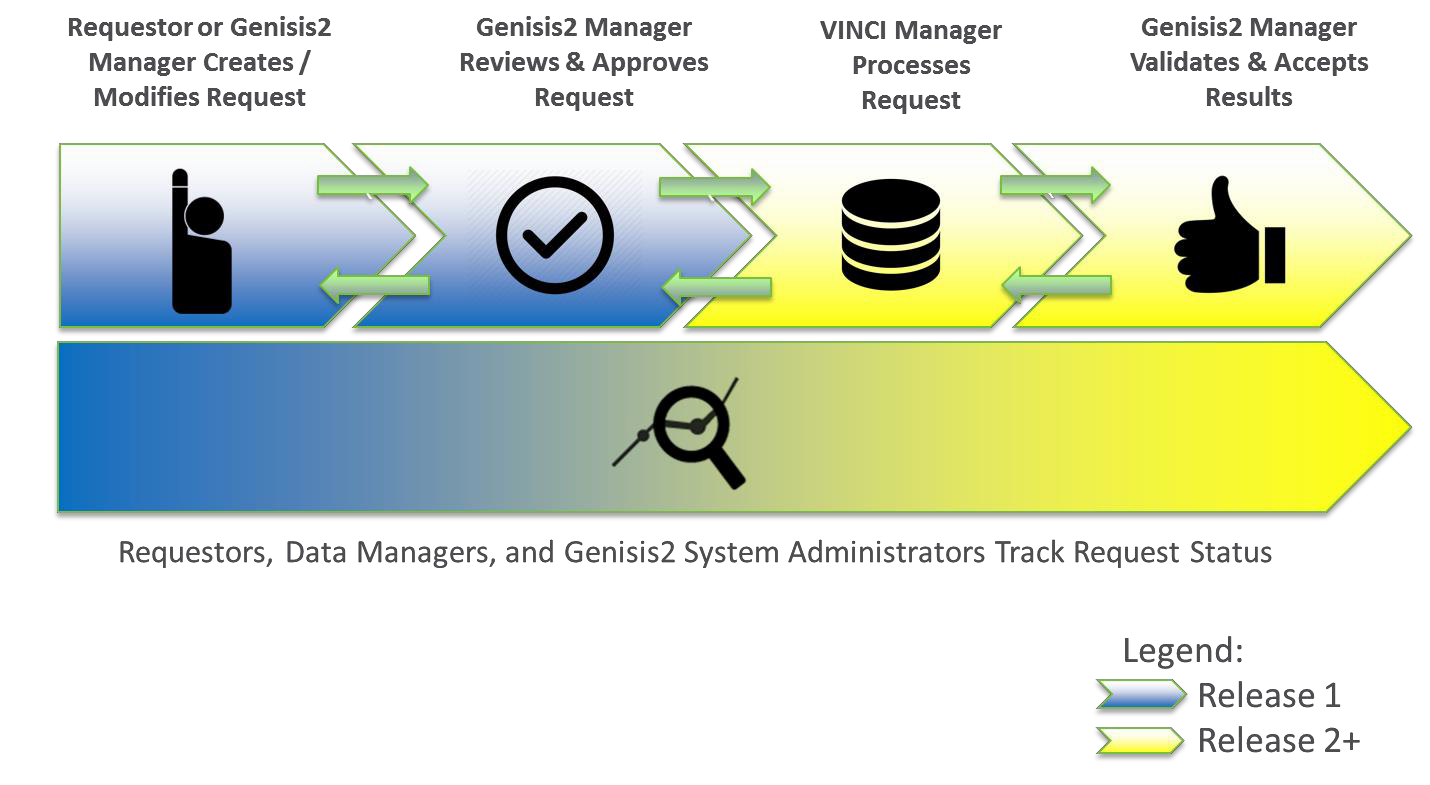
* Researcher logs in using VA two-factor authentication and creates a request for data.
* A Genisis2 Manager coordinates the questions and clarifications with the Researcher, mediates data asked for and data available kinds of issues and finally submits a request for Genomic, Clinical, and Survey data to be prepared and copied over to the Study Mart.
* Genomic, Clinical and Survey data pass through an Honest Broker system that removes all Personally Identifiable Information (PII) and maintains a Study Mart Id to Patient ID mapping for future refreshes of data.
* A Genisis2 System Administrator can login with enhanced privileges, create and register new users, modify their roles, disable them if necessary and be a super user that can perform the role of all other users when needed.

***Smart orchestration and automation of workflows and ETL processes will allow GenISIS 2.0 to preserve data integrity and scale to meet escalating demands of a national resource***



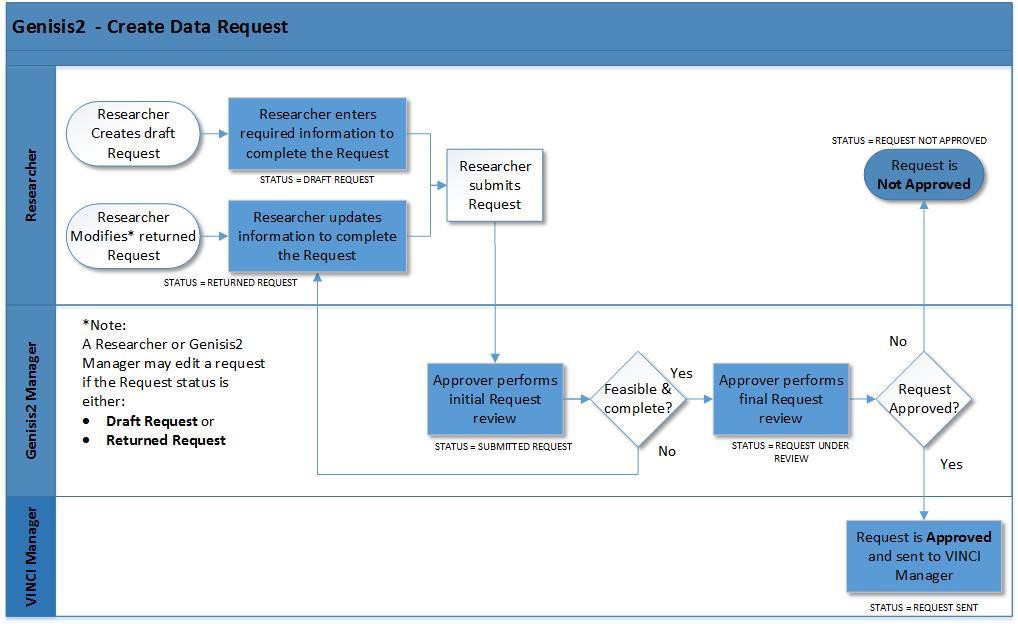
**Figure 3: Genisis Data Requests - Overall Structure of the System and Epics**

A high-level flow of the system is represented in Figure 4, showing the pathways a request for data can take.



**Figure 4: High-Level Data Request Flow Process**

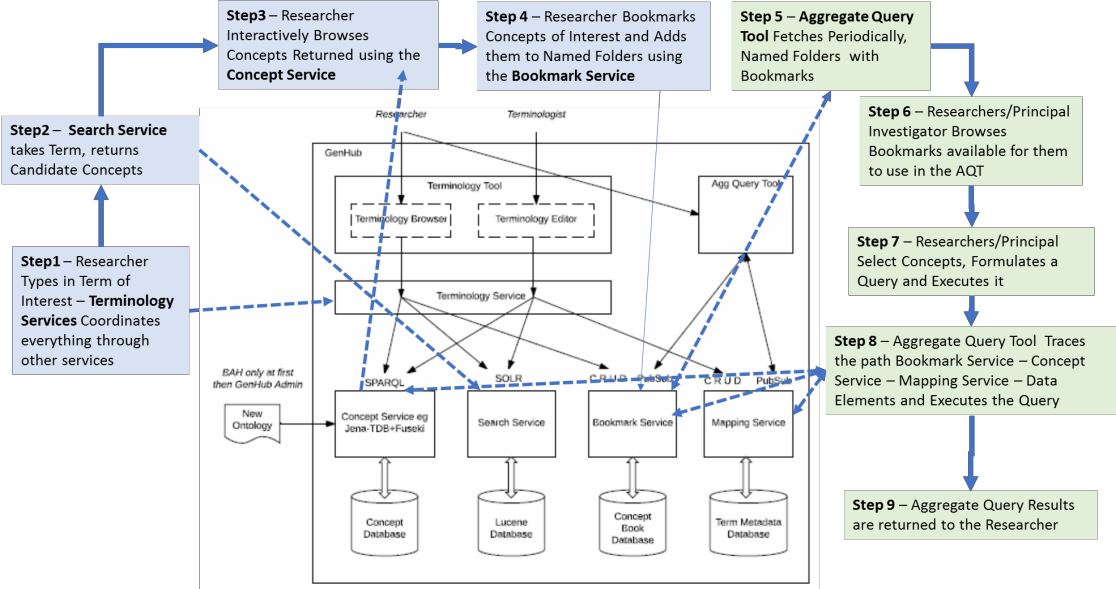
The key workflow that is central to the implementation of the high-level flow in Figure 4 is represented in the processes illustrated by Figure 5.



**Figure 5: Data Request Flow**

### Genisis2 Terminology Service Process Flow

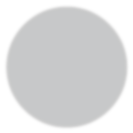
A high-level process flow for the Genisis2 Terminology Service is as shown in Figure 6. It represents the workflow for a Researcher to start searching for Data elements in Ontologies, in terms that they are familiar with, ending up with data they need. Once they have put together the Data Elements that makes sense to them they go forward with executing queries that involved them.



**Figure 6. Genisis2 Terminology Service Process Flow**

### Application Context

Figure 7 depicts the Application Context Diagram (ACD) for the Genisis2 application.



National DNA Lab

MVP Coordination Sites

System Administrator Dashboard

Other VA Databases (Brain Bank)

Scan Form Data Provider

MAVERIC

Non VA Data Sources

Active Directory

CDW

Public Datasets

VistA

VINCI

GeneSIS v2.0

**Figure 7: High-Level Application Context Diagram**

Table 3 describes the information in the Application Context Diagram in four sections. Note that a single object (in the center of Figure 7) represents the system for which this design applies.

Therefore, it is not referred to in Table 3.

##### Please note that this section outlines all the Data Sources that Genisis2 may interface with in the future, not necessarily within the scope of this project now.

**Table 3: (Grouping) Application Context Description**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **ID** | **Name** | **Description** | **Interface Name** | **Interface System** |
| Scan Form Data Provider | Scan Form Data Provider | Scan form data providers process form data from study participants and present forms in tabular form. VA Westhaven Research Lab and Synovate, Inc. are two current scan form data providers. | Form | Genisis2 |
| CDW | VA Corporate Data Warehouse | The VA Corporate Data Warehouse presents data through relational database views. | CDW | Genisis2 |
| VistA | Veterans Health Information Systems and Technology Architecture (VistA) electronic health records system | VistA is the VHA electronic health record system. | VistA | Genisis2 |
| Public Datasets | Public Data sets | Genomic annotations and ontologies are publicly available and necessary for Genomics research. | Public Datasets | Genisis2 |
| Active Directory | VA Active Directory | The VA Active Directory provides authentication services to Genisis2. | VaAd | Genisis2 |
| Brain Bank | Brain Bank | The Brain Bank is a bio specimen repository maintained by the VA. | BrainBank | Genisis2 |
| National DNA Lab | VA National DNA  Laboratory | The VA National DNA Laboratory is the repository for bio specimens and DNA collected from the VA Genomic Medicine Program. | DnaLab | Genisis2 |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **ID** | **Name** | **Description** | **Interface Name** | **Interface System** |
| VINCI | Veterans Informatics and Computing Infrastructure | VINCI is a partner with the Corporate Data Warehouse (CDW) and hosts all data available through CDW as well as some unique data. | VINCI | Genisis2. |
| MAVERIC | The Massachusetts Veterans Epidemiology Research and Information Center | MAVERIC is an interdisciplinary research and development organization with the goal of creating a learning healthcare system within VA through application of research resources and methodologies to important clinical problems. | MAVERIC | Genisis2 |
| Non-VA Data Sources | Non-VA Data Sources | National Death Index (NDI), Centers for Medicare and Medicaid Services (CMS) and other national data. | Non-VA Data Sources | Genisis2 |
| MVP  Coordination Sites | Million Veteran Program Coordination Sites | MVP Surveys and Consent Forms. | MVP | Genisis2 |
| Genisis2 System Administrator Dashboard | Genisis2 System Administrator Dashboard | Dashboard allows Researchers to request study data and track the status of their study throughout the lifecycle, and allow system administrators to monitor the Genisis2 application. | Genisis2 System Administrator Dashboard | Genisis2 |

**Interfaces External to OI&T**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **ID** | **Name** | **Related Object** | **Input Messages** | **Output Messages** | **External Party** |
| N/A | Public Datasets | Public Data Sets | Data records encoded in structured text. The data is publicly available reference data, which is cached in Genisis2 for use in genomic analysis. | N/A | National Institutes of Health, Universities, Government- sponsored research institutions. |

**Interfaces Internal to OI&T**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **ID** | **Name** | **Related Object** | **Input Messages** | **Output Messages** | **External Party** |
| N/A | Form | Scan Form Data Provider | Scan form data is presented as a simple database table with one row per subject. | N/A | VA  Westhaven Research Lab |
| N/A | CDW | CDW | CDW subset data is presented as database views from the Austin Information Data Center. | N/A | AITC |
| N/A | ActiveDirectory | Active Directory | LDAP query responses. This is part of VA Windows Authentication System. | LDAP queries to lookup users to the Genisis2 application.  This is part of VA Windows Authentication System | VA Active Directory Infrastructure |
| N/A | BrainBank | Brain Bank | The Brain Bank inventory records is presented as a set of relational database views. | SQL queries to look up bio specimens stored within the Brain Bank | VA Brain Bank |
| N/A | DNALab | VA National DNA Lab | The National DNA Lab presents bio specimen data as a set of relational database tables. | SQL queries to look up bio specimens stored within the National DNA Lab | VA National DNA Lab |
| N/A | VINCI | Veterans Informatics and Computing Infrastructure | VINCI data sets are CDW extractions from VistA, MedSAS in SAS and SQL, and DSS in SAS and SQL. | SQL queries to look up VINCI data sets | AITC |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **ID** | **Name** | **Related Object** | **Input Messages** | **Output Messages** | **External Party** |
| N/A | MAVERIC | The Massachusetts Veterans Epidemiology Research and Information Center | Genotype data. |  | VA Boston Healthcare System |
| N/A | Non-VA Data Sources | Non-VA Data Sources | National Death Index (NDI), Centers for Medicare and Medicaid Services (CMS) and other national data. |  | NDI, CMS  and other |
| N/A | MVP  Coordination Sites | MVP  Coordination Sites | MVP Surveys and Consent Forms. | N/A | Nationwide |
| N/A | Genisis2 System Administrator Dashboard | Genisis2 System Administrator Dashboard |  |  | Virtual |

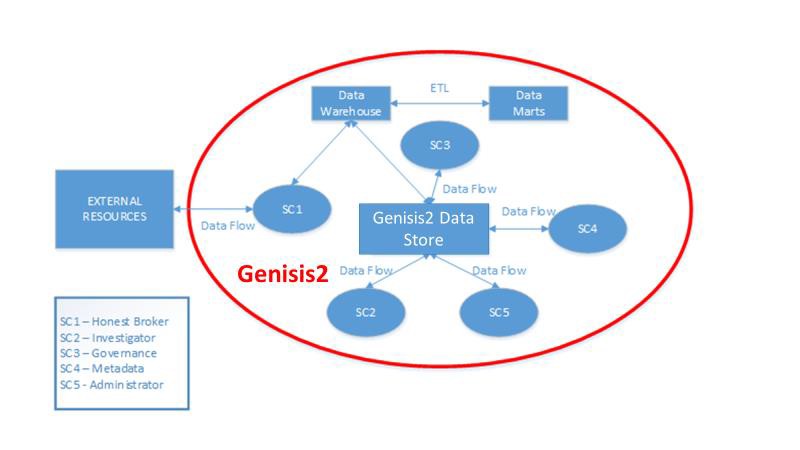
**Externally Shared Data Stores**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **ID** | **Name** | **Data Stored** | **Owner** | **Access** |
| N/A | N/A | N/A | N/A | N/A |

### High-Level Application Design

The High-level Application Design depicted in the Figure 8 identifies the major components of the application and the relationships of the major application components to each other and to surrounding applications.

##### Please note that the following descriptions are providing only the CONTEXT for the Genisis2 application. The other applications such as Honest Broker, Investigator, Governance, Metadata and Administrator are all under the scope of other development projects, many of them in development as of now. That is the reason for Not Applicable (NA) or to be Determined (TBD) entries in the table, since their exact status as of this update is not clearly known at the time of this release.



**Figure 8: High-Level Application Design**

Tables 4 and 5 provide information regarding these components.

**Table 4: Objects in the High-Level Application Design**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **ID** | **Name** | **Description** | **Servi ce or Lega cy Code** | **External Interface Name** | **External Interface ID** | **Internal Interface Name** | **Internal Interface ID** | **SDP**  **Sections 1&2** |
| N/A | External Resources | External Data | N/A | Legislative Information Management System (LIMS)  Backend, West Haven, Synovate, CDW, VistA, VINCI, MAVERIC,  MVP Sites, Public Datasets | TBD | N/A | N/A | N/A |
| N/A | Data Warehouse | Survey & bio- specimen data | N/A | N/A | N/A | N/A | N/A | TBD |

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **ID** | **Name** | **Description** | **Servi ce or Lega cy Code** | **External Interface Name** | **External Interface ID** | **Internal Interface Name** | **Internal Interface ID** | **SDP**  **Sections 1&2** |
| N/A | Data Mart(s) | Study specific survey & bio- specimen data | N/A | N/A | N/A | N/A | N/A | TBD |
| SC 1 | Honest Broker | The Honest Broker receives external data and uploads it to the data warehouse | N/A | N/A | N/A | Honest Broker | SC1 | TBD |
| SC 2 | Investigator | Interface to create and manage Genisis studies | N/A | N/A | N/A | Investigat or | SC2 | TBD |
| SC 3 | Governance | Maintain Study States | N/A | N/A | N/A | Governa nce | SC3 | TBD |
| SC 4 | Metadata |  | N/A | N/A | N/A | Metadata | SC4 | TBD |
| SC 5 | Administrat or | Provides System Administrator functionality | N/A | N/A | N/A | Administr ator | SC5 | TBD |

**Table 5: Internal Data Stores**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **ID** | **Name** | **Data Stored** | **Steward** | **Access** |
| Genisis Data Store | Genisis | Stores all data to support all Genisis2 application components. | Genisis Web Applications | Create, Read, Update, and Delete |

### Application Locations

Table 6 specifies the locations where the Genisis2 application components are hosted and Table 7 provides the users.

**Table 6: Application Locations**

|  |  |  |  |
| --- | --- | --- | --- |
| **Application Component** | **Description** | **Location at Which Component is Run** | **Type** |
| LIMS Backend | MVP bio sample, genomic data | DNA Repository Lab at MAVERIC, Boston, MA | Business Logic / Data Logic / Interface Code |
| Consent Form | MVP Consent forms | VA West Haven, CT | Business Logic / Data Logic / Interface Code |
| Synovate | MVP Survey Form Data | VA West Haven, CT | Business Logic / Data Logic / Interface Code |

|  |  |  |  |
| --- | --- | --- | --- |
| **Application Component** | **Description** | **Location at Which Component is Run** | **Type** |
| CDW | Clinical data | VA Austin, TX (AITC) | Business Logic / Data Logic / Interface Code |
| VistA | TBD | VA Austin, TX (AITC) | Business Logic / Data Logic / Interface Code |
| Data Warehouse | Survey & bio- specimen data | VA Pittsburgh Information Technology Center | Data Logic |
| Data Mart(s) | Study specific survey & bio- specimen data | VA Pittsburgh Information Technology Center | Data Logic |
| Honest Broker | The Honest Broker receives external data and uploads it to the data warehouse | VA Pittsburgh Information Technology Center | Interface Code |
| Investigator | Interface to create and manage Genisis studies | VA Pittsburgh Information Technology Center | Presentation Logic / Business Logic |
| Governance | Maintain Study States | VA Pittsburgh Information Technology Center | Business Logic |
| Metadata | N/A | VA Pittsburgh Information Technology Center | Interface Code |
| Administrator Dashboard | Provides System Administrator functionality | VA Pittsburgh Information Technology Center | Presentation Logic / Business Logic |

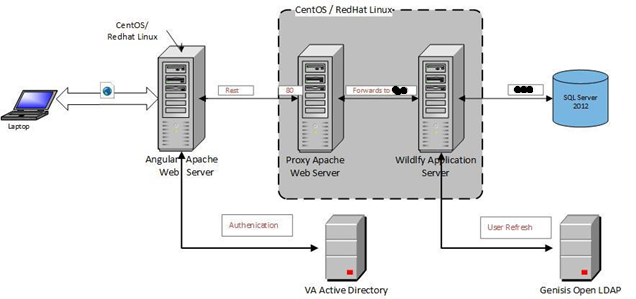
**Table 7: Application Users**

|  |  |  |
| --- | --- | --- |
| **Application Component** | **Location** | **User** |
| Recruitment & Enrollment applications | 50+ locations across the country. All major VA hospitals | Site Coordinators, Lab techs, MVP staff, report writers |
| Scientific Environment | Across the US | Scientists, MVP staff, Researchers |

### Preliminary Architecture

#### Architecture for Genisis2 Data Workflows

Figure 9 presents the architecture for handling Genisis2 Data Requests Workflows and the other external components with which the application interacts.



**Figure 9: Server Architecture for Genisis2 Data Request Workflows**

The Genisis2 architecture consists of the following components:

* + - * + **Genisis2 Core Application –** The Genisis2 core application will be implemented using an integrated Workflow tool such as Activiti. This core application will run in the Web/ Application Server and will consist of:

**Genisis2 Workflow Process Manager –** This will implement the functions needed by the different types of users of this system – Researcher, Genisis2 Manager, VINCI Manager and Genisis2 System Administrator. Each user will have their own dashboard and access to functions that they are authorized to perform. The dashboard will be their login page showing status of their requests or workflow steps, tasks they need to perform and Notifications from various activities. The workflow tool selected will have the appropriate workflow design and business rules implementation capabilities within it. Genisis2 will use these capabilities to design and implement the initial workflow and business rules in force today.

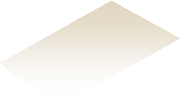
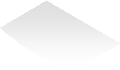
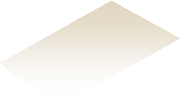
**Notifications Management: –** Requirements that need email or mobile/text message notifications will be handled by existing VA communications management capabilities. Genisis2 will integrate programmatically with these communication mechanisms to create and send out these notifications or text messages.

* **Data Operations and Tools –** As outlined in the Conceptual Application Design Principles section, Genisis2 Managers would be performing their data cleanup, approval and movement operations using scripts from command line interfaces. ***Genisis2 will allow them to be started from the application as needed.***

#### Terminology Service Workflows

Figure 10 presents the architecture for handling Terminology Service workflow and the other external components with which the application interacts.

Centos/RedHat Linux



**Port**

**Port 80**

**Web Server/ Application Server Running**

* **Apache HTTP**
* **Tomcat**
* **Jena/Fuseki**
* **SOLR**

**Database Server**

* **Windows Server 2008 R2**
* **Microsoft SQL Server 2012**

**Figure 10: Terminology Service Server Architecture**

The Terminology Service architecture consists of the following components:

* + - * + **Terminology Service Application –** The Terminology Service core application is implemented using SOLR search services and Jena/Fuseki Triple Store and Query capabilities.

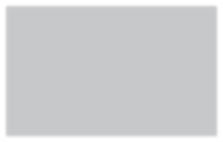
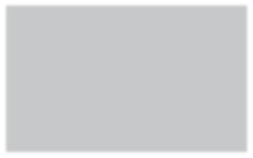
**Notifications Management: –** Requirements that need email or mobile/text message notifications will be handled by existing VA communications management capabilities. Genisis2 will integrate programmatically with these communication mechanisms to create and send out these notifications or text messages.

* **Database Server** – All database storage and querying needs will be implemented using a Microsoft SQL Server 2012 software component running on a server which has the Microsoft Windows Server 2008 R2 software as the Operating System.

## Conceptual Data Design

### Project Conceptual Data Model

The conceptual data model for the Genisis2 application is based on the central entity called Request, as shown in Figure 11. Each Request has attributes for the person who created the request (Researcher), the type of request (either aggregate data or detail data), when the request was created, when it was fulfilled, etc. Each request can also have pointers to data tables or flat files within the data file repository (e.g., either the Genisis landing Zone or a Researcher Study Mart) that satisfied the request. Importantly, the status history of the request is tracked as the request progresses through its workflow. Each transition, (e.g., from submitted by Researcher to approved by Genisis2 Manager) is recorded (along with the person who made the transition), when the transition occurred, and any additional comments or notations. The request may have comments or notations added without a status change; the person that added the notation and the date the notation was added would also be recorded.



Data File

Request

Request Type Requestor Current Status Date Submitted Date Fulfilled

Associated Requests Approval Information Comments

Data File

Annotations

Researcher

Name Date Notation

Name

Contact Information Affiliation

Workflow History

Status Date Name

Comments

Data File Repository

**Figure 11: Project Conceptual Data Model**

### Database Information

Table 8 lists the existing database inventory.

**Table 8: Database Inventory**

|  |  |  |  |
| --- | --- | --- | --- |
| **Database Name** | **Description** | **Type** | **Steward** |
| **Genisis\_Activiti\_DB** | All workflow related data is handled in tables created by the Activiti workflow system. **Contains workflow data only.** | Relational Database | ORD |
| **Genisis\_DB** | The **Main Genisis Database** that handles all data not handled by the Workflow System above (like Comments table). | Relational Database | ORD |
| **Jena/Fuseki** | The Main Terminology Service Triple Store database that stores all Ontologies / Concepts that have been loaded to date. | Triple Store | ORD |

|  |  |  |  |
| --- | --- | --- | --- |
| **Database Name** | **Description** | **Type** | **Steward** |
| **TS\_DB** | The Terminology Service Database that deals with User Specific activities such as Bookmarks, Labels, etc. | Relational Database | ORD |

### User Interface Data Mapping

Table 9 provides the Database Mapping.

**Table 9: Database Mapping**

|  |  |  |  |
| --- | --- | --- | --- |
| **User Interface Type** | **Description** | **Database Mapping** | **Additional Comments** |
| **Researcher/Genisis2 Manager Landing Page/ Request Screens** | Researcher or Genisis2 Manager looking at a status of all their requests, drilling down into one request or creating a new request for data. | **Genisis\_DB/ Genisis\_Activiti\_DB** | All workflow related data is handled by the **Genisis\_Activiti\_DB** automatically while all other data go into tables in the **Genisis\_DB** |
| **Data Table Copy Statuses** | When Research Data is ready and is being copied from one Data Source to the Genisis Databases, statuses of the copy are logged. | **Management\_Table** in the **Genisis\_DB** Database | **Number of Rows** being copied and a **Checksum** calculated both **BEFORE** the copy and **AFTER** the copy and stored in this log. If these do not match, the table is dropped with a **Table not copied** error message and also logged in this table. |
| **Terminology Service User Specific (Researchers, Terminologists) data such as Bookmarks, Labels, etc.** | When users bookmark concepts for their use and create labels to classify them. | **TS\_DB** is the database that handles storage and retrieval of this data | **Bookmarks, Labels** |

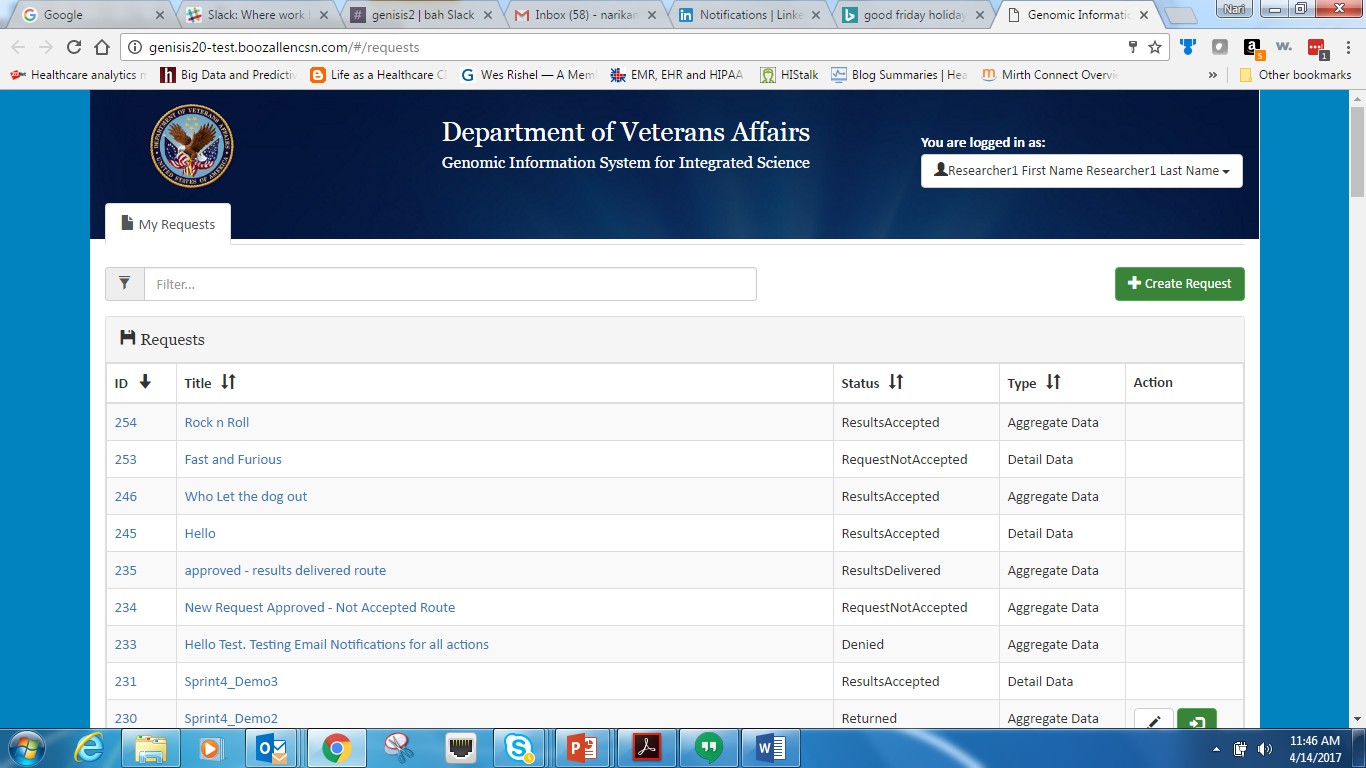
### Application Screen Interface

The following sections provide high-level Application Screen designs as of Build 5. Where features are not yet implemented, mockups at the level of detail currently available are provided.

#### Genisis Data Request Workflows

#### Researcher Landing Page/Dashboard

Figure 12 provides a screenshot of the Researcher Login/Dashboard. From this view, a Researcher can track the progress of their requests.



**Figure 12: Researcher Landing Page/Dashboard Design**

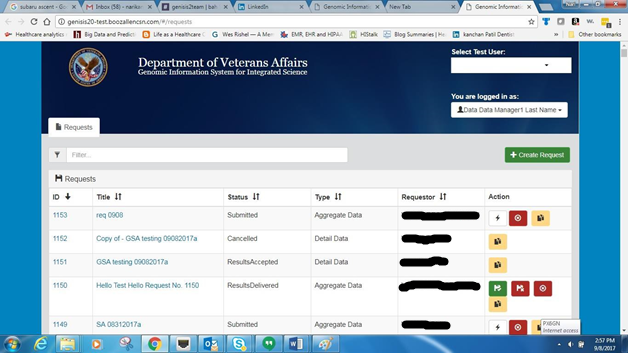
**Researcher Dashboard:** The Researcher Dashboard is also the landing page when a Researcher logs in using 2-factor authentication. It shows in various widgets the Tasks that are waiting for the Researcher to perform, request status changes and notifications from Genisis2, and other external components integrated with Genisis2.

**Request Operations Tab:** The Request Operations tab leads to options that can help a Researcher create a new request for data, modify a submitted request (internal state should be modifiable; for example, if a request has already been submitted to VINCI, it may not be modifiable) or cancel a request. **(Future build feature: will be part of this screen, but not part of Build 5.)**

**Data Operations Tab:** This tab allows a Researcher to open a current request and store information about any data operations they may be performing. For example, if they have developed a script to query a study mart or a query mart for some information or cleaning up some data, they can associate this script or query with a request, store additional metadata such as description, inputs, outputs, or processing changes intended. **(Future build feature: will be part of this screen, but not part of Build 5.)**

#### Genisis2 Manager/VINCI Manager Landing Page/Dashboards

Figure 13 is a screenshot of the Genisis2 Manager Login/Dashboard. From here, Genisis2 Managers can follow up on the status of the requests they are managing or perform actions such as move the workflow to the next step in the process. Alternatively, they can copy data from remote databases to the Genisis2 Database (Landing zone) for further processing before sending them off to Researcher study marts.

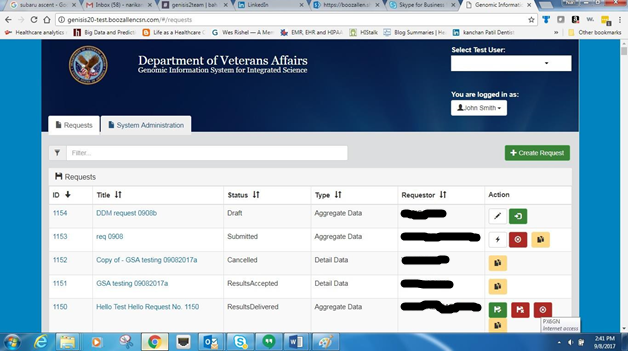


**Figure 13: Genisis2 Manager Landing Page/Dashboard Design**

**Genisis2 Manager Dashboard:** This is also the landing page when a Genisis2 Manager Genisis2 Manager logs in using 2-factor authentication. This shows in various widgets, the Tasks that are waiting for the Genisis2 Manager to perform, request status changes and notifications from Genisis and other external components integrated with Genisis. The VINCI Manager Dashboard is very similar to this; however, the VINCI Manager does not have the same access as the Genisis2 Manager so some of the buttons will not be available to the VINCI Manager.

#### Genisis2 System Administrator Landing Page/Dashboard

##### Figure 14 is a preliminary mock-up/wireframe of the Genisis2 System Administrator Landing Page/Dashboard. Once the preliminary design is approved and formal design is complete, this section will be updated again.



**Figure 14: Genisis2 System Administrator Landing Page/Dashboard**

**Genisis2 System Administrator Dashboard:** This is also the landing page when a Genisis2 System Administrator logs in using 2-factor authentication. This landing page has two main tabs

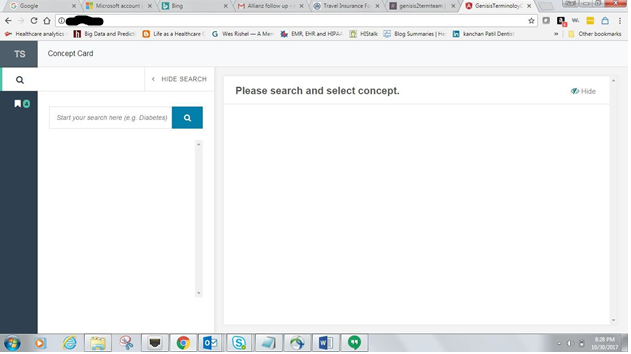
– Requests and System Administration. The Request tab enables all of the functions as that of any of the other roles – Researcher, Genisis2 Manager or the VINCI Manager can perform. As a super user, a Genisis2 System Administrator is able to perform all of these activities. The Systems Administrator tab will lead to two other tabs – User Management (Statistics about how many users are currently registered, broken down by role) and Request Metrics (How many Requests are in each stage of the workflow with the ability to drilldown to any one request and see all its details).

**Manager Workflow Tab:** This tab leads to features that allow a Genisis2 System Administrator to enable additional workflows. This feature may be seldom used after the initial implementation, but is available only to the Genisis2 System Administrators if needed. **(Future Release feature: will be part of this screen, but not part of this Release.)**

#### Terminology Service Workflows

#### Terminology Service Landing Page

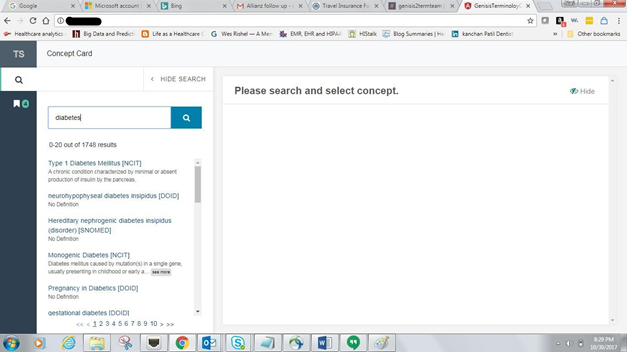
Figure 15 is a screenshot of the Terminology Service landing page when a Researcher or a Terminologist accesses that page.



**Figure 15: Terminology Service Landing Page**

Terminology Service consists of the backend loading a number of standard Ontologies like SNOMED-CT, NCIT, and DOID. These Ontologies make available to the user a number Clinical and Genomic Concepts for them to start their search for Data Elements. For example, if their research deals with Type I Diabetes, they start searching for Diabetes and navigate their way to data elements that may be attached to concepts that they are interested in. Figure 15 shows the landing page when users access Terminology Service. Users type in the terms they are interested in within the search box at the top left side of the landing page.

#### Search Results Page

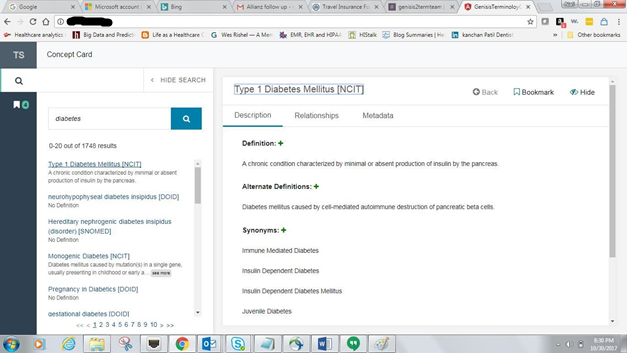


**Figure 16: Search Results Page**

Figure 16 shows a user typing in Diabetes in the Search box and concepts from all ontologies that deal with Diabetes are shown in the Search Results box on the left. Each result is shown with the Concept Title and some number of characters in the definition. The Ontology from which a concept was picked up is also displayed next to the Concept Name like [NCIT] or [DIOD].

The user can choose a concept that they want to explore further and click on the Concept name. This takes them to the Concept Card display in the next screenshot.

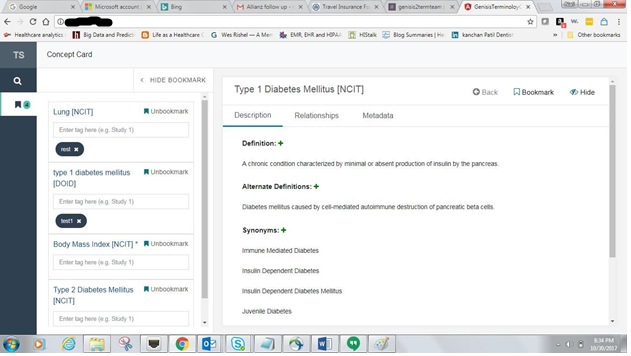
#### Concept Card Details



**Figure 17: Concept Card Page**

Figure 17 shows the Concept Card of one of the concepts selected for detailed display. It shows various properties of this concept under different tabs such as Description, Relationships and Metadata. The user navigates these tabs and to determine if this concept is the one they are looking for. If so, then the user can Bookmark it as shown in the next screenshot.

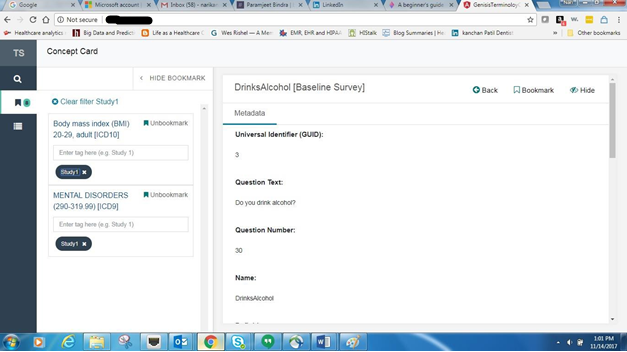
#### Bookmark a Concept



**Figure 18: Bookmark a Concept Page**

Figure 18 shows a page where the user bookmarks a Concept that they find of interest. If Data elements from MVP Clinical, Survey and Genomic data are attached to this concept, users can utilize the collection of bookmarked concepts that are of interest to them to query the data further for their research. The Bookmark Icon on the black bar on the left enables them to Bookmark the selected Concept. It also enables them to attach Labels to their bookmarks (shown in dark ovals in the middle panel. Labels enable a Researcher to classify their bookmarks and search for them at a later time.

#### Search Bookmarks by Label



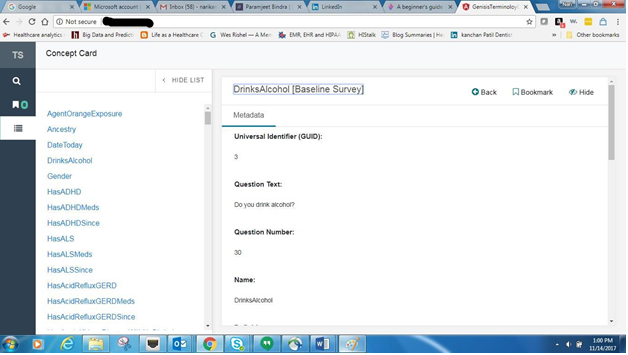
**Figure 19: Search Bookmarks by Label**

Figure 19 shows how a user can search for a bookmark by label. By clicking on a label of interest (“Study1” in the figure above), all bookmarks that have that label are filtered and shown in a list on the left as above. The user can select any one of the bookmarks and see the details.

The button “Clear Filter Study1” when clicked takes the user back to seeing all of the bookmarks they made.

#### Browsing Data Elements

The user can click on the List Icon on the left panel. This displays the data elements list in the central panel as shown in Figure 20. Selecting any one of the Data Elements displays the Concept Card in the right panel showing all the details of that data element. Each data element could be potentially mapped to one or more Concepts in the various Ontologies loaded into the system. By starting the search for Concepts by typing in Terms in the Search Box, the user can navigate down to Concepts that have Data Elements attached to them. They can bookmark these Concepts and use them to create a Cohort of VA MVP Volunteers for their research.



**Figure 20: Browsing Data Elements**

## Conceptual Infrastructure Design

The conceptual infrastructure design is a high-level overview of the infrastructure used to support the Genisis2 system. ***These activities are not in the scope of the Genisis2 project but provided here only for context. It is assumed that all of these are in place already or will be made available by VA on time for the Genisis2 development team.***

The current infrastructure supporting Genisis will be expanded as follows:

* + - Add Petabyte (PB) storage progressively through the years 2016 to 2018, with a projection of 45 PBs by June 2018.
    - Increase the bandwidth by at least 2X (doubling) to add throughput between VINCI and Genisis, and to meet the “big data” traffic demands anticipated with the addition of new studies.
    - Add computing capacity by interfacing Genisis to a public/private hybrid cloud with the capability to load-balance during “high burst” periods of intensive large-scale computing, and to provision cloud services for archival storage. Hybrid cloud deployments must be

documented in the Federal Risk and Authorization Management Program (FedRAMP) and requires a FedRAMP Authority to Operate (ATO).

* + - * An implementation strategy developed to demonstrate the feasibility of a hybrid cloud in a pilot instance to:
        + Test integration of the hybrid cloud in the pilot instance;
        + Optimize the hybrid cloud for performance and adaptive/perfective maintenance; and
        + Test hybrid cloud capabilities to support load-balancing during “high burst” periods of intensive large-scale computing, and provisioning for archival storage.

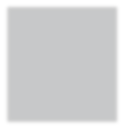
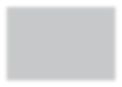
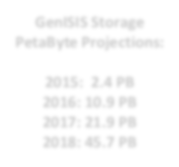
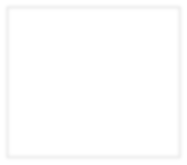
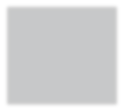
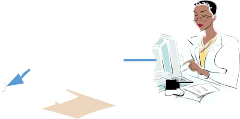
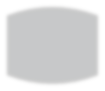
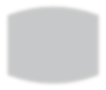
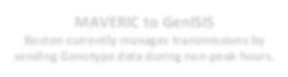
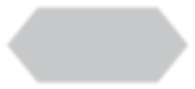
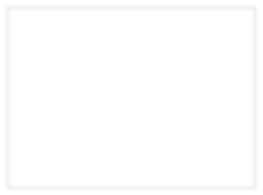
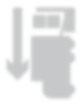
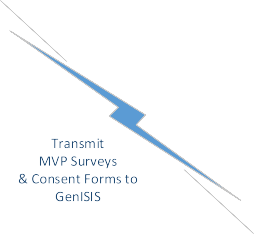
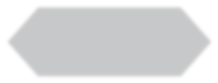
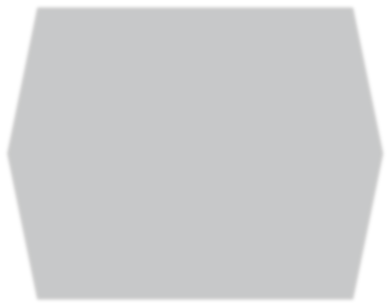
Figure 21 provides a high-level illustration of the requested expansion and enhancements for Genisis2.

MVP

Coordination Sites

**Scientific Public/Private Hybrid Cloud**

**Science DMZ – Allocate Cloud during Peak “High Burst” Demands**



System Administrator Dashboard

**GenISIS Storage**

**PetaByte Projections:**

**2015: 2.4 PB**

**2016: 10.9 PB**

**2017: 21.9 PB**

**2018: 45.7 PB**

WholeGenotype Exome

Genome SNP

**Study**

**Mart**

**Study**

**Mart**

**VINCI/CDW**

**VINCI Landing Zone**

**Redistribute Loads to Hybrid Cloud during Peak Demands**

**GenISIS System**

**Administrator**

**GenISIS Landing Zone**

**Honest Broker**

**Expand VINCI to GenISIS Bandwidth 2X**

GenISIS

**PIV + 2nd**

**Token**

**Blood Samples**

**from**

**Enrollees (DNA PREP shipped**

**To Sequencing Vendors)**

De-Couples PII, PHI and other Identifiable data

external inputs into GenISIS

**GenISIS**

**Scientific**

Researcher

Dashboard

**Request VINCI Data**

**Authentication**

**GenISIS Researcher**

**Data Warehouse**

**Genomic**

**Data**

**& Track Study**

**Access Control Lists (ACL) to all Study Marts**

**Study**

**Mart**

**Sequenced Data**

**Genotype Data**

**MAVERIC to GenISIS**

**Boston currently manages transmissions by**

**sending Genotype data during non-peak hours.**

**MAVERIC**

**Figure 21: Genisis2 “To Be” With Expanded Bandwidth, Storage, and Hybrid Cloud**

### System Criticality and High Availability

The Genisis2 application is not mission critical. The Recruitment and Enrollment Platform will support national users during business hours from 8:00am – 8:30pm ET, with 98% uptime.

However, the Scientific Platform (HPC cluster and analysis stations) will support Researchers with 100% uptime. The Scientific Platform is used across multiple geographic time zones, and must support concurrent, long-running computational jobs that may be scheduled at any time during a 24x7 period. The Genisis requirements for Availability, Allowable Downtime, Recovery Time Objectives (RTO) (for disaster recovery) and Recovery Point Objectives (RPO) (for disaster recovery) are identified in the Genisis2 RSD*.*

### Special Technology

Special Technology has not been identified for this system. Table 10 is a placeholder for this information should it be identified in the future.

**Table 10: Special Technology Requirements**

|  |  |  |  |
| --- | --- | --- | --- |
| **Special Technology** | **Description** | **Notional Location** | **TRM Status** |
| N/A | N/A | N/A | N/A |

### Technology Locations

This section describes the various technology components that will be used and their locations, as shown in Table 11. The host site is located at the Pittsburgh, PA Information Data Center with the MAVERIC facilities at the Boston VA Healthcare System identified as the Disaster Recovery (DR), development, and test environment. ***N/A below is Not Applicable in general or at least as of this Release. If the status changes, updates will be made as necessary in the future.***

**Table 11: (Grouping) Technology Location Details**

|  |  |  |
| --- | --- | --- |
| **Technology Component Production 1** | **Location** | **Usage** |
| Workstations | All VA research sites | Genisis access via web browsers |
| System Hardware | VAMC – Pittsburgh, PA | * Genisis host site web server * Application server * Database server * Data Mart server |
| Special Hardware | N/A | N/A |
| Interface Processors | N/A | N/A |
| Legacy Mainframe | N/A | N/A |
| Legacy Application Server | N/A | N/A |
| Legacy Databases | N/A | N/A |
| Other | N/A | N/A |

|  |  |  |
| --- | --- | --- |
| **Technology Component Production 2** | **Location** | **Usage** |
| N/A | N/A | N/A |

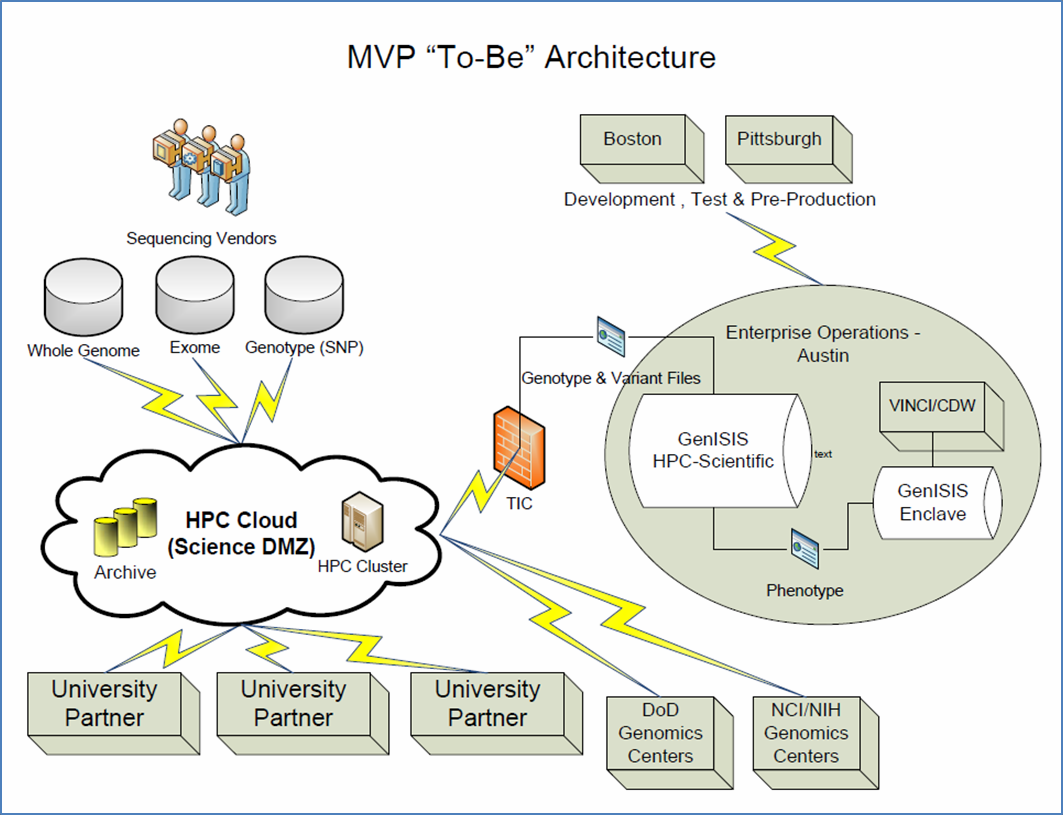
|  |  |  |
| --- | --- | --- |
| **Technology Component Certification** | **Location** | **Usage** |
| N/A | N/A | N/A |

|  |  |  |
| --- | --- | --- |
| **Technology Component Education** | **Location** | **Usage** |
| N/A | N/A | N/A |

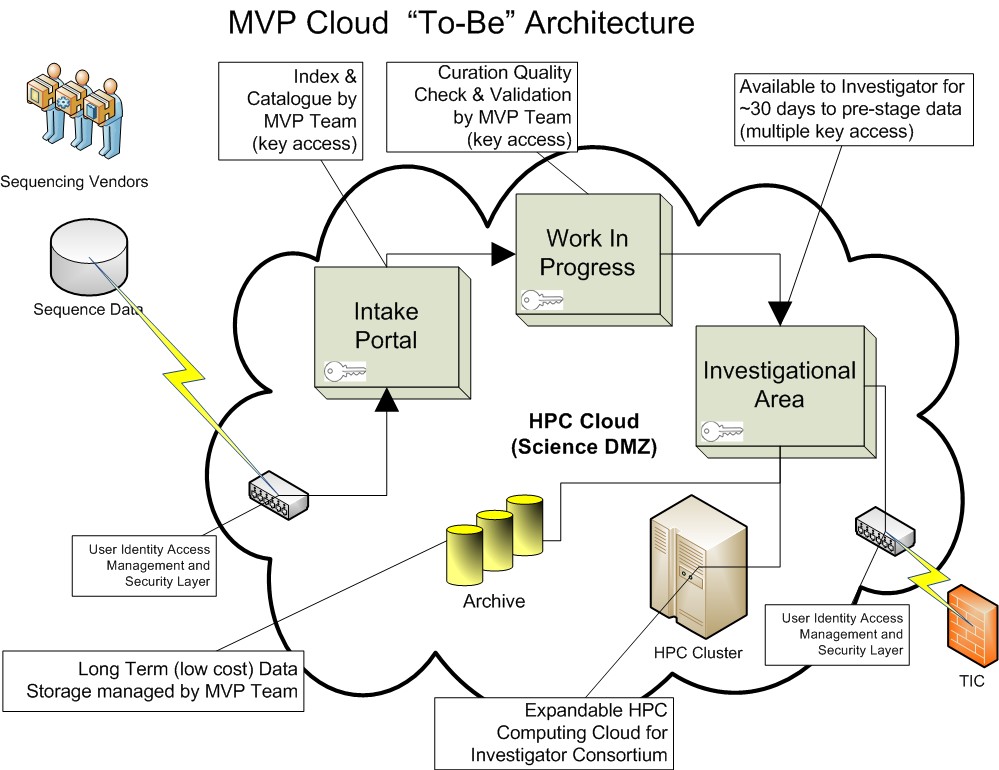
### Conceptual Infrastructure Diagram

The following diagrams depict a concept for the future Genisis infrastructure, which provides flexible, on-demand access to MVP cloud services. This cloud model is expected to enable implementation of ongoing, changing requirements to expand or contract capabilities as needed.

This includes HPCC and PB storage resources. Figure 22 illustrates how all groups access cloud resources, while Figure 23 focuses on the workflow of the HPC environment. ***Note that these are future states envisioned and not part of the project as of this release.***



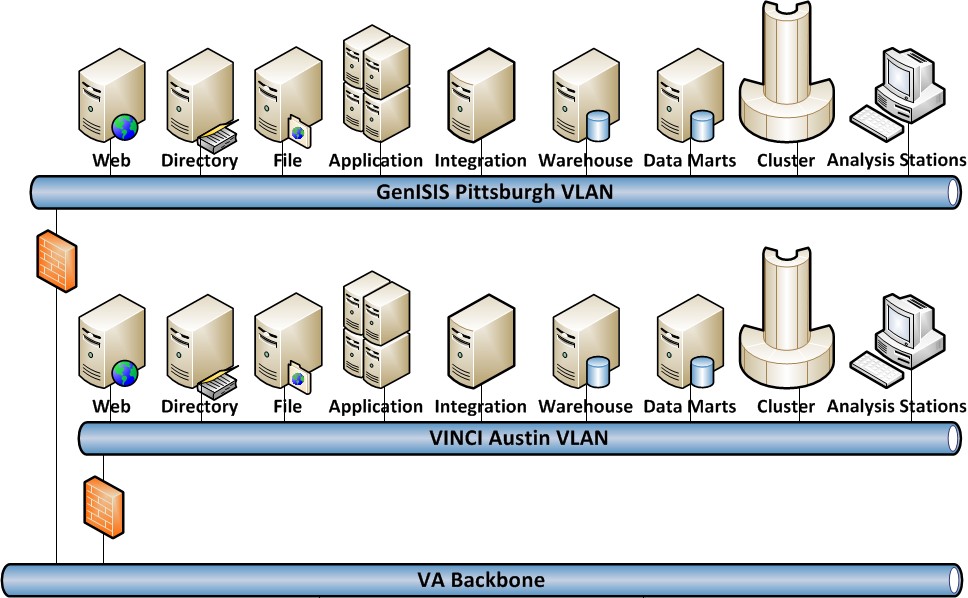
**Figure 22: Concept of Cloud for VA Genomics and Precision Medicine**



**Figure 23: High Performance Computing Cloud as a Science DMZ**

#### Location of Environments and External Interfaces

Figure 24 depicts the environments that will be supported, including the local networks to which they will be attached, and the locations at which they will be installed. Each of the external interfaces is shown in terms of where they enter the network.

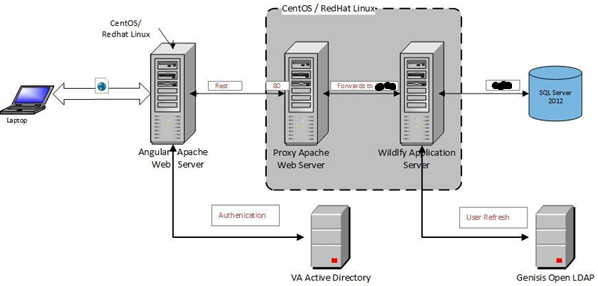


**Figure 24: Conceptual Networks and Environments**

#### Conceptual Production String Diagram

#### Genisis Workflows Production String Diagram

The Conceptual Production Environment for the Genisis Data Request Workflows consists of three tiers as shown in Figure 25.



**Figure 25: High-Level System Design for Genisis Data Request Workflows**

Genisis2 architecture for the Data Request Workflows consists of the following three tiers:

**Presentation/Webserver Layer:** Apache Webserver on Linux – does 2-Factor Authentication using a PIV card and Windows Authentication with the VA LDAP system. This webserver handles browser-based access from users. Traffic from this is redirected to a second webserver in the next layer. The second webserver is for access (not used currently but in place for the future) from other systems within the VA network to access Genisis2 application through an API (Application Programming Interface).

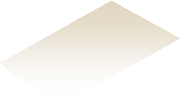
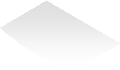
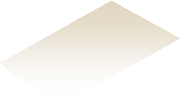
**Application/Business Logic/Workflow Layer:** Wildfly server on Linux– Application code in Java using the Activiti Workflow engine to coordinate the workflow.

**Database Layer:** Microsoft SQL Server 2012 on a Microsoft Windows 2012 R2 server.

#### Terminology Service Production String Diagram

The Conceptual Production Environment for Terminology Service consists of three tiers as shown in Figure 26.

Centos/RedHat Linux



**Port**

**Port 80**

**Web Server/ Application Server Running**

* **Apache HTTP**
* **Tomcat**
* **Jena/Fuseki**
* **SOLR**

**Database Server**

* **Windows Server 2008 R2**
* **Microsoft SQL Server 2012**

**Figure 26: High-Level System Design for Terminology Service**

Genisis2 architecture for Terminology Service consists of the following three tiers:

**Presentation/Webserver Layer:** Apache Webserver on Linux – This handles browser based access from users and sends them as REST calls to the Tomcat Application Server in the same Virtual server.

**Application/Business Logic/Workflow Layer:** Tomcat server on Linux– This application server is resident in the same Virtual Machine as the Webserver. Application code in Java implements the logic portion of Terminology Service.

**Database Layer:** Microsoft SQL Server 2012 on a Microsoft Windows 2012 R2 server.

# System Architecture

The Genisis2 application is designed to serve as a research computation platform for the VA Genomic Medicine Program. In order to leverage the experience of the worldwide genomic research community, the Genisis architecture follows open source and COTS architecture that is currently in wide use.

Logically, Genisis2 is divided into the following major subsystems:

* Relational Database, which provides the Genomic Data Warehouse and Data Query applications;
* A large capacity file store for storage of large flat files from genomic analysis;
* Web-based applications, which provide governance management, and mediate the data query and the computational job submission applications;
* Virtual analysis workstations for ad-hoc data analysis and job preparation; and
* An HPCC with large-capacity, parallel file system for performing genomic analysis.

## Hardware Architecture

Genisis2 hardware architecture is a centralized design that comprises server and storage components to meet the needs of large-scale genomic analysis. This section and Section 6.1 describe the server and storage components in more detail. Refer to Genisis2 RSD for requirements and projections to increase PB storage capacity from fiscal year (FY) 2016 through FY 2018.

A survey of major centers of genomic research showed that the predominant storage systems used were Isilon clustered file systems connected by GigE networks. The streaming performance characteristics of the Isilon system were used as the basic requirements of the system.

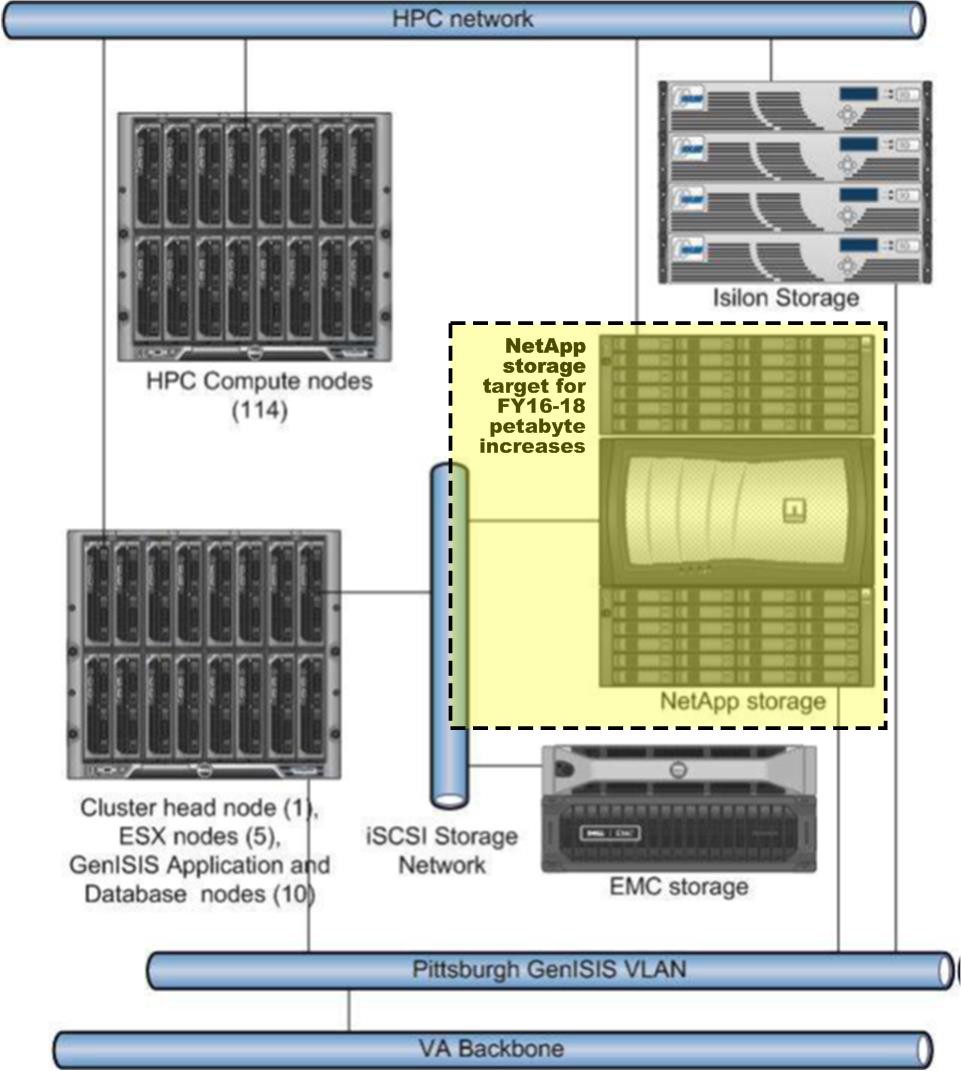
Genisis storage design consists of three classes of storage:

*Local Disk:* Each computational node is deployed with 300GB of Statistical Analysis Software (SAS) storage, which may be upgraded during a future refresh.

*SAN Storage:* A 30TB EMC CX4 Storage array is connected to a dedicated 10GigE iSCSI network to provide virtualized storage for the HPCC head node and the virtual analysis workstations.

*Clustered File Store:* 60TB Isilon and 2.7PB NetApps storage appliance provides a high- bandwidth analysis workspace for genomic analysis. Storage is presented via NFS over the non-blocking 10GigE HPCC data network.

The HPCC consists of commodity Linux servers that are chosen to maximize the performance- to-price ratio. The first iteration of Genisis consists of 114 identical Dell M610 Blade servers (HPCC Compute nodes). Each server has dual hex-core Intel Xeon processors, with 96GB memory, packaged in a dense, half-height blade. Computational servers chosen for future expansion will be based on best performance-to-price ratio available. Figure 27 depicts the Genisis hardware architecture.



**Figure 27: Genisis Hardware Architecture**

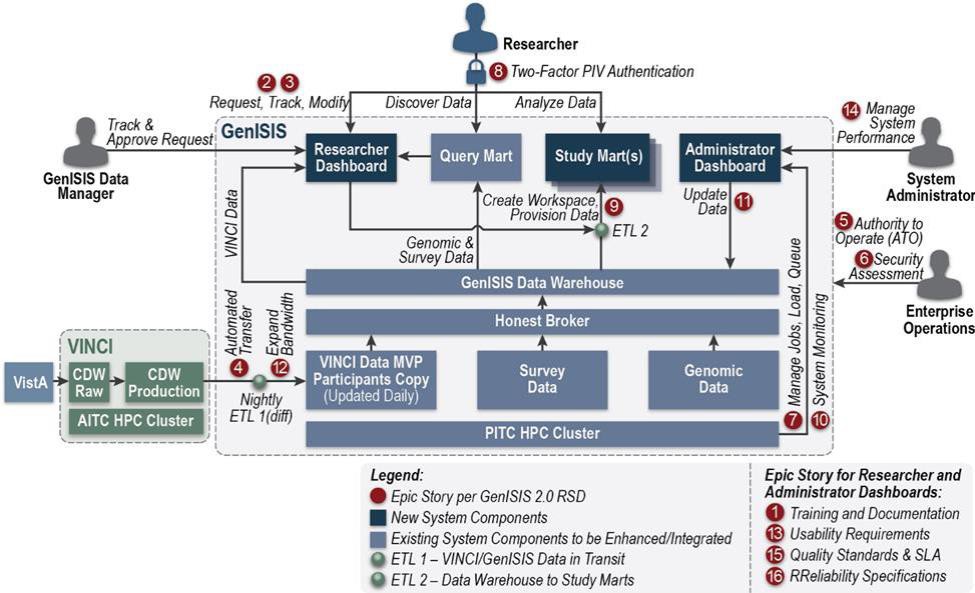
## Software Architecture

The Genisis2 front-end is accessed by a VA standard web browser and uses the existing authentication that is obtained from logging into a VA domain with valid user credentials. The VA has multiple domains, each with its own users. The goal for Genisis2 is to implement the new requirement of VA PIV access control for all VA domain users.

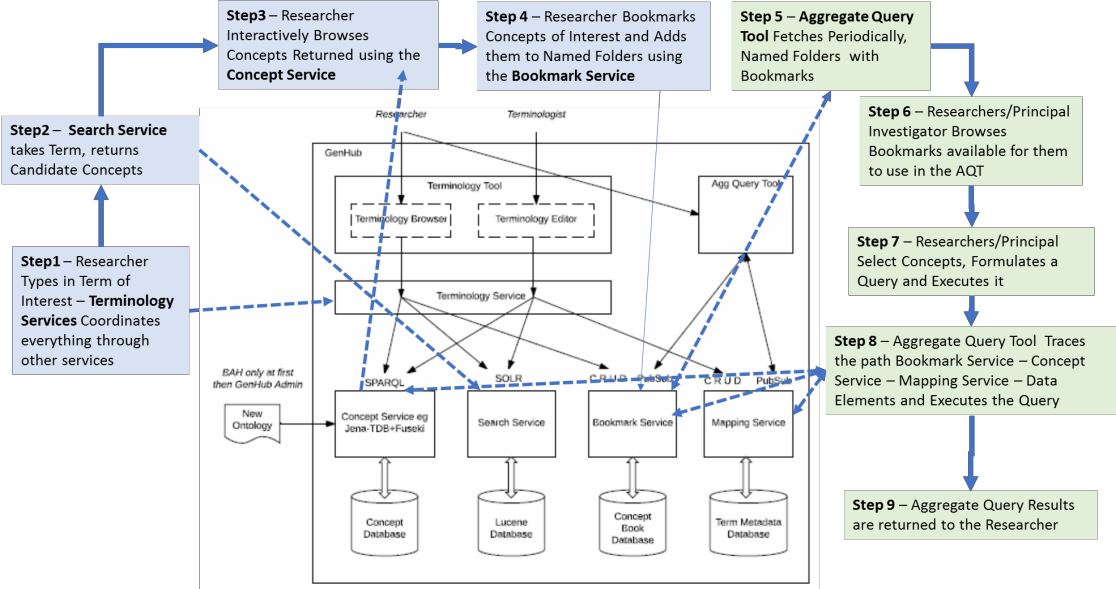
All Genisis2 interactions use Hypertext Transfer Protocol Secure (HTTPS), which requires signed server certificates from the VA’s security authority. The Apache Web Server will handle all authentications. User VA credentials used to authenticate will be passed to VA LDAP, along with the request; the container called depends on the URL of service call. For security purposes, Apache will only listen on localhost and are limited to Apache JServ Protocol (AJP) connections. Other ports will be disabled, as needed, if not required by the application server.

The Apache Web Server is configured to use Kerberos—the network authentication protocol, which provides the Single Sign-On (SSO) capability. Kerberos requires the creation of a key tab file on the Windows Domain Controller (DC), which acts as the Key Distribution Center (KDC). This key tab file contains Kerberos principals and their encrypted keys; therefore, it must be well protected, but readable, by the Hypertext Transfer Protocol daemon (HTTPD) process.

After the user is successfully authenticated, the user’s authorizations must be gathered and evaluated. Since the Active Directory does not store information associated with Genisis, the Genisis database itself is used to provide group membership. Figure 28 depicts the software architecture and context of Genisis2 Data Request workflows, and Figure 29 illustrates the software architecture and context of Genisis2 Terminology Service.



**Figure 28: Software Architecture and Context of Genisis2 Data Request Workflows**



**Figure 29: Software Architecture and Context of Genisis2 Terminology Service**

### Software Development Approach Analysis

#### Genisis2 Data Request Workflows

The following approach was taken to analyze the options available for software development based on the architectural principles outlined in Section 3.1. Our team:

1. Identified **Custom Software Development** as one of the options available to us for the development of Genisis2 application.
2. Identified a number of candidate **Commercial Off-the-Shelf (COTS)** and **Open Source** software products that could be possible candidates for ***alternatives for Custom development.***
3. Identified key features that these software products should have for being considered for Genisis2 implementation as ready-made features, or features that can be modified readily for use.
4. Identified the status of the above software products according to the One-VA Technical  [Reference Model v16.8, Technology Standard List.](https://www.va.gov/TRM/) Chosen for consideration were those products that were either **Approved** or **Approved with Constraints** AND for **CY2017** and **CY2018**.
5. In the case of Custom Software Development, the assumption was made that key features considered in the products for comparison could be developed within the software application itself.

***In addition to the Genisis2 application, we also considered COTS or Open Source products that can be used in the Data Operations part of the project (Copying files between one Landing Zone to another and between a Landing Zone and a Study Mart or a Query Mart).*** The following analyses present the development team’s findings for these two parts of Genisis2.

#### Analysis for the Genisis2 Application

The results of the analysis of alternatives for the Genisis2 Application are illustrated in Figure 30.

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Option** | **TRM Approval Version**  **or Approval Approved Cost Model with for CY2017**  **Constraints** | | | **General Notes** | **Cost** | **Real-time Communi Ability to Alerts and ty Customize Notificatio Adoption**  **ns** | | | **General Maturity Integration**  **Capabilities** | | **Workflow/ Process Flow** | **Active Directory Integratio n** |
| **Custom Development** | NA | NA | NA | Necessary Features Can be | 3 | 4 | 4 | NA | NA | 4 | 3 | 3 |
|  |  |  |  | Built |  |  |  |  |  |  |  |  |
| **Microsoft SharePoint Portal** | 3 | 2013  and Above | COTS | Intranet Focused, Content Management based Workflows | 3 | 3 | 3 | 3 | 3 | 2 | 2 | 3 |
| **LifeRay Portal** | 3 | 7.0 and Above | Only COTS version is Viable | Intranet Focused, Content Management based Workflows | 2 | 2 | 2 | 1 | 1 | 2 | 2 | 2 |
| **Alfresco One/Activiti Workflow** | 3 | 201605  GA  Version | Only COTS version is Viable | Intranet Focused, Content Management based Workflows | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 |
| **Nuxeo** 0 | | | | | | | | | | | | |
| **Plone** 0 | | | | | | | | | | | | |
| **Open Atrium/Drupal** 0 | | | | | | | | | | | | |
| **Portofino** 0 | | | | | | | | | | | | |
| **MindTouch** 0 | | | | | | | | | | | | |
| **Trac Prohibited** | | | | | | | | | | | | |
| **for use** | | | | | | | | | | | | |

**Figure 30: Summary of Analysis of COTS Alternatives for the Genisis2 Application**

Salient highlights of the analysis results:

* We did not analyze further if a COTS or an Open Source product did not have TRM approval (e.g., Nuxeo, Plone, etc.) or if it was prohibited for use within VA by OIT (Trac). They merit an empty circle.
* All of the COTS and Open Source products that were approved had workflow designs that were tied to document workflows or content workflows (such as approval of web pages across multiple people, for example). That is the reason they merit only half a circle for workflows in the chart above.
* Products like LifeRay and Alfresco had community editions that had limited functionality and only COTS versions of them had the features needed for Genisis2. This is the reason they both merit the half circle for Cost.

#### Analysis for the Data Operation Tools

The results of the analysis of COTS alternatives for the Genisis2 Application are illustrated in Figure 31.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **TRM**  **Approval**  **or Version**  **Option Approval Approved Cost Model**  **with for CY2017 Constrain**  **ts** | | | | **General Notes** | **Real-time**  **Alerts and Community Notifications Adoption** | | **General Active**  **Maturity Integration Directory**  **Capabilities Integration** | | |
|  |  |  | COTS - | Most widely used COTS |  |  |  |  |  |
| **Titan SFTP** | 3 | 16.X and | $1950 for | product for automating | 3 | 4 | 4 | 3 | 3 |
| **Enterprise** |  | above | Enterprise | Secure File Tranfer Uses |  |  |  |  |  |
|  |  |  | Edition |  |  |  |  |  |  |
| **WinSCP** | 3 | 5.7.7 and | Open | Custom Code Built Around the | 2 | 4 | 3 | 2 | 1 |
|  |  | above | Source | Product |  |  |  |  |  |

**Figure 31: Analysis of COTS Alternatives: Genisis2 Research Data Handling Tools**

Salient highlights of the results:

* The development team found many alternatives for transferring files securely.
* The choices were narrowed down to those products that can recognize files when they appear in folders, capable of real-time notifications when transfers are completed and supported SFTP (Secure File Transfer Protocol).
* The choices were further narrowed down to the most widely used products represented by the Community Adoption full circles. WinSCP has a large following in the users of Open Source products but on lighter set of simpler tasks. The Titan SFTP Enterprise is an enterprise-ready tool used widely in large organizations and more mature (V16 vs V5.7 for WinSCP).
* The Titan SFTP Enterprise product has many enterprise-class features and is very widely used in organizations with great reviews, especially for use with Windows platforms.

#### Terminology Service

Component selection for Terminology Service was a much simpler process because of the limited availability of qualified, open-source components. In addition, a number of components were common with the Genisis2 Data Request workflow system. The final components chosen were:

* + - * + **Apache HTTP** – For handling browser requests through this webserver component. The analysis performed for Genisis2 Data Request Workflows was useful for this selection since it was identified as one of the better components.
        + **Jena/Fuseki** – This is the Triple Store database that was chosen since it is one of the few Triple Store database software that is open-source, TRM approved, and already in use within the VA. A formal performance evaluation was made to make sure this has the necessary performance characteristics necessary for use.
        + **SOLR** – This is the text search engine component chosen. It is also open-source, TRM approved, and in widespread use within the VA.
        + **Microsoft SQL Server 2012** – This component was chosen for the Database Server since the analysis for Genisis2 Data Request Workflows already identified it as one of the best options.

### Software Development Approach Chosen

The Development team recommends the following approaches based on the analysis of alternatives above and their results:

#### Genisis2 Application

The development team recommends the Custom Development approach for the Genisis2 application. The reasons are as follows:

* All of the other alternatives, even though they have many needed features at first blush are oriented towards simple Intranet, document-based or content-based workflow applications. Genisis2 automates the data request, preparation and fulfillment process and is, as such unique. Trying to force-fit this workflow on to a tool designed for content management applications may be a mismatch.
* Genisis2 workflows may evolve over a period with many subtle variations between one and the other. Custom development may be the best way to keep these features up-to-date and without bending to the quirks of the product chosen.
* Custom code can be tuned for high performance while if they are implemented with COTS or Open Source products may come with unnecessary performance penalties with respect to features within the product that the Genisis2 application may never use.

#### Product for Data Operations

The development team recommends the Titan SFTP Enterprise product for Data Operations. The reasons are as follows:

* The Titan SFTP Enterprise product includes features such as recognizing when a file appears in a directory and automating actions following that with scripts that can be invoked.
* This product contains real-time notifications needed by email or mobile SMS/Texts.
* It seems to be used widely in organizations, stable and long standing (version 16 is the one approved with constraints by VA OIT TRM).

Even though this is a COTS product with a nominal cost, the development team thinks that this will help move Genisis2 Data Operations to a higher level of automation, especially when it comes to moving data between landing zones or between landing zones and marts (study or query).

## Network Architecture

Communications between system entities are divided into the following categories:

* Web browsers to web application
* Web application to web service
* Web application/service to database
* HTTP server (Apache) to Application Server (Apache)
* Web application/service to LDAP server
* HTTP server (Apache) to Active Directory Domain Controller

All communications will be encrypted using Transport Layer Security (TLS) version 1.2 or higher. A signed server certificate will be assigned to the system hosting the Apache HTTP server. Any web browser connecting to the Apache HTTP server will be redirected to a TLS connection, if not already using TLS. Hypertext Transfer Protocol (HTTP) 1.1. will be the data exchange protocol. User identities will be authenticated and user authorizations will be enforced.

Connections between web applications and web services will use Representational State Transfer (ReST) mechanisms over HTTP. Data exchanges will be typically Extensible Markup Language (XML) documents. Connections to external web services will be authenticated and authorized.

Web application/service connections to database will use TCP/IP connections. All connections will be authenticated and authorized to protect the contents. TCP/IP is used to support remote database connections.

Connections to LDAP servers will use LDAPs to ensure secure exchanges for user information. Connections will be authenticated and authorized.

Single-sign-on will be supported from the HTTP server using Kerberos authentication with an Active Directory Domain Controller.

## Service Oriented Architecture / ESS

Genisis2 Presentation Layer talks to the Application Layer only through an Application Programming Interface (API), which is documented and presented along with the delivery of the software. This SOA approach can be useful for any other external software system to call Genisis2 the same way its own User Interface calls it.

## Enterprise Architecture

Genisis2 will continue to retain its status as a “Medical Device” program, with the intent to remain exempt from the EA Technology Reference Model (TRM) requirements. Because of the high-computational/scientific computing aspects related to investigational genomics, Genisis is “isolated” and behind a Virtual Local Area Network (VLAN)/ Access Control List (ACL), and does not sit on the VA Backbone; therefore, Genisis2 shall meet the criteria required of a Medical Device Isolation Architecture.

The following is a list of the COTS software planned for the Genisis2 enterprise development environment, with TRM approval status and versions. The development team may update or alter as necessary, but must also abide by Architecture Engineering Review Board (AERB) constraint: “The project will ensure that all the tools in use are compliant with the TRM. If a tool is not compliant, the project will seek a waiver for its use.” Table 12 provides the technology location details.

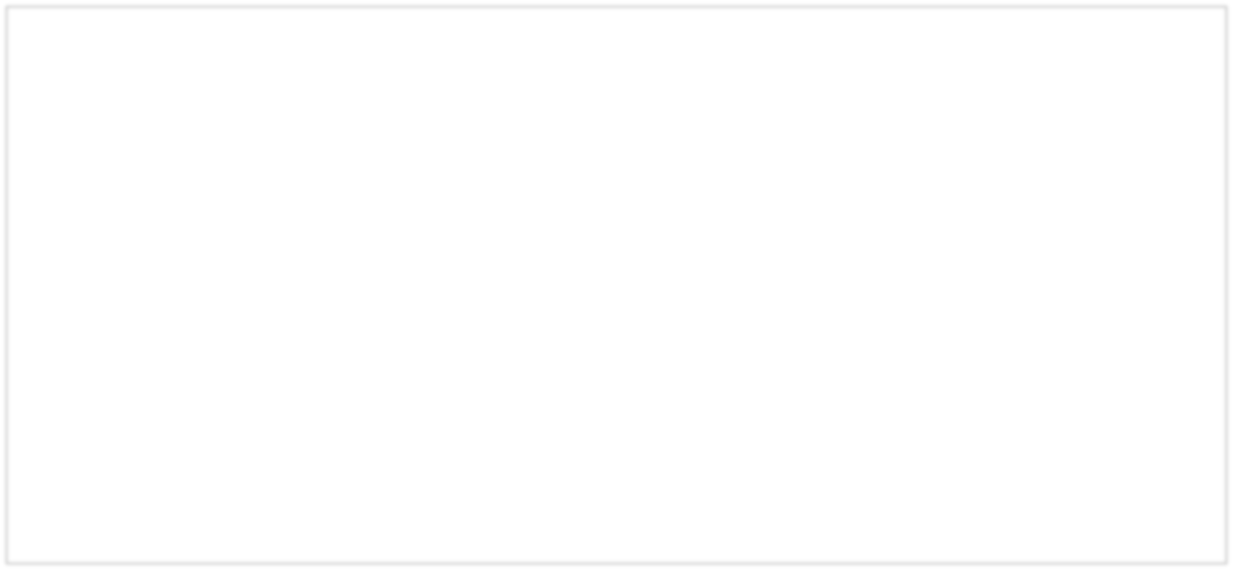
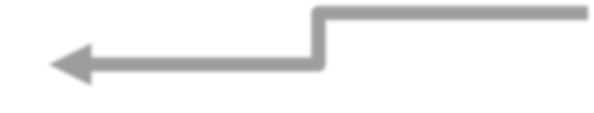
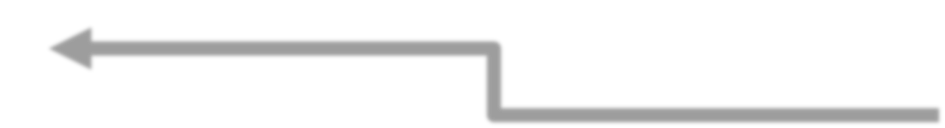
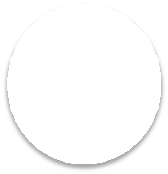
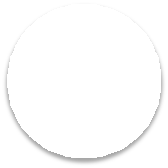
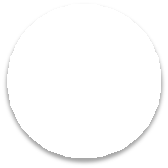
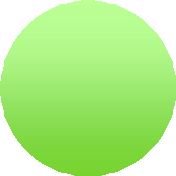
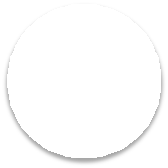
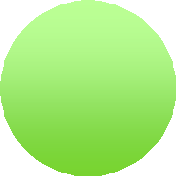
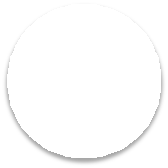
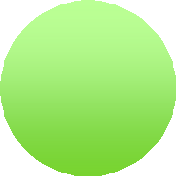
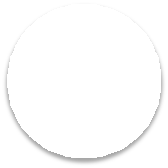
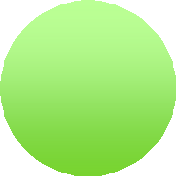
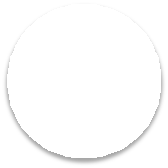
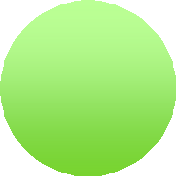
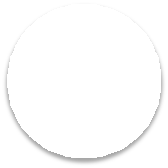
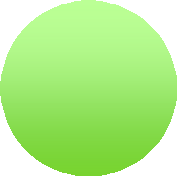
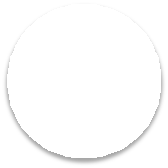
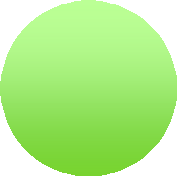
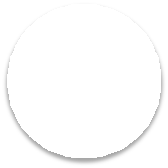
**Table 12: (Grouping) Technology Location Details**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **#** | **Product Name** | **Version** | **TRM**  **Approved (Y/N)** | **Description/Use/TRM link** |
| 1 | Apache HTTP Proxy Server | 2.3 + | Y | Web Server, Authentication https://www.va.gov/TRM/ToolPage.aspx?tid=5009 |
| 2 | WildFly | 10.0.0-  Final | Y | Web Application Container, Genisis Web Apps <https://www.va.gov/TRM/ToolPage.aspx?tid=240> |
| 3 | Apache Tomcat | 9.0 | Y | Web Application Container, Genisis Web Apps https:/[/www.va.gov/TRM/ToolPage.aspx?tid=5451](http://www.va.gov/TRM/ToolPage.aspx?tid=5451) |
| 4 | Apache Jena/Fuseki | 3.4.0 | Y | Triple Store Database and Query Engine https:/[/www.va.gov/TRM/ToolPage.aspx?tid=9476](http://www.va.gov/TRM/ToolPage.aspx?tid=9476) |
| 5 | Apache SOLR | 6.0.0 | Y | Text Search Engine https:/[/www.va.gov/TRM/ToolPage.aspx?tid=7005](http://www.va.gov/TRM/ToolPage.aspx?tid=7005) |
| 6 | Microsoft SQL Server | 2012 | Y | Relational DBMS for Genisis Not **used for external connections.**  <https://www.va.gov/TRM/ToolPage.aspx?tid=5020> |
| 7 | Red Hat Enterprise Linux | 7.x | Y | Operating System <https://www.va.gov/TRM/ToolPage.aspx?tid=6367> |
| 8 | Windows Server | 2012 | Y | Operating System <https://www.va.gov/TRM/ToolPage.aspx?tid=35> |
| 9 | JDK – Sun  Java SE Development Kit | 1.8.0\_92 (8u92) | Y | Java Runtime Environment and Utilities <https://www.va.gov/TRM/ToolPage.aspx?tid=5161> |

# Data Design

Figure 32 represents the request data status over its lifecycle in Genisis2 Data Request Workflows.

**Figure 32: Request Status Lifecycle**



**Data Manager VINCI Places Data Manager Data Manager**

**Data Request Completes Data Request Data in VINCI Copies Data Copies Data Researcher Submitted Request Sent to VINCI Landing Zone to GenISIS to Query Accepts Data**

**Review Landing Zone Mart/Study**

**– Reviews it Mart after QC**

**Data Manager Needs Data**

**Request is Modifications**

**Modified**

**Researcher Needs Data Modifications**

Status Changes of a Request

Figure 33 shows in more detail the various State Transitions and Status changes.



Data Request Workflow – Status Transitions

Results Accepted

Request Not Approved

Request Not Accepted

Results Not Accepted

New Request

Submitted Request

Conditionally Approved

Processing

Request Accepted

Request Delivered

Returned Request

|  |  |  |  |
| --- | --- | --- | --- |
| **From Status** | **To Status** | **Actor** | **Description** |
| New Request | Submitted Request | Researcher or DM | A Researcher or DM may create a new Request |
| Returned Request | Submitted Request | Researcher or DM | A Researcher or DM may modify a Request and re-submit |
| Submitted Request | Conditionally Approved | Data Manager | A DM may conditionally approve a Request |
| Submitted Request | Returned Request | Data Manager | A DM may send a Request back for correction |
| Conditionally Approved | Processing | Data Manager | A DM may approve a Request and send it to VINCI or MVP |
| Conditionally Approved | Request Not Approved | Data Manager | A DM may not approve a Request |
| Processing | Request Accepted | VINCI / MVP | VINCI/MVP may accept a Request from a DM |
| Processing | Request Not Accepted | VINCI / MVP | VINCI/MVP may not accept a Request from a DM |
| Request Accepted | Request Delivered | VINCI / MVP | VINCI/MVP may fulfill a Request & provide Results to a DM |
| Request Delivered | Results Accepted | Data Manager | A DM may review and accept the Results from VINCI/MVP |
| Request Delivered | Results Not Accepted | Data Manager | A DM may review and not accept Results from VINCI/MVP |

GenISIS 2.0 Preliminary Design Review

**MVP-GenISIS 2.0 -HighLevelOrientation- *DRAFT (for discussion purposes***

***Contract No.: VA118-11-D-1008/VA118-1008-0101***

***only)***

Oct 17, 2016 15

**Figure 33: Status Transitions**

## Human Machine Interface

### Genisis Data Request Workflows

The following provides details of users and what they accomplish using screens in the Human Machine Interface:

* **Researchers –** Researchers are typically Researchers who are requesting Genomic Data, Clinical Data and Survey Data collected from Veteran volunteers in the MVP program. They log in using two-factor authentication, type in their request, attach documents explaining what they need in more detail, and submit it to the Genisis2 Manager. Once they get notification that the data is ready and copied to their study mart, they can start their research efforts with the data provided.
* **Genisis2 Managers –** Genisis2 Managers are data coordinators that referee requests for data between sources of that data and the Researchers. Some data may not be available and some data may have some clarifications that get passed back and forth before the data can be made available. There may be some data cleaning efforts that is done on the data that is obtained from the source. Genisis2 does not handle those activities but Genisis2 Managers may add a comment or two regarding those activities.
* **VINCI Managers –** These managers are sources within the VA for Clinical Data and Survey Data. Genisis2 Managers coordinate with VINCI Managers in reconciling the data that’s available and the data requested by researchers. In some cases, the same kind of data may be available in more than one location within the Clinical Data Warehouse (CDW), for example. Height and Weight may be recorded in VA Primary Clinics, In Patient Hospitals or VA Specialist divisions, being stored in different parts of the CDW. The Data Source and Destination Managers make sure the right information in extracted and provided to researchers (Researchers in the Genisis2 sense).
* **Genisis2 System Administrators –** These are super users that can perform all of the functions that a Researcher or a Genisis2 Manager can do. In addition, they can update from the Genisis2 Open LDAP server, the latest set of users. Permissions such as Researcher, Genisis2 Manager, VINCI Manager or another Genisis2 System Administrator are assigned in the Genisis2 OpenLDAP server (central repository for permissions for many other systems including Genisis2).

### Terminology Service

The following provides details of users and what they accomplish using screens in the Human Machine Interface:

* **Researchers/Terminologists –**Genisis2 Terminology Service consists of the backend loading a number of standard Ontologies like SNOMED-CT, NCIT, and DOID. These Ontologies make available to the user a number Clinical and Genomic Concepts for them to start their search for Data Elements. For example, if their research deals with “Diabetes”, they start searching for “Diabetes” and navigate their way to data elements that may be attached to Concepts that they are interested in. The user can then bookmark these Concepts. They can also label them for future grouping and search. Users then use

those bookmarked Concepts attached to Data Elements for requesting MVP Clinical, Survey and Genomic data.

### Interface Design Rules

Below are some of the conventions and standards for designing the GUI:

* It should follow the general look and feel of the VA Office of Research and Development (ORD) websites and tools.
* It should follow the general Logo and other Identity related colors that the Veteran’s Administration uses in all of its tools and websites.
* It should provide the standard set of links at the footer of each screen that most of the VA sites and tools have in their footers.
* It should be fairly easy to navigate and use.
* It should be Section 508 compliant.

### Inputs

Below are some of the characteristics of the Inputs to Genisis2:

* All inputs to Genisis2 are planned to be by way of the keyboard and mouse clicks only. No forms are scanned in. Documents such as PDF, Microsoft Word, PowerPoint, Excel spreadsheets (all of Microsoft Office formats) may be attached to what they type in on the Genisis2 screens.
* No forms packages are used in the GUI.
* Messages – Positive or Negative are displayed at the Top Left corner of all screens in the Genisis2 application when inputs are wrong.
* Access restrictions are implemented by Roles – Researcher, Genisis2 Manager, VINCI Manager or Genisis2 System Administrator. Different tabs are displayed or taken away depending on who logs in.
* Security is enforced at login time using the VA PIV Card based Two-Factor Authentication. This and the VA email address is used to authenticate users. Authorization is controlled by the Roles that a user has in the system.

### Outputs

Below are some of the characteristics of the outputs from the Genisis2 application:

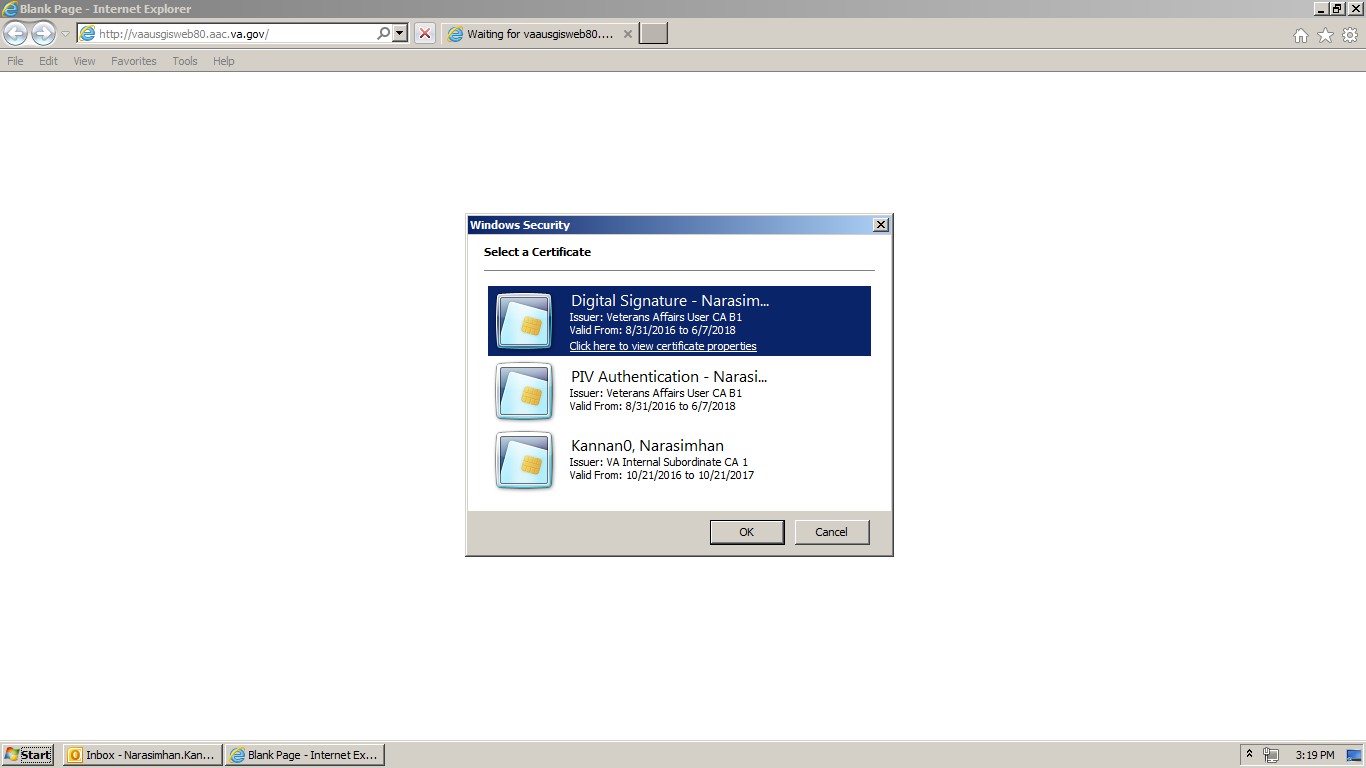
* Genisis2 could transfer data automatically behind the scenes because of actions on the Genisis2 screens.
* Messages – Positive or Negative are displayed at the Top Left corner of all screens in the Genisis2 application when operations are successful or not successful.
* Outputs could also be reports generated as Txt or CSV files in the future when reporting functions are added in a future release. Reports envisioned currently are ad-hoc, on demand generated but could also be scheduled periodically in the future.

### Navigation Hierarchy

The navigation hierarchy will be elaborated on as the application becomes more fully developed.

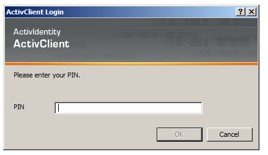
### Login Screens

When the user types in the appropriate URL for Genisis2, they reach a screen that looks like this. They need to insert their PIV Card into the Card slot; appropriate certificates all pop up as shown in Figure 34. The user selects the VA PIV certificate.



**Figure 34: Initial Login**

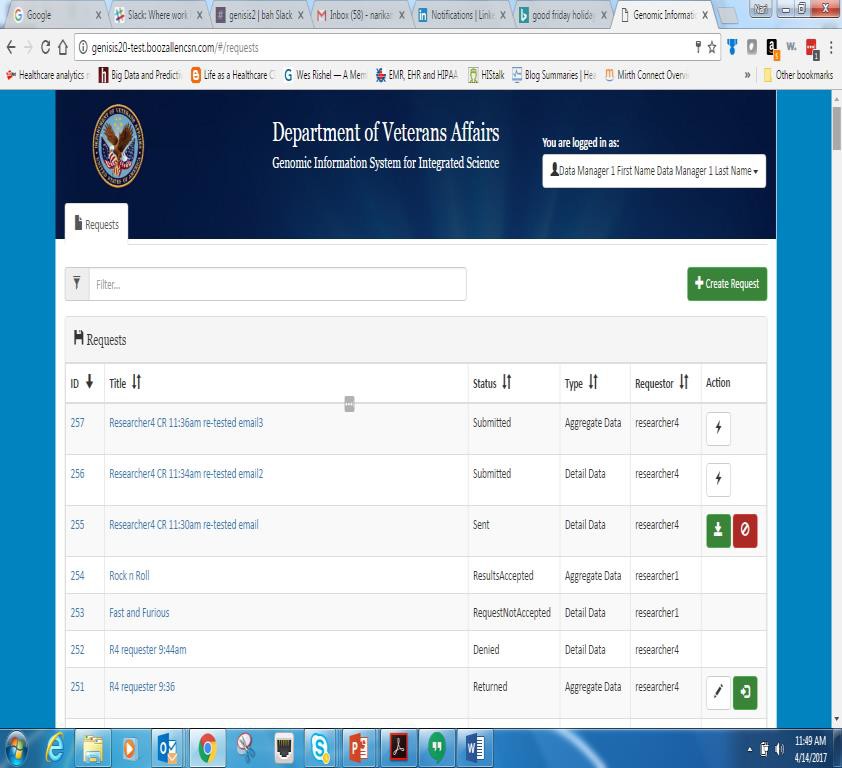
This prompts for the PIV Pin and then logs the user in as shown in Figure 35.



**Figure 35: PIN Prompt**

#### Landing Zone and Navigation Aspects – Genisis Data Request Workflows

Once the user logs in, they are placed in a Landing zone and the various aspects of navigation are as shown in Figure 36. Annotations are also provided.



Username, Profile, Help Pulldown Menu

Create a New Request

Request It, Link to Details of a Request

Request Status

– Submitted, Returned, Accepted, etc

Tabs – Different

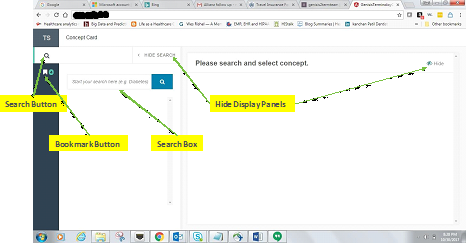
User Types See different ones

**Figure 36: Landing Zone and Navigational Aspects**

#### Landing Zone and Navigation Aspects – Terminology Service

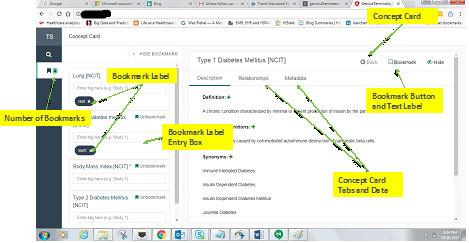
When the user accesses the Terminology Service, they reach a landing page as shown in Figure

37. This is where the user starts their search for concepts by typing in terms in the Search Box.



**Figure 37: Terminology Service Landing Zone and Navigational Aspects**

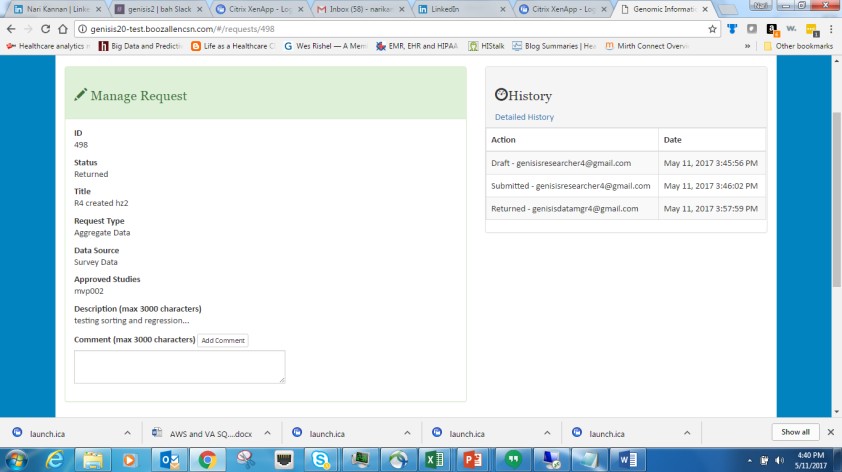
Figure 38 shows the Navigational Aspects when the user has found a Concept of interest and is interested in Bookmarking that Concept



#### Figure 38: Terminology Service – Bookmarking Concept

### Details of Any One Data Request

Figure 39 is a screen shot of the details of any one request.



Detailed History Link – Comments and every step the request has taken

Details of a

Request

Status Changes

**Figure 39: Navigating a Single Request**

## DBMS Files

Table 13 summarizes the Databases involved in the Genisis2 Application.

**Table 13: Genisis2 Databases**

|  |  |  |  |
| --- | --- | --- | --- |
| **Database Name** | **Description** | **Type** | **Steward** |
| **Genisis\_Activiti\_DB** | All workflow related data is handled in tables created by the Activiti workflow system. **Contains workflow data only.** | Relational Database | ORD |
| **Genisis\_DB** | The **Main Genisis Database** that handles all data not handled by the Workflow System above (such as Comments table) | Relational Database | ORD |
| **TS\_DB** | The main **Terminology Service Database** that handles user related data such as Bookmarks, Labels, etc. | Relational Database | ORD |

## Non-DBMS Files

Terminology Service creates a Triple Store database that stays in memory after all the Ontologies are loaded. The Ontologies are in RDF format.

## Data View

Table 14 provides a high-level view of the tables in each of the databases involved:

**Table 14: Genisis2 Tables in the Genisis2 Databases**

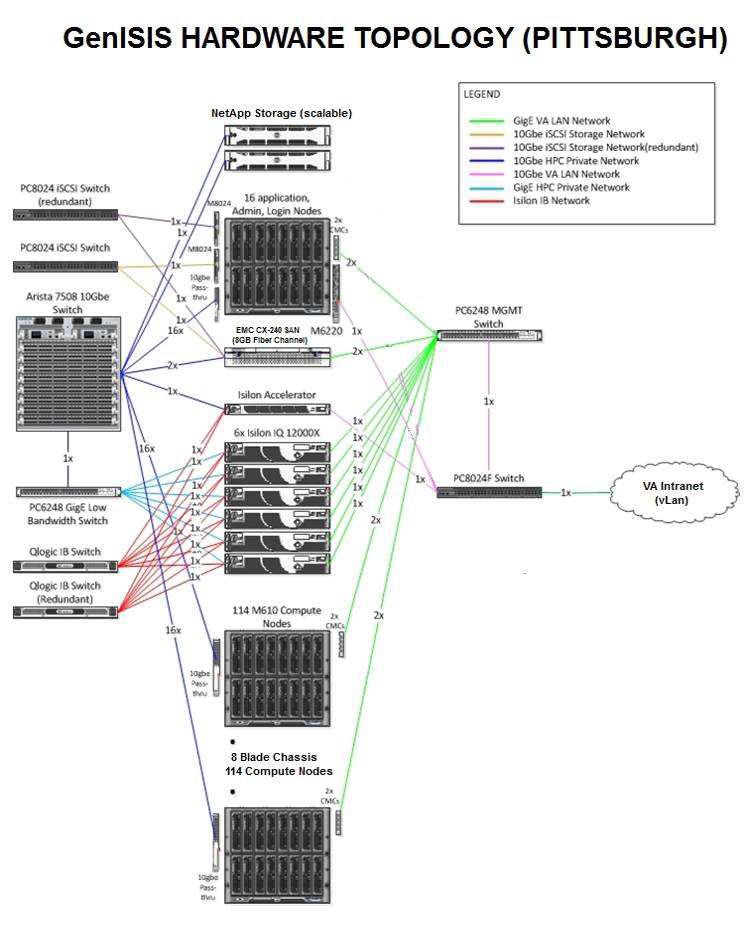
|  |  |  |
| --- | --- | --- |
| **Genisis\_DB** | **Genisis\_Activiti\_db** | **TS\_DB** |
| Comment\_History | ACT\_EVT\_LOG | Bookmarks |
| Management\_Table | ACT\_GE\_BYTEARRAY | Concept\_mapping |
| Request | ACT\_GE\_PROPERTY | Concept\_mapping\_data\_elements |
| Request\_History | ACT\_HI\_ACTINST | Data\_element |
| RequestType | ACT\_HI\_ATTACHMENT | Data\_element\_components |
| Role\_Type | ACT\_HI\_COMMENT | Data\_element\_source |
| Source | ACT\_HI\_DETAIL | Data\_element\_type |
| Study\_Approval | ACT\_HI\_IDENTITYLINK | Labels |
| User\_Approver | ACT\_HI\_PROCINST | Simple\_data\_element |
| User\_Role\_Type | ACT\_HI\_TASKINST |  |
| User\_Type | ACT\_HI\_VARINST |  |
| Users | ACT\_ID\_GROUP |  |
| Workflow\_Status | ACT\_ID\_INFO |  |
|  | ACT\_ID\_MEMBERSHIP |  |
|  | ACT\_ID\_USER |  |
|  | ACT\_RE\_DEPLOYMENT |  |
|  | ACT\_RE\_MODEL |  |
|  | ACT\_RE\_PROCDEF |  |
|  | ACT\_RU\_EVNT\_SUBSCR |  |
|  | ACT\_RU\_EXECUTION |  |
|  | ACT\_RU\_IDENTITYLINK |  |
|  | ACT\_RU\_JOB |  |
|  | ACT\_RU\_TASK |  |
|  | ACT\_RU\_VARIABLE |  |

# Detailed Design

## Hardware Detailed Design

Figure 40 depicts a view of the hardware components, network segments, bandwidth and connectivity that make up the existing Genisis environment. Storage and network components may be modified in increments to align with and implement VA infrastructure expansion plans.

Refer to Section 2.6.4 to Expand Genisis2 Storage, Bandwidth, and Computing in the Genisis2  [RSD](http://vhaispwww3/warboard/anotebk.asp?proj=1870).



**Figure 40: Hardware Topology**

## Software Detailed Design

As a platform for Genomic research, Genisis2 consists of Linux and Windows analysis stations and a high performance-computing cluster. Investigators shall be provided with COTS and open source tools used for genomic research. The palette of tools shall be in constant flux reflecting the rapid pace of change in genomic research.

In addition to existing analysis tools, Genisis2 introduces the Unified Genisis2 System Administrator—Researcher Dashboard, a web application described in further detail in the following sections.

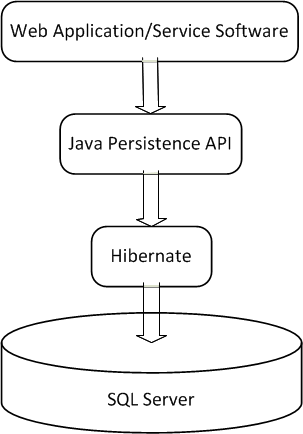
### Conceptual Design

The software design is can be divided into two categories:

* + - 1. Web Application
      2. Web Service

Each web application will follow the same basic design and components, as will web services. Web applications provide connections for web browsers to access web pages implemented in the application. Web services provide connections to custom clients and are not intended for typical user access, only other services and web applications.

Database access will be via JDBC packages that interact directly with the DBMS. Hibernate will be used to provide a higher level of abstraction for managing data stored in the Genisis database. Program-level access to database contents will be provided by the Java Persistence API (JPA) package. JPA functionality will integrate with Hibernate to create a complete path to the database. Using Hibernate and JPA helps alleviate the need to manage every detail of data access. The application can concentrate on the business requirements. These packages manage the creation and update of database tables, along with streamlining data update and retrieval. The JPA mechanisms primarily target the persistence and retrieval of Java-based objects, but JPA can also support complex queries to the database, along with calling stored procedures. These are typically known as native queries and tend to be database specific. Since Genisis does not contain time-critical applications, the performance impact will be minor.



**Figure 41: Data Access**

All database-related Java objects are defined in a common library that is available to each web application/service. These Java objects are annotated with directives that instruct the framework on how to store and retrieve the objects.

Accounts will be used to prevent unauthorized access to the Genisis database system. Each application/service must present valid credentials. These credentials are stored locally and are protected from external access. Sensitive files will be placed in a location that is unavailable to external access, except for a local system administrator login. Both web applications and services are designed to operate in standard servlet containers. For Genisis, the container will be Wildfly. The primary development language is Java.

The Wildfly server will be configured to limit access from outside of the host computer to increase security of the applications residing on Wildfly. Wildfly will only listen on the localhost address. All outside access will occur via the Apache HTTP Server. Wildfly will interact with the Apache HTTP Server via an AJP protocol-based connection. Applications and services running on Wildfly will assume that identities of all external users have been authenticated.

Wildfly will use these trusted identities to authorize access to each web page or service endpoint. Identities will be transferred using the HTTP header (REMOTE\_USER).

Web services will follow the ReST model of interfacing. ReST provides a simplified access to a web service via HTTP methods of GET, PUT, POST, and DELETE. Spring Framework MVC supports building RESTful services and associated clients. Using annotations methods are associated with HTTP methods and URLs. The Genisis web services will exchange XML documents defined in a schema.

#### Product Perspective

Each web application is built upon various open source packages. The primary package is a subset of the overall Spring Framework:

* + - * + Spring Framework
        + Spring Webflow
        + Spring Security
        + Spring LDAP

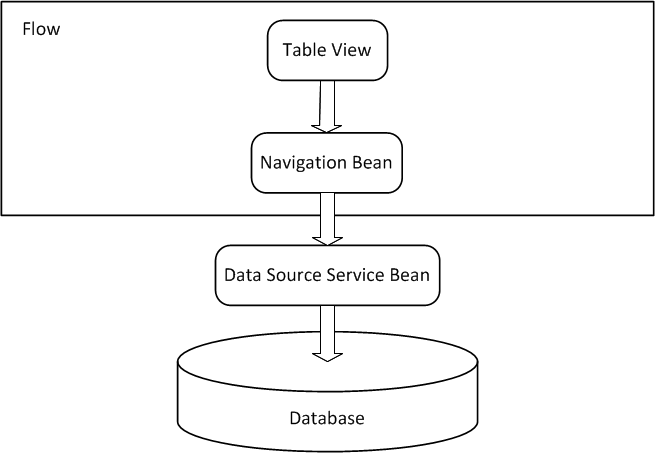
Web page implementation is provided by Java Server Faces (JSF) and Primefaces. Primefaces provides a set of widget definitions for common web application controls that either enhance existing JSF controls or implement new functionality. In addition, web pages will contain Hyper Text Markup Language (HTML) and Java Facelets constructs. Asynchronous JavaScript and XML (AJAX) will be used to enhance user interactions with the web pages. XML manipulation is provided by XMLBeans and JAXB packages, which provide XML parsing and generation based on schemas. All web applications follow the same basic architecture. External web page requests are routed through Spring’s Model View Controller (MVC) architecture to the appropriate web page, based on the uniform resource locator (URL) entered. Spring Security examines the URL to determine if the user is authorized to access the requested resource.

Security authorization will be implemented by each application for each resource it exposes. Authorization is role-based. User roles will be stored in a local Lightweight Directory Access Protocol (LDAP) server. The local LDAP server will be queried by the web application to gather all assigned roles when the user first enters the application. Access to the local LDAP is implemented using the Spring LDAP extension. The web application security will compare the user’s roles to the required role for the resource. If the user is not authorized, a page indicating the error will be sent back to the user instead of the requested resource. Each application will be designed to avoid authorization errors by only presenting options that the user is authorized to access. Menu options will be presented based on user roles. After authorizing access, Spring will direct the request to the web flow processing.

Spring Webflow is an extension to the Spring Framework that implements a state-machine architecture that manages views, which are basically web pages. Each view is associated with a flow. The flow defines the transitions between views, which guides a user through an execution of a business task. A flow may have one or more views associated with it. An application consists of one or more flows. Genisis uses Webflow to enforce transitions between views when necessary.

Generally, each web flow is backed by one or more Spring beans. The Spring beans provide the business logic behind the view and access to the data sources. The beans are typically defined in the flow configuration. When the application enters a flow, any defined beans will be created according to the rules of the flow. They will then be available to view within the workflow.

These beans exist during the life of the flow. There are other beans, such as data access interfaces, which exist during the life of the application. For example, when information is shown in a table, a “navigation” bean is defined. The navigation bean provides objects, such as data models, to support the table displayed in the view. The model contains the list of data elements shown in the view. When the data sets are potentially large, lazy loading will be implemented to enhance the user’s experience.

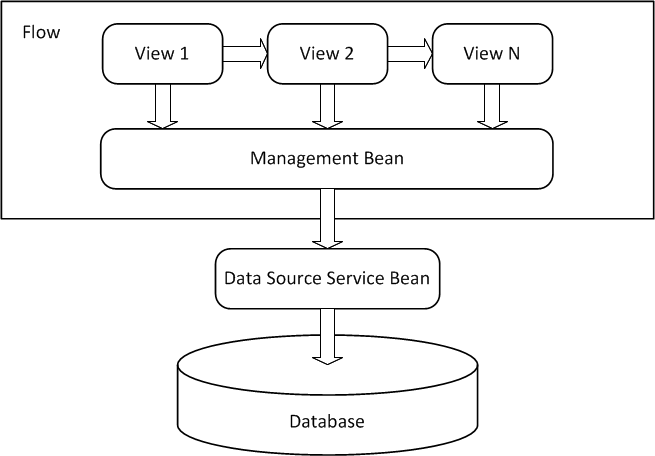


**Figure 42: Web Flow**

This is a typical architecture. For the case where data is entered, a “management” bean is

defined. Again, the purpose is to provide the business logic and data for the view or views. When business flows require more than one view, the flow defines the transitions between views.

Based upon user interactions, a view may return a new state. Figure 18 illustrates the concept. When the flow is entered, the management bean is created. The management bean lives during the entire flow. The first view shown to the user is “View 1.” After satisfying the requirements of the first view, the user presses, for example, “Next.” The view defines the “Next” button and results in a “next” state, by returning that value to the flow. The flow has defined the “next” state for “View 1” as “View 2.” Spring Webflow then presents “View 2” to the user. This pattern is followed throughout the flow. After completing the requirements for the flow, the flow is exited, and we return to a top-level flow, which is typically a landing page for the application.



**Figure 43: Wildfly Servlet**

Each flow extends a parent flow definition. The parent flow defines common transitions for all the child flows.

#### User Interfaces

Genisis2 enhances the existing user interface (UI) configuration through the new functions of the Unified Genisis2 System Administrator – Researcher Dashboard. This enhanced UI serves as a portal for both Researcher and system administrator activities. The dashboard will help Researchers request study data and track the status of their study throughout the lifecycle, and allow system administrators to monitor the Genisis2 application. Diagrams illustrating UI characteristics for the dashboard design are included the following figures in this document as of this Build.

3.3.4.1.1 Researcher Landing Page Dashboard Design

3.3.4.1.2 Genisis2 Manager/VINCI Manager Landing Page Dashboard Design

3.3.4.1.3. Genisis2 System Administrator Landing Page Dashboard Preliminary Design

3.3.4.2.1 Terminology Service Landing Page

3.3.4.2.2 Search Results

3.3.4.3.3 Concept Card Details

3.3.4.3.4 Bookmark a Concept

#### Hardware Interfaces

There are no hardware interfaces involved with Genisis2. This section is N/A.

#### Software Interfaces

Genisis2 has a Service Oriented Architecture and has an Application Programming Interface (API) that even its own Browser interface calls. Documentation for this is generated

automatically. Details on individual calls and parameters for each are documented in a separate document, **Genisis2\_VIP\_Build 5\_API\_Manual\_11032017.pdf**.

#### Communications Interfaces

There are no communication interfaces involved with Genisis. This section is N/A.

#### Memory Constraints

There are no memory constraints involved with Genisis. This section is N/A.

#### Special Operations

There are no special operations involved with Genisis. This section is N/A.

#### Product Features

Overall planned product features and those implemented in Builds 1 through 5 are covered in Sections 1.1.1 and 1.1.2 of this document.

#### User Characteristics

New requirements for Genisis2 include infrastructure expansion to support MVP, a program to enroll up to one million users of the VHA into an observational cohort. The goal is to support robust genomic and bioinformatics-related data analysis, including over 50 concurrent studies, and translating to over 100 research users and 500 computational jobs. Genisis shall serve the VA genomic research community, whose members have working experience with academic computational platforms. Specialized user support, for both genomic Researcher and Genisis2 System Administrator groups, is implemented as the Unified Genisis2 System Administrator— Researcher Dashboard requirement.

#### Dependencies and Constraints

There are no dependencies or constraints involved with Genisis. This section is N/A.

#### Database Repository

The Database Repositories and individual tables for Genisis2 are outlined in Section 5.3.

#### System Features

System Features are extensively described in sections 4 and 5 of this document – The overall goals, the workflow, status changes, user interface screens, etc.

#### Design Element Tables 6.2.1.1.13. Routines (Entry Points)

Genisis2 has a Service Oriented Architecture and has an Application Programming Interface (API) that even its own Browser interface calls. Documentation for this is generated automatically. Details on individual calls and parameters for each are documented in a separate document, **Genisis2\_VIP\_Build 5\_API\_Manual\_10032017.pdf**.

#### Templates

Genisis2 does not use any templates. This section is N/A.

#### Bulletins

Genisis2 does not use any bulletins. This section is N/A.

#### Data Entries Affected by the Design

Genisis2 does not use any variables or constants that affect design or its execution. This section is N/A.

#### Unique Record(s)

Genisis2 requirements do not use unique records in the design or execution. This section is N/A.

#### File or Global Size Changes

Genisis2 requirements do not use file or global sizes in the design or execution. This section is N/A.

#### Mail Groups

Genisis2 does not use file or global sizes in its design or its execution. It uses users VA email addresses for notifications of workflow status changes, but does not use mail groups per se. This section is N/A.

#### Security Keys

Genisis2 users their VA PIV Cards for authentication and the VA LDAP. There are no other security keys involved. This section is N/A.

#### Options

Genisis2 requirements do not use options in its design or its execution. There is an online help section that is available right next to the User Login/Profile details. That provides help regarding the User Interface. This section is N/A.

#### Protocols

Genisis2 has a Service Oriented Architecture and has an Application Programming Interface (API) that even its own Browser interface calls. Documentation for this is generated automatically. Details on individual calls and parameters for each are documented in a separate document, **Genisis2\_VIP\_Build 5\_API\_Manual\_11032017.pdf**. Protocol involved is a **RESTful API** call. It is documented in detail in this document.

#### Remote Procedure Call (RPC)

Genisis2 has a Service Oriented Architecture and has an Application Programming Interface (API) that even its own Browser interface calls. Documentation for this is generated automatically. Details on individual calls and parameters for each are documented in a separate document, **Genisis2\_VIP\_Build 5\_API\_Manual\_11032017.pdf**. Protocol involved is a **RESTful API** call. It is documented in detail in this document

#### Variables Defined in Interface

Genisis2 has a Service Oriented Architecture and has an Application Programming Interface (API) that even its own Browser interface calls. Documentation for this is generated automatically. Details on individual calls and parameters for each are documented in a separate

document, **Genisis2\_VIP\_Build 5\_API\_Manual\_11032017.pdf**. Variables involved are documented in detail in this document

#### Types Defined in Interface

Genisis2 has a Service Oriented Architecture and has an Application Programming Interface (API) that even its own Browser interface calls. Documentation for this is generated automatically. Details on individual calls and parameters for each are documented in a separate document, **Genisis2\_VIP\_Build 5\_API\_Manual\_11032017.pdf**. Types involved are documented in detail in this document

#### GUI

The Graphical User Interface elements are described in detail in the automated Java Docs generated by the environment automatically at any time. Not applicable in this section.

#### GUI Classes

The Graphical User Interface classes are described in detail in the automated Java Docs generated by the environment automatically at any time. Not applicable in this section.

#### Current Form

Forms are not used in Genisis2. This section is N/A.

#### Modified Form

Forms are not used in Genisis2. This section is N/A.

#### Components on Form

Forms are not used in Genisis2. This section is N/A.

#### Events

Events are outlined in detail in the automated Java Docs generated by the environment automatically at any time. Not applicable in this section.

#### Methods

Methods are outlined in detail in the automated Java Docs generated by the environment automatically at any time. Not applicable in this section.

#### Special References

Special References are outlined in detail in the automated Java Docs generated by the environment automatically at any time. Not applicable in this section.

#### Class Events

Class Events are outlined in detail in the automated Java Docs generated by the environment automatically at any time. Not applicable in this section.

#### Class Methods

Class Methods are outlined in detail in the automated Java Docs generated by the environment automatically at any time. Not applicable in this section.

#### Class Properties

Class Properties are outlined in detail in the automated Java Docs generated by the environment automatically at any time. Not applicable in this section.

#### Uses Clause

Uses Clauses are outlined in detail in the automated Java Docs generated by the environment automatically at any time. Not applicable in this section.

#### Forms

Forms are not used in Genisis2. This section is N/A.

#### Functions

Functions are outlined in detail in the automated Java Docs generated by the environment automatically at any time. Not applicable in this section.

#### Dialog

Dialogs are outlined in detail in the automated Java Docs generated by the environment automatically at any time. Not applicable in this section.

#### Help Frame

Help frames are outlined in detail in the automated Java Docs generated by the environment automatically at any time. Not applicable in this section.

#### HL7 Application Parameter

HL7 is not used in Genisis2 anywhere. This section is N/A.

#### HL7 Logical Link

HL7 is not used in Genisis2 anywhere. This section is N/A.

#### COTS Interface

The only COTS interface involved is that the JAVA code makes JDBC calls to the database server, Microsoft Windows 2012 R2/Microsoft SQL server 2012. This follows standard JAVA and Microsoft SQL query formats.

## Network Detailed Design

Section 4.1 and Figure 27 cover this topic.

## Security and Privacy

### Security

Genisis2 will adhere to all VA security requirements in accordance to VA Directive and Handbook 6500, Federal Information Processing Standard (FIPS) 199 and National Institute of Standards and Technology (NIST) Special Publication (SP) 800-60, recommended Security Categorization. Genisis2 is housed at the Pittsburgh Information Technology Center (PITC), and information security controls are covered under the Pittsburgh facility General Support System (GSS). The implementation of a Continuous Readiness in Information Security Program

(CRISP) is in place in order to review key security controls for regular and continuous monitoring.

The Genisis2 Security Categorization will drive the initial set of minimal security controls required for the information system. Minimum security control requirements are addressed in NIST SP 800-53, Revision 4, and VA Handbook 6500, March 2015, Appendix C: (References), Appendix E: (VA System Privacy Controls), and Appendix F: (VA System Security Controls).

For additional information, please reference Appendix C of the Genisis2 Requirements  [Specification Document (RSD).](http://vhaispwww3/warboard/anotebk.asp?proj=1870)

### Privacy

Genisis2 will adhere to all proposed VA Privacy requirements and controls, Identity Management and Security requirements, including VA Handbook 6500, March 2015, Appendix C: (References), Appendix E: (VA System Privacy Controls), NIST SP 800-53, Revision 4, Privacy Controls; any privacy or data security constraints that should be addressed in accordance with VA directives and HIPAA Privacy Act. Efforts that involve the collection and maintenance of Personal Identifiable Information (PII) must be covered by a Privacy Act system of records notice.

When a Privacy Impact Assessment (PIA) has been completed for Genisis2, a link to the PIA will be added.

## 6.5. Service Oriented Architecture / ESS Detailed Design

Genisis2 has a Service Oriented Architecture and has an Application Programming Interface (API) that even its own Browser interface calls. Documentation for this is generated automatically. Details on individual calls and parameters for each are documented in a separate document – **Genisis2\_VIP\_Build 5\_API\_Manual\_11032017.docx**.

# External System Interface Design

Genisis2 has a Service Oriented Architecture and has an Application Programming Interface (API) that even its own Browser interface calls. Documentation for this is generated automatically. Details on individual calls and parameters for each are documented in a separate document – **Genisis2\_VIP\_Build 5\_API\_Manual\_11032017.docx**.

# Attachment A – Approval Signatures

This section is used to document the approval of the System Design Document. The review should be conducted face to face where signatures can be obtained ‘live’ during the review. If unable to conduct a face-to-face meeting, then it should be held via LiveMeeting and concurrence captured during the meeting. The Scribe should add /es/name by each position cited. Example provided below.

The Business Sponsor and Project Manager are required to sign.

Sumitra Muralidhar Date:

Business Sponsor

Katie Thomas Date:

Project Manager

1. **Additional Information**

## Identification of Technology and Standards

Details of this can be found in Section 4.5. It has a table that lists the components used in Genisis2 and their TRM approval statuses.

## Constraining Policies, Directives and Procedures

Genisis2 has no additional constraining policies, directives, or procedures. This section is N/A.

## Requirements Traceability Matrix

The Genisis2 RTM will be updated with the new requirements using IBM Rational Tools.

## Packaging and Installation

**The Genisis2 Deployment Guide** contains extensive systematic instructions to set up any of the environments for this project – **DEV/TEST/SQA/PRE-PROD/PROD**.

## Design Metrics

Design metrics and approaches are outlined in Section 3.1 in this document.

### Template Revision History

|  |  |  |  |
| --- | --- | --- | --- |
| **Date** | **Revision** | **Description** | **Author** |
| June 2015 | 2.10 | Changed Heading 1 default setting to eliminate page break before | Process Management |
| May 2015 | 2.9 | Edited for Section 508 conformance and remediated with Common Look Office tool | Process Management |
| February 2015 | 2.8 | Incorporates revisions from PMAS Reform Lockdown; namely removing requirements for information that can be obtained from other PMAS authoritative sources. | Andrew Slawter, Office of Technology Strategies |
| September 2014 | 2.7 | Adds Enterprise Shared Services terms and requires AERB Compliance Certificate attachment. | Process Management |
| August 2014 | 2.6 | Signature block update authorized by AERB CR\_018934 | Process Management |
| March 2014 | 2.5 | Section 508 repairs to new version approved by AERB Chair approved | Process Management |
| August 2013 | 2.3 | Replaced the Service Architecture sub- section with new sub-sections for consumed and provided services. Also, applied miscellaneous feedback from VA team. | ASD Enterprise Shared Services (ESS) Work Group |
| June 2013 | 1.3 | Upgraded to MS Office 2007-2010 format | Process Management |
| June 2013 | 1.2 | Address inconsistencies in Section 3, Conceptual Design, Correct headings | Process Management |
| March 2013 | 1.1 | Formatted to documentation standards and edited for Section 508 conformance | Process Management |
| January 2013 | 1.0 | Initial Document | PMAS Business Office |

See TOGAF® 9.1, Part III: ADM Guidelines & Techniques, Gap Analysis on TOGAF website at <http://pubs.opengroup.org/architecture/togaf9-doc/arch/chap27.html>