

Department of Veterans Affairs

Veteran's Enterprise Management System (VEMS)

VEMS Simulated Artificial Data Environment White Paper



September 9th, 2015

John Dawson, President

Exact Data, LLC

200 Canal View Blvd., Suite 100

Rochester, NY 14623

John.dawson@exactdata.net

(585) 781-4220



VEMS Simulated Artificial Data Environment White Paper

Exact Data, LLC

200 Canal View Blvd., Suite100

Rochester, NY 14623

For profit independently NWBOC certified Women Owned Small Business

Background

VEMS is a modernization project for OSDBU (Office of Small & Disadvantaged Business Utilization) whose goal is to increase veteran contracts as mandated by congress. VEMS is a major issue today because of time involved to process activity using current workflows which is driven by a largely manual process and lack of integrated functionality. Use of ExactData manufactured artificial data to create simulated data environments within the EDE (Enterprise Development Environment) will reduce overall program risk, reduce development cost and timelines and create a higher quality end solution by immediately provisioning the correct test data for the development, integration, evaluation and testing of the new solution including interfaces without the need to access any VA internal, US Federal external or commercial databases.

The VEMS project would also be used to pilot an Enterprise implementation of simulated data environments for the VA's Enterprise Development Environment. The benefits for an Enterprise wide implementation of this technology at the VA include:

- Reduce the cost of test data creation by 90%. A typical agency has over 500 full time equivalents involved in this process yielding savings of \$50M per year.
- Reduce project development timelines by up to 50%. This directly translates to cost savings for that program that will be measured in \$100M's per year.
- Increase use case coverage and measure error rates in your development environment. Correcting errors in the development process and reducing the cost of escape errors will save your agency \$100M a year.
- Eliminate the cost and risk of managing confidential and private information in your development environments.
- Create collaborative Industry environments where challenge databases are released to the public and Industry and prototype and prove value before you purchase.

Technical Information

ExactData[™] is proposing the VA simulate the VEMS data environment – including all modes and logic workflows. This would create a large volume of realistic reference data for each database required for the VEMS solution for development and interface testing including batch processing. Simulated DevOps solutions could now be more effectively deployed to execute transactions against the simulated reference databases. ExactData integrates with standard commercial simulated DevOps products such as CA Technologies LISA[®] to create simulated data referenced to request/response pairs if required.



ExactData manufactures customized data using a sophisticated rules engine designed for the specifics of your system's requirements for realism, complexity, and scale. Our fully engineered data includes the unique features of longitudinal consistency, internal consistency, consistency across disparate data sets, and perfectly known ground truth. This enables comprehensive system performance measurement, scoring, and algorithm testing.

We generate test data with absolutely no confidentiality or privacy risks. Our patented Dynamic Data Generator[™] technology doesn't require any access to production data or live records. This makes it ideal for testing current applications and processes, as well as future systems development and testing.

This proposal will significantly reduce the cost and time to develop, test, deploy, and maintain the VEMS system. The VA will be able to procure systems and evaluate vendors for both speed and error rates, implementing Service Level Agreements for on-going best value. The VA will be able to run the entire VEMS solution before deployment, significantly reducing project risk. Confidentiality will be protected through the use of manufactured data from procurement through development to deployment. Sister agencies such as Department of Defense will not need to provide sensitive data as part of the development process.

The objectives of the proposed simulated VEMS data environment are:

- Simulate the entire VEMS data including all modes and downstream logic workflows for the projected VA population.
- Enabling Cloud Implementation/Migration through the use of non-confidential data that can be independently characterized
- Eliminate all Confidentiality/Privacy risks and the management costs associated with securing test data.
- Enable accelerated development and testing of new systems through the generation of realistic Future State VEMS data
- Enable comprehensive system performance measurement and benchmarking including error rates
- Thoroughly exercised predictive analytics and detection algorithms and then enhance to reduce error rates

The method of approach and extent of effort to be employed:

ExactData would recommend that the VA directly control the manufacture of the data and systems response files and not release this control through a subcontract to another industry vendor.

The following databases would be simulated:

VEMS Con Ops





List of Estimated Manufactured Databases

Database	Comments
VA Active Directory	Personnel Directory Information
Lexus Nexus	Credit worthiness, other personnel information
D&B	Basic business information
SAM	SAM information
VIP Data Submission	
Document Submission, Fax, Scan, Portal	Meta Data associated with the document
	submission
VEMS User Directory	Subset of VA Active Directory
User Access, Security Roles	Security access codes associated with Users
Applicant Org Listing	Simulated organization
Veteran Contact	Person and contact information
Other Contact	Person and contact information
Military History	Basic military history, enlist date, discharge date,
	citations, etc
CRM Case Information	Case number, veteran, case handler basic
	information
Task Service Request	Task information
Activity Logs: Phone, Letter, Fax, Chat, Walk In, To	Activity logs
Do	



BIRLS Data	U.S., Department of Veterans Affairs BIRLS Death
	File
DEERS/DMDC	Defense Enrollment Eligibility Reporting System

Assumptions: No Vista interface or equivalent complex system to interface, development is sequential, i.e., functionality is added one at a time

Technical Approach

The proposed technical approach is for the manufacture of realistic artificial test data on demand, versus a tradition data masking or deidentification process that can be facilitated by Extract Load Transform (ETL) technologies. Client application requirements would be gathered for the specific application being tested or developed. Current data model templates would be accessed, modified or new templates created based on these requirements. ExactData Dynamic Data Generator[™] technology would manufacture a model universe of test data designed for the specifics of the system's requirements for realism, complexity, and scale which is then applied through a data output filter that would create data outputs specific to the requirements of the Systems Under Test (SUT) for ease of ingest. (Figure 1: VEMS Manufactured Artificial Test Data Process) A Data Request and Commissioning portal would be established to provision the data to the appropriate client. (Figure 2: Request and Commissioning Portal)

Example Successful Agency Deployments

Extensive realistic simulated data environments for the US Army, IRS, DHA, and Bureau of Census. Agency wide Enterprise deployments will save \$100M's of dollars by making data processing IT procurement, development and maintenance better, faster and more efficient.

Figure 1: VEMS Manufactured Artificial Test Data Process



VEMS Manufactured Artificial Test Data Process

The engineered data includes the unique features of future state, longitudinal and internal consistency, consistency across disparate data sets, unstructured data and perfectly known ground truth. The data



has no confidentiality or privacy risks and can be shared with industry outside private/classified environments.

Figure 2: Request and Commissioning Portal

Request And Commissioning



22

The nature and extent of the anticipated results include:

- Eliminate the need for the EDE to provide external connections
- Eliminate the need to pay for access fees to external databases such as Dun & Bradstreet during the development process
- Ability to simulate the entire VEMS data environment
 including interfaces, all modes and downstream logic workflows
- Ability to run non-confidential challenges to evaluate best of breed technologies to address confidential needs
- Creation of non-confidential data that can be independently characterized, enabling Cloud Implementation/Migration
- Ability to create any current or future database to unique systems requirements without access to an existing data source
- Completely eliminate all Confidentiality/Privacy risks and the management costs associated with securing test data
- Accelerated development and testing of new systems through the generation of realistic Future State data.
- The ability to automatically create data attributes of realism, complexity, scale, internal consistency, consistency across disparate data sets, chronological consistency, and perfectly known ground truth which:
 - Enables comprehensive system performance measurement and benchmarking.



- Predictive analytics and detection algorithms can be thoroughly exercised and then enhanced to measure and reduce error rates
- Ability to perform rigorous testing with appropriate volumes of records millions instead of thousands.
- Ability to automatically manufactured artificial data records that don't lose their interconnectedness and compromise test quality like data records built from de-identified or masked production data.

This proposed simulated VEMS data environment will have a dramatic positive effect supporting the VA's mission:

- Data Quality: this proposed solution will enable the VA to measure systems error rates, at all levels with the system, in a statistically justified quantifiable manner.
- Costs: The VA will be able to make better purchasing decisions with the ability to measure cost of errors. Vendors will be more effectively managed through quantifiable SLA agreements. The proposed solution will reduce traditional Extract Transform Load (ETL) or manual data creation costs from approximately 20-30% of the total development costs to less than 5% while increasing use case coverage from below 50% to 70-80%. Using precision test data in the development process that is rapidly created will reduce overall development schedules and costs. Fewer errors will escape the development process which will significantly reduce the overall cost of enumeration.
- Privacy: Manufactured synthetic data cannot be re-identified. Confidentiality will be protected through the use of manufactured data from procurement through development to deployment. Sister agencies such as the DoD will not need to provide sensitive data as part of the development process.
- Risk: Simulating VEMS with data that mirrors the production data will significantly reduce the VA's risk by enabling the exercising of all systems and logic paths, at production level volumes and beyond. Under performing vendors can be proactively monitored and replaced before they become a critical path risk factor. DevOps environments will contain simulated data, removing a major threat vector for hacking into Census systems.

Required VA Support

The VA will need to provide supporting personnel as subject matter experts for requirements into the creation of the data model templates for the synthetic data generation process. VA will also need to help evaluate the manufactured data for feedback and corrective action. ExactData runs an Agile development process and daily participation of a VA key contact will be required for a short 15 minute daily SCRUM stand up meeting.

Pilot Project Proposal

Our pilot project proposal involves creating a simulated data environment for a subset of the VEMS current state databases to be used along with the Microsoft CRM solution pilot and verification project. The simulated data environment could be hosted on single provisioned servers or multiple unique servers at the Austin EDE, Enterprise Development Environment. ExactData would populate the databases, usually with a flat file schema, and the integrator (Microsoft CRM) would manage use of the databases for development, test and evaluation including any external and internal interfaces



The pilot would involve generating simulated Veterans, SAM, BIRLS, MVI (Master Veterans Index) and DEERS databases. Other databases could also be simulated upon review and mutual agreement. (Figure 3: VEMS EDE Simulated Data Pilot)



Figure 3: VEMS EDE Simulated Data Pilot

The period of performance would be approximately 4 months starting in the first of November timeframe. Initial dataset would be provided after month one with data models evolving and finalizing over the 3 month period. (Figure 4: VEMS Manufactured Artificial Test Data Process Pilot)



VEMS Manufactured Artificial Test Data Process Pilot



Figure 4: VEMS Manufactured Artificial Test Data Process Pilot

Pilot Project Objectives:

- Validate a Development, Test and Evaluation environment, based on manufactured artificial test data can be established within the EDE. Validate establishment of this capability in the EDE would benefit all of the VA.
- Validate that connections can be simulated for internal and external interfaces for test and evaluation before deployment
- Validate that the simulated data environment can be used by numerous other VA development projects such as those projects currently requiring access to BIRLS data.
- Validate confidentiality and privacy risk and costs can be eliminated within the development environment through the use of simulated data within a DT&E of the EDE
- Validate overall program and schedule risk will be reduced through the use of simulated data within a DT&E of the EDE
- Quantify the impact on compressing development timelines for the larger VEMS project, extrapolating to all VA projects.
- Quantify the cost savings associated with automating the test data and use case creation process for the larger VEMS project, extrapolating to all VA projects.
- Quantify the cost savings associated with reduced development errors escaping to the production systems through the use of simulated data environments for the larger VEMS project, extrapolating to all VA projects.

Pilot Project Savings Estimates that will be validated with the Project:

• Total pilot project Microsoft CRM Development costs are estimated at \$700,000.00



- Reductions in test data creation costs from standard ETL or manual creation processes are estimated to reduce the overall project costs by 10% or \$70,000.00 in savings
- Development timelines are estimated to be compressed by 10% for \$70,000.00 in savings
- Reductions in Cost of errors that are corrected in the development versus production systems are estimated at another 10% and \$70,000.00 in savings
- Confidentiality and privacy risk and costs would be eliminated within the development environment
- Overall program and schedule risk will be reduced

Similar savings amounts for the entire implementation, estimated at \$20M in costs, would yield cost savings of \$6M for the entire project.

ExactData's Previous Experience and Relevant Past Performance

ExactData's clients are primarily US Federal Agencies in the Healthcare, Financial Services, Population Statistics, and Intelligence application spaces. ExactData generates data customized for the specific systems under test. Privacy risks and costs associated with the data were eliminated. Application development schedules were always met or exceeded with significantly lower costs than with traditional test data creation methods. Clients and integrators involved in these projects include: Department of Defense, IRS, DARPA, U.S. Census Bureau, The Office for National Statistics, United Kingdom, Statistics Canada, and Major Federal Systems Integrators including Lockheed Martin and IBM. ExactData is a Unisys and Lockheed Martin Mentor Protégé Company. ExactData with team Leidos was just awarded a \$4.3B contract to modernization the Department of Defense Electronic Healthcare Record system.