Overview of Data & Analytics at the Department of Veterans Affairs

Jack Bates
VA Office of Information & Technology

August, 2016
Veterans Health Administration (VHA)

The Largest Integrated Healthcare System in the Country

VHA Points of Care (1,748)
- Integrated Healthcare Networks: 21
- Major Medical Centers: 152
- Outpatient Clinics: 990
- Vet Centers: 370
- Domicillaries: 102
- Community Living Centers: 134

Patient Population
- Enrollees: 8.8M
- Active Patients: 6M
- All Time Patients: 22M
- FY15 Outpatient Visits: 84M
- FY15 Inpatient Admissions: 703K
VA Electronic Health Record (EHR)

EHR – Provider Generated Data

- **VistA (Veterans Health Information Systems and Technology Architecture)**
  - An integrated Electronic Health Record (EHR) data system with application modules that share a common data store and common internal services
  - There are **130 hospital based instances of VistA**
  - There are approximately 200 modules comprising VistA
    - Clinical Modules (e.g. Lab, Pharmacy, Radiology, Vital Signs)
    - Administrative Modules (e.g. Enrollment, Scheduling, Admission-Discharge-Transfer, Outpatient Encounters)
    - Financial Modules (e.g. Payroll, Beneficiary Travel, Fee Basis)

**Horizon:**
- PGD – Patient Generated Data (Mobile, Web, Kiosk)
- MGD – Machine Generated Data (Connected Devices, IoT)
CDW System Facts:
- Source system:
  - VISTA EHR: 130
  - VISTA Admin: 7
- Extract tools:
  - VISTA NRT Journal Reader
  - VISTA Batch Extractor
- Data facts:
  - Domains of information: 68
  - Rows of data: 2 trillion
  - Columns of data: 30,000+
  - Tables of data: 900+
- Data quality program
- Active Users: 80,000/Month
- Vibrant user community

CDW Sample Data Facts:
- Unique Veterans: 20 million
- Outpatient encounters: 1.6 billion
- Inpatient admissions: 9 million
- Clinical orders: 3.2 billion
- Lab tests: 5.6 billion
- Pharmacy fills: 1.5 billion
- Radiology procedures: 162 million
- Vital signs: 2.3 billion
- Text notes: 2.0 billion

CDW Analytic Enclaves:
- GP: General Purpose
- BI: Business Intelligence
- AN: Analytics and Informatics
- RD: Health Services R&D (VINCI)
- FR – Field Reporting

CDW Analytic Capabilities:
- Primary/Secondary/Data Mart Structures
- Data Standardization
- Metadata Services
- Business Intelligence Reporting & Dashboards Tools
- Geospatial Mapping Tools and Images
- SAS/Grid High Performance Compute Grid
- Natural Language Processing Engines
- Hadoop Cluster

VA Analytic Ecosystem
Common Data ♦ Common Infrastructure ♦ Common Tools ♦ Common Security
Corporate Data Warehouse (CDW) 
Facts & Figures

- (>16 years) Longitudinal data on >22.3M veterans
  - > 200 million rows of data/day → 2.4 trillion - (80% of all VA data, doubling annually)
- Business: Accounting (AP/AR/GL/MCA), Acquisition/Logistics, Assets, HR, Emergency Management,
- Clinical: VistA (100%), EDIS, MVP Genomic
- Veteran: Patient generated, Enrollment, Demographics, Telephony, CRM, Survey
- 80,000 users throughout VA
  - >650 user-driven analytic solutions
  - 125,000 reports and 2,500 dashboards
  - 125 training events in FY15
    - e.g. BI, Analytics, Data Mgmt, Best Practices
- Community Collaboration/Crowdsourcing
  - Subject Matter Expertise
  - Sharing of lessons learned & best practices
  - Data sharing (e.g. Data Object Exchange)
CDW EHR Based Data Examples

<table>
<thead>
<tr>
<th>Patients: 22 M/16 years</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lab Results</td>
</tr>
<tr>
<td>7.7B</td>
</tr>
<tr>
<td>Pharmacy Fills</td>
</tr>
<tr>
<td>2.2B</td>
</tr>
<tr>
<td>Radiology Proc</td>
</tr>
<tr>
<td>202 M</td>
</tr>
<tr>
<td>Appointments</td>
</tr>
<tr>
<td>1.4B</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Production</th>
</tr>
</thead>
<tbody>
<tr>
<td>Allergy</td>
</tr>
<tr>
<td>Appointment</td>
</tr>
<tr>
<td>Consult</td>
</tr>
<tr>
<td>CPRS Orders</td>
</tr>
<tr>
<td>Data Profiling</td>
</tr>
<tr>
<td>Dental</td>
</tr>
<tr>
<td>Health Factors</td>
</tr>
<tr>
<td>Immunization</td>
</tr>
<tr>
<td>Inpatient</td>
</tr>
<tr>
<td>Lab-Microbiology</td>
</tr>
<tr>
<td>LabChem</td>
</tr>
<tr>
<td>Mental Health Assessment</td>
</tr>
<tr>
<td>Non-VA Meds</td>
</tr>
<tr>
<td>Outpatient</td>
</tr>
<tr>
<td>Patient</td>
</tr>
<tr>
<td>Patient Associated</td>
</tr>
<tr>
<td>Patient Enrollment</td>
</tr>
<tr>
<td>PCMM (Primary Care Management Module)</td>
</tr>
<tr>
<td>Pharmacy/BMC (Bar Code Medication Administration)</td>
</tr>
<tr>
<td>Pharmacy Outpatient</td>
</tr>
<tr>
<td>Pharmacy Patient</td>
</tr>
<tr>
<td>Purchased Care (formerly fee)</td>
</tr>
<tr>
<td>Pyramid IIS</td>
</tr>
<tr>
<td>Recall Reminders</td>
</tr>
<tr>
<td>S Patient</td>
</tr>
<tr>
<td>Staff</td>
</tr>
<tr>
<td>STaff</td>
</tr>
<tr>
<td>Surgery FRE Table (Not the Surgery Raw domain)</td>
</tr>
<tr>
<td>Vista Compensation &amp; Pension</td>
</tr>
<tr>
<td>Vista Watlist</td>
</tr>
<tr>
<td>Vital Signs</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>RAW*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bill Claims (CBO)</td>
</tr>
<tr>
<td>CAPRI Audit Trail table</td>
</tr>
<tr>
<td>ClinicComp</td>
</tr>
<tr>
<td>Echocardiogram</td>
</tr>
<tr>
<td>Emergency Dept. Int. Software (EDIS)</td>
</tr>
<tr>
<td>Equipment Inventory</td>
</tr>
<tr>
<td>FBCL (Fee Basis Claim System)</td>
</tr>
<tr>
<td>FFCAP (Integrated Funds Control, Accounting, and Procurement)</td>
</tr>
<tr>
<td>Inpatientmed (IV)</td>
</tr>
<tr>
<td>Oncology</td>
</tr>
<tr>
<td>PAID (Personnel and Accounting Integrated Data System)</td>
</tr>
<tr>
<td>Prosthetics</td>
</tr>
<tr>
<td>Pulmonary Function Test (PFT)</td>
</tr>
<tr>
<td>Radiology</td>
</tr>
<tr>
<td>RxUD (Unit Dose)</td>
</tr>
<tr>
<td>Surgery</td>
</tr>
<tr>
<td>Travel</td>
</tr>
<tr>
<td>VAGAA (Veterans Choice Program Eligibility)</td>
</tr>
</tbody>
</table>
CDW Use Case Examples

- **Operational Business Intelligence**
  - Access Management
  - Resource Management
  - Electronic Clinical Quality Measures

- **Operational Analytics**
  - Hospital Readmission Predictive Models
  - Patient Fall Predictive Models

- **Point of Care Decision Support**
  - Panel Management
  - Disease or Condition Management

- **Population Health & Surveillance**
  - Hospital Acquired Infections
  - Biosurveillance

- **Health Services Research**
  - Research on patient care, care delivery, health outcomes, cost, and quality
Use Case: Situational Awareness/Disaster Response
Twin Hurricanes Lester (C4) and Madeline (C1)
Operational Business Intelligence Use Case: Veteran Choice Act Support

• **Data/Infrastructure**
  – VACAA Patient Geocoding and Drive Times
  – VACAA Choice Patient Wait Times
  – VACAA Patient Eligibility
    • Registry of VACAA Eligible Patients and related demographics per CBO
  – Fee Basis Claims (FBCS) for Community Care

• **Reports/Tools**
  – Drive Time Analysis
  – Community Provider Lookup
  – VISN/Facility Choice Patient Eligibility
  – Patient Wait Times
Predictive Analytics Use Case: Care Assessment Need Score

- Complex predictive model - 40 variables incl. demographics, diagnoses, lab, vitals, meds, health care utilization, DoD data, census data
- Highly accurate predictions of death and hospitalization for all VA primary care pts (~6M)
- Updated weekly and available to all primary care providers via VistA
- Tightly linked with web-based care coordination software – Patient Care Assessment System

Veterans in highest %ile of risk have 58% probability of admission, 23% probability of death, and 64% probability of either event.
Use Case: Health Services Research

Project Examples:
- Collaborative for Improving Hypertension Management
- Primary Care Quality and Service Customization of the Homeless
- Using Knowledge Discovery Strategies to Identify Fall-Related Injuries
- Post-Deployment Sentinel Event Surveillance
- Monitoring and Management of Metabolic Side Effects of Antipsychotics
- In-Theater Medical Treatment and Long-Term Health Outcomes of Recent Military Amputees
- Outcomes, Costs and Trends in Dialysis Timing in VA
- PTSD, Depression and Substance Use Disorders among U.S. Veterans Returning from Iraq and Afghanistan
- Traumatic Brain Injury among Homeless Veterans

FY15 Statistics:
- 5,547 Users (50% increase)
- 871 Projects (30% increase)

Capabilities:
- National Scale Data
- Prep to Research Tools
- Data Support
- SAS/Grid for Statistical Processing
Precision Medicine/Genomics Use Case: MVP-CHAMPION

MVP-CHAMPION

VA Million Veteran Program (MVP)

Computational Health Analytics for Medical Precision to Improve Outcomes Now
MVP-CHAMPION: The Vision

1. To improve the lives and well-being of our Nations veterans by discovering new or improved treatment regimens for diseases and conditions that are prevalent in the Veteran community.
2. To usher in a new era of personalized and precision medicine that will change the science and practice of medicine.
3. To make quantum leaps in the knowledge and treatment of our most prevalent diseases and conditions.
4. To advance the state of the art in data science, informatics, and analytics in medicine.
5. To build a sustainable data science workforce that have the skill to tackle the most complex problems in medicine.

<table>
<thead>
<tr>
<th>Data</th>
<th>Actions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Symptom-based</td>
<td>Intuition-based Medicine</td>
</tr>
<tr>
<td>Pattern-based</td>
<td>Evidence-based Medicine</td>
</tr>
<tr>
<td>Algorithm-aided</td>
<td>Precision-Medicine</td>
</tr>
</tbody>
</table>
MVP-CHAMPION *FastTrack* Goal:

Make a discovery that advances the state of the art in prostate cancer treatment within 18-months

**VA Cases of Common Cancers FY2005-FY16**

As of 3/16/2016

<table>
<thead>
<tr>
<th>Rank</th>
<th>ICD-9</th>
<th>Common Cancers</th>
<th>VA Cases (Patients)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>185</td>
<td>Malignant neoplasm of prostate</td>
<td>362,256</td>
</tr>
<tr>
<td>2.</td>
<td>162</td>
<td>Malignant neoplasm of trachea, bronchus, and lung</td>
<td>105,731</td>
</tr>
<tr>
<td>3.</td>
<td>153</td>
<td>Malignant neoplasm of colon</td>
<td>61,797</td>
</tr>
<tr>
<td>4.</td>
<td>188</td>
<td>Malignant neoplasm of bladder</td>
<td>61,262</td>
</tr>
<tr>
<td>5.</td>
<td>200-202</td>
<td>Lymphoma</td>
<td>40,165</td>
</tr>
<tr>
<td>6.</td>
<td>198</td>
<td>Secondary malignant neoplasm of other specified sites</td>
<td>35,141</td>
</tr>
<tr>
<td>7.</td>
<td>189</td>
<td>Malignant neoplasm of kidney and other and unspecified urinary organs</td>
<td>32,694</td>
</tr>
<tr>
<td>8.</td>
<td>204-208</td>
<td>Leukemia</td>
<td>31,354</td>
</tr>
<tr>
<td>9.</td>
<td>172</td>
<td>Malignant melanoma of skin</td>
<td>27,184</td>
</tr>
<tr>
<td>10.</td>
<td>197</td>
<td>Secondary malignant neoplasm of respiratory and digestive systems</td>
<td>20,058</td>
</tr>
</tbody>
</table>

**Prostate Cancer: The Paradox of Choice**

- Prostate Cancer is a difficult and complex disease, and the choices facing a man who is diagnosed with it — what treatment to have, indeed whether to be treated at all — are complex, perhaps more complex than in any other major cancer.

- The goal of this project is within 18-months to make significant progress in being able to categorize prostate cancer patients based upon genetic factors to be able to be more precise (fewer, more precise) options in their treatment and treatment regimens.
Methodological Challenges

• How to scale “deep” phenotyping to large data sets (how-to use data to characterize):
  • Stage
  • Severity
  • Social context
  • Disease course
  • Treatment response
  • Differential diagnosis

• Need for NLP plus structured data!

• Need to integrate phenotype and genotype data
  – Scale of genotype data
Big Data Challenges

- **Things we want to do but haven’t**
  - Analytic work flow “factory” (democratize analytics)
  - Real-Time Point of Care Decision Support (“Amazon” like patient care “suggestions”)
  - Mainstream genomic data into analytic pipelines

- **Things we want to do but can’t**
  - Open source community collaboration with VA data
  - Population scale Health Services Research
  - Rapid analytic model prototyping

- **Things we are doing and need to do better**
  - 360° View of the Veteran (all interactions)
  - Advanced analytics education and training of workforce (Data, Technique, Tools)
  - New generation computing platforms (processor, memory, storage, networking)
  - Information protection (Rogue States and individuals, Ransomware)
Toward a Veteran Centric “Big Data” Capability

VA Line of Business Data

Benefits
Benefits Data Warehouse
(Benefits, Education, Loan Data)

Health
Health Data Warehouse
(Healthcare, Phenome, Genome Data)
(Patient Generated, Machine Generated)

Cemetery
Cemetery Data Warehouse
(Internment and Location Data)

360° Analytic View of the Veteran (and the VA)

VA Data, The Whole Picture

Analytic Capabilities:
• Descriptive
• Diagnostic
• Predictive
• Geospatial
• Natural Language Processing

High Performance Computing (DoE Oak Ridge National Lab)
(Infrastructure, Knowledge, Skill, AoA)

Cloud Computing (IaaS, PaaS, SaaS)
(Infrastructure)

Customer Master Data
• Master Veteran Demographics

Customer Experience Data
• Customer Sentiment
• Customer Relationship Management
• Telephony
• Vets.gov

VA Business Data
• Accounting (AP/AR/GL/MCA)
• Acquisition
• Assets (Facilities/Equipment)
• Human Resources/Payroll
• Emergency Management
• Information Technology

Non VA Data
• Department of Defense
• Department of Homeland Security
• Department of Energy
• Department of Commerce
• Health & Human Services
• Community Care
• Lenders, Schools
• Private Sector

CDW

Infrastructure

Customer Master Data

Customer Experience Data

VA Business Data

Non VA Data

Benefits

Health

Cemetery
Conclusion

VA’s goal is to leverage our vast array of clinical and genomic data to:

1. **Improve the lives and well-being of our Nations veterans** by discovering new or improved treatment regimens for diseases and conditions that are prevalent in the Veteran community.

2. **Usher in a new era of personalized and precision medicine** that will change the science and practice of medicine.

3. **Make quantum leaps** in the knowledge and treatment of our most prevalent diseases and conditions.

4. **Advance the state of the art in data science, informatics, and analytics in medicine.**

5. **Build a sustainable data science workforce** that have the skill to tackle the most complex problems in medicine.

6. **All of the above will NOT be possible without a strong partnership with the Department of Energy** and their domain expertise in high performance computing and big data analytics.