Leveraging Clinical Data in the VHIE to Build a Clinical Data Warehouse

Mike Gagnon
VITL CTO
Background

• VITL collects clinical data from many VT healthcare organizations today as part of regular VHIE operations
• Over 4M clinical data messages per month now being processed
• Data includes patient demographics, patient events, labs, transcribed reports, medications, immunizations and care summaries
Background (cont’d)

• The data collected is used for:
  – Patient identification (MPI)
  – Clinical data at the point of care (VITLAccess)
  – Processing transactions (lab orders, results, immunizations)
  – Populating state registries (Blueprint and VDH)
  – Supporting ACOs clinical data needs
Why is a Clinical Data Warehouse Important?

- Aligns with the scope of our State Innovation Model grant
- Strong support from the Blueprint, state agencies, ACOs, hospitals, providers, and others to improve the “quality” of data coming from providers for use in reporting and analytics
  - Data for reporting needs to be complete, accurate and consistent
  - Our experience with the Blueprint identified issues with data reporting
- Would provide a similar capability for clinical data that is currently provided by VHCUREs for claims data
- Would be a step forward in combining clinical and claims data for analysis
Alignment with SIM (VHCIP)

This capability was clearly defined in the VHCIP Scope:

Technology and Infrastructure: Integrated Platform & Reporting System Scope

- Development of an integration repository;
- Development of data management, normalization, person and provider identification, and data merging capabilities;
- Development of reporting portals and dashboards;
- Incorporation of refined analytic methods, algorithms, and reporting formats
Possible Use Cases

• Blueprint data normalization and data quality improvements
• ACO data normalization and quality improvements for all 3 organizations
• Behavioral Health Network
• Regional Lab Testing Analysis
• Reporting for organizations or groups connected to the VHIE
Data Flow and Terms

• Data flows from Source Organizations to the VHIE to the Warehouse and then to downstream systems for analysis
  – Data can be cleaned and improved at each step in the process
• Data Quality is a general term which describes data completeness, accuracy and consistency
• Data Normalization is the process of mapping and converting local terms to common standard terms (i.e. lab test HgA1C to 55454-3)
• Data Marts are subsets of the larger warehouse which can be quickly created to perform specific analytics
Warehouse Concept

- Use clinical data already collected In VHIE
  - Extract this data to create the warehouse
  - Build the data model
  - Capture live data to keep warehouse current
- Analyze the data for quality and perform “cleansing”
- Perform data normalization to map terms to standards
- From this core warehouse create smaller data sets (marts) for analysis
  - Send these data sets to participant organizations to perform their own analysis, or
  - Provide data reporting and analytics tools for organizations to use
Details on the Concepts

- Use clinical data already collected
  - The data collected in the VHIE is not in a form that is ready for reporting and analysis. We need to develop the warehouse for this purpose
- Analyze the data for quality and perform “cleansing”
  - The data then needs to be analyzed for completeness, accuracy and consistency
  - We can then provide a “report card” on data quality for each organization
- Perform data normalization to map terms to standards
- From this core warehouse create smaller data sets (marts) for analysis
Data Quality Analysis

• VITL has selected SAS as its core data warehouse tool
• SAS is a highly scalable platform for data management, reporting, statistical analysis and analytics
• One key feature of the SAS product is its data quality analysis capability
Determine people, processes, technologies and data sources
- Acceptable results
- Build a roadmap
- Explore metadata
- Profile and validate data
- Build workflows, jobs and business rules
- Create data model
- Integrate DQ services
- Reuse DQ services across enterprise
- Refine and adjust business rules and definitions
- Create reports with DQ KPIs
- Examine trends
- Archive data not in use
- Communicate next steps and success
- Explore metadata
- Profile and validate data
- Build workflows, jobs and business rules
- Create data model
- Evaluate
- Control
- Design
- Discover
- Plan
SAS Data Quality

**IMPROVE - PARSE**

- Convert “semi-structured“ data into structured data
- Identify attributes within a string of data
- Supports customer Information, product data and others
- Locale sensitive

nathan smith, CEO; n-smith@windowsoft.com 219 w 49th street New York, ny 10019, 6th floor; 212.541.8457, windowsoft Svcs Inc.
SAS Data Quality

**IMPROVE - STANDARDIZE**

- Create a consistent representation of attributes throughout the organization
- Standardize data attributes in batch and real-time
- Upper case, lower case, mixed case, abbreviations, number patterns
- All data types
  - Person names, Legal entities, organizations, titles, address information, phone numbers, SSO, material and product names, codes,…
- Locale sensitive
- Custom standardizations

### First Name: Nathan
### Last Name: Smith
### Street: 219 West 49th Street
### 2nd Address line: 6th floor
### City: New York
### Zip Code: 10019
### State: NY
### Phone: +1 (212) 541.8457
### Position: CEO
### Company: Windowsoft Services Incorporated
SAS Data Quality

- Identify Data Quality issues in your data
- Data Profiling
  - Intuitive and easy to use
  - Rich set of standard profiling metrics
  - Define individual business rules for custom quality checks
  - Tabular and graphical profiling results
  - Print and drill
  - “Drill down” to data record
- Field relationship maps
- Redundancy analysis
- Analyze root cause of data issues
SAS Data Quality

**IMPROVE - ENRICHMENT**

- Enrich data with missing information to fit new and upcoming analytical and operational needs
  - To ensure address attributes are reliably
- Postal validation for over 240 countries
  - Gender:
  - Add geocoding information to address
- Validate and append other attributes like
  - Gender
  - Product classification like e-C@ss; UNSPSC

<table>
<thead>
<tr>
<th>First Name:</th>
<th>Nathan</th>
</tr>
</thead>
<tbody>
<tr>
<td>Last Name:</td>
<td>Smith</td>
</tr>
<tr>
<td>Street:</td>
<td>219 West 49th Street</td>
</tr>
<tr>
<td>City:</td>
<td>New York</td>
</tr>
<tr>
<td>Zip Code:</td>
<td>10019</td>
</tr>
<tr>
<td>State:</td>
<td>NY</td>
</tr>
<tr>
<td>Phone:</td>
<td>+1 (212) 541.8457</td>
</tr>
<tr>
<td>Position:</td>
<td>CEO</td>
</tr>
<tr>
<td>Company:</td>
<td>Windowsoft Services Inc.</td>
</tr>
<tr>
<td>Segment:</td>
<td>Premium</td>
</tr>
</tbody>
</table>
SAS Data Quality

**IMPROVE - MATCH AND MERGE**

- Eliminate unneeded and outdated records to increase trust in data
- Identify duplicate data records using fault-tolerant matching
  - Locale sensitive
  - Matching & De-Duplication of data records
  - All types of data (customer, product,....)
- Merge attributes to build “Golden Record” through
  - Automated process
  - Manually based on system generated suggestions
  - Household identification
- Fault tolerant comparison against “blacklists” & “whitelists”
Components of the Warehouse

- Core database
- Messaging engine
- VITL Staff
- Terminology mapping (normalization) services
- Data warehouse with data quality services
- Reporting modules
- Tools for Analytics
- Hardware
Future Phase 2 Capabilities

• **Scope (shorter term)**
  – Build data marts for specific customers
  – Combined claims-clinical MPI
  – Pilot projects for claims-clinical analysis

• **Scope (longer term)**
  – Claims–clinical data integration
  – Predictive analytics for disease monitoring and progression
Mike Gagnon
CTO, VITL
mgagnon@vitl.net
Questions?