Virginia Longitudinal Data System (VLDS) – Overview

Social Policy Research Associates (SPRA)

Today’s Objectives
• Answer your questions
• Provide an overview of VLDS, to include the primary business process and governance for requesting and releasing data
• Review Agency and Researcher roles
• Explain the security built into VLDS

Part I – VLDS Overview
Current Partners

- VDOE – K-12 student data
- SCHEV – Student data for two and four year higher education institutions
- VEC – Wage, Trade Act Assoc., Wagner-Pyser, and Unemployment Insurance Benefits data
- VCCS – Workforce Investment Act data

What it took...

- Strong personal and professional alliances between the founding agencies
- Establishing cross-agency governance procedures
- Everyone’s commitment (time and will) to make it happen

VLDS is...

- A web portal to access education and workforce data, which
- Retrieves and matches data from agency databases, while
- Maintaining the confidentiality of individual student/teacher/employee data, to be
- Used for reporting and analytic purposes.
High Level Component Overview

- Data resides within each agency
- Agencies control access to the data using customer Management software (aka Workflow)
- Researchers request and access VLDS data using the VLDS Portal

Part II – The Business Process

VLDS Process Overview
Notifications

Emails are sent to your inbox for everything requiring an action

- Emails (Action & Info) for:
  - Portal Accounts
  - Research Purpose
  - Team Members
  - RUDA
  - Amendments
  - Data Packages
  - Data Results
  - Documents
  - Closing

VLDS Portal Symbols

- Information that CAN be accessed at this time
- Information that CANNOT be accessed at this time
- This step is WAITING on APPROVAL.
- This step is currently APPROVED and COMPLETE

[Image of VLDS Portal Symbols]

[Image of VLDS Portal Interface]
Show me the demo!
Part III - Roles & Security

Built-in Security Features

- Portal access is progressively permitted
- Agencies control data release through RUDA and Data Package approval
- NDA must be on file to download data results
- Data Results do not contain Personally Identifiable Information (PII), e.g. name, address, SSN, and STI
- Random unique identifiers are created for each Data Package
- Data Results removed from the VLDS server after 10 days
Getting Data Ready for “De-Identified Federation”

How De-Identification Works in VLDS System

Agency 1 Exposure Database

VLDS Shaker

Agency 2 Exposure Database

The Data Adapters prepare the data (including de-identification) before leaving the agency firewall.

Agency Firewalls

Privacy Protection

• Two step process

At the data adapter (Agency)

1. De-identify personal information

At the Shaker (VLDS)

2. Further protect it by creating random identifiers

Step 1 is re-run every 90 days. Step 2 occurs with each data request.

An Example of De-Identification...

• Data Adapter De-identifies
  • De’Smith-Barney IV

  Data scrambling

  C57S78XCEBF9WECP2AA9DK59N1CO27QBES54HFD

  Occurs every 90 Days

• Shaker Double De-identifies

  Double De-identify (VLDS-Shaker)

  Ready for transport to the researcher

  C57S78XCEBF9WECP2AA9DK59N1CO27QBES54HFD

  Occurs with every data request
Conclusion

- VLDS is a multi-component application which supports agency and researcher business processes by managing access to double de-identified cross-agency data in a secure and private environment.

The End

Back Up Slides
VLDS Components

- Portal
- Security
- Data Dictionary
- Data Dictionary Tool
- Data Request
- Workflow
- Shaker / Data Adaptor
- Agency Data

VLDS Application Architecture

Portal Features

- MyVLDS
  - Team member management
  - Research information management (Who, What, Where, When, Why)
  - Data dictionary & selection tool
  - Data request and retrieval
  - Document management (NDAs, research papers, etc.)
  - Ability to check status, modify or cancel account and/or data request
  - Help / Training
  - Password reset
Workflow Features

- **Automated** Email Notifications and Tasks
- Manage Contacts (Researchers)
- Create / Edit / Approve / Disapprove:
  - Research Purpose
  - Restricted Use Data Agreement (RUDA)
  - Data Package
  - Artifact (Documents)
  - Non-Disclosure Agreement (NDA)
- Document **Storage**
- Audit Logs
- **Integration** with Microsoft Outlook

Roles

- **Sponsor Agency**
  - Creates Portal Account
  - Approves Research Purpose, Restricted Use Data Agreement (RUDA), Research Team Members, Data Packages, Artifacts, and Closes Research Purpose
- **Partner Agency**
  - Approve RUDA and Team Members
  - Notified regarding Data Packages and Artifacts
- **Principal Project Officer (PPO)**
  - Submits Research Purpose, RUDA and Amendments
- **Researchers**
  - Provides General info, creates Data Requests and Data Packages

An Example of De-Identification...

- **What the Data Adapter Does:**
  - DeSmith-Barney
  - Uses Substitution Cipher
  - Hashing Key – Dynamically Generated every 90 days and sent from Matching Engine
  - CLEANED & ENCODED for TRANSPORT
A different way to look at it...

### Agency Data

<table>
<thead>
<tr>
<th>INT_ID</th>
<th>FIRST_NAME</th>
<th>MIDDLE_NAMES</th>
<th>LAST_NAME</th>
<th>GENDER</th>
<th>DOB_MONTH</th>
<th>DOB_DAY</th>
<th>DOB_YEAR</th>
<th>RECORD_DATE</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>John</td>
<td>Tyler</td>
<td>Smith</td>
<td>M</td>
<td>12</td>
<td>23</td>
<td>1990</td>
<td>6/1/2011</td>
</tr>
<tr>
<td>A</td>
<td>John</td>
<td>Tyler</td>
<td>Smith</td>
<td>M</td>
<td>12</td>
<td>23</td>
<td>1990</td>
<td>7/3/2012</td>
</tr>
<tr>
<td>A</td>
<td>John</td>
<td>Tyler</td>
<td>Smith</td>
<td>M</td>
<td>11</td>
<td>23</td>
<td>1990</td>
<td>8/5/2013</td>
</tr>
<tr>
<td>B</td>
<td>Bill</td>
<td>Matthew</td>
<td>Rogers</td>
<td>M</td>
<td>5</td>
<td>10</td>
<td>1991</td>
<td>7/3/2012</td>
</tr>
<tr>
<td>B</td>
<td>Bill</td>
<td>Matthew</td>
<td>Rogers</td>
<td>M</td>
<td>5</td>
<td>10</td>
<td>1991</td>
<td>8/5/2013</td>
</tr>
<tr>
<td>C</td>
<td>Sara</td>
<td>Crane</td>
<td>Parker</td>
<td>F</td>
<td>8</td>
<td>6</td>
<td>1990</td>
<td>6/1/2011</td>
</tr>
<tr>
<td>C</td>
<td>Sara</td>
<td>Crane</td>
<td>Parker</td>
<td>F</td>
<td>8</td>
<td>6</td>
<td>1990</td>
<td>7/3/2012</td>
</tr>
<tr>
<td>C</td>
<td>Sara</td>
<td>Crane</td>
<td>Parker</td>
<td>F</td>
<td>8</td>
<td>6</td>
<td>1990</td>
<td>8/5/2013</td>
</tr>
<tr>
<td>D</td>
<td>Mary</td>
<td>Rachel</td>
<td>Moore</td>
<td>F</td>
<td>10</td>
<td>19</td>
<td>1990</td>
<td>6/1/2011</td>
</tr>
<tr>
<td>D</td>
<td>Mary</td>
<td>Rachel</td>
<td>Moore</td>
<td>F</td>
<td>10</td>
<td>19</td>
<td>1990</td>
<td>7/3/2012</td>
</tr>
<tr>
<td>D</td>
<td>Mary</td>
<td>Rachel</td>
<td>Moore</td>
<td>F</td>
<td>10</td>
<td>19</td>
<td>1990</td>
<td>8/5/2013</td>
</tr>
</tbody>
</table>

Data Adapter Reduction (Done at the Agency server)

<table>
<thead>
<tr>
<th>INT_ID</th>
<th>UEID</th>
<th>FIRST_NAME</th>
<th>MIDDLE_NAMES</th>
<th>LAST_NAME</th>
<th>GENDER</th>
<th>DOB_MONTH</th>
<th>DOB_DAY</th>
<th>DOB_YEAR</th>
<th>RECORD_DATE</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>A395825930</td>
<td>John</td>
<td>Tyler</td>
<td>Smith</td>
<td>M</td>
<td>12</td>
<td>23</td>
<td>1990</td>
<td>7/3/2012</td>
</tr>
<tr>
<td>B</td>
<td>B402924567</td>
<td>Bill</td>
<td>Matthew</td>
<td>Rogers</td>
<td>M</td>
<td>5</td>
<td>10</td>
<td>1991</td>
<td>8/5/2013</td>
</tr>
<tr>
<td>C</td>
<td>C940293847</td>
<td>Sara</td>
<td>Crane</td>
<td>Parker</td>
<td>F</td>
<td>8</td>
<td>6</td>
<td>1990</td>
<td>8/5/2013</td>
</tr>
<tr>
<td>D</td>
<td>D421095241</td>
<td>Mary</td>
<td>Rachel</td>
<td>Moore</td>
<td>F</td>
<td>10</td>
<td>19</td>
<td>1990</td>
<td>8/5/2013</td>
</tr>
</tbody>
</table>

Data Adapter Hashed & Reduced Data (Done at the Agency server)

<table>
<thead>
<tr>
<th>INT_ID</th>
<th>UEID</th>
<th>FIRST_NAME</th>
<th>MIDDLE_NAMES</th>
<th>LAST_NAME</th>
<th>GENDER</th>
<th>DOB_MONTH</th>
<th>DOB_DAY</th>
<th>DOB_YEAR</th>
<th>RECORD_DATE</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>A395825930</td>
<td>hash</td>
<td>hash</td>
<td>hash</td>
<td>hash</td>
<td>hash</td>
<td>hash</td>
<td>hash</td>
<td>hash</td>
</tr>
<tr>
<td>B</td>
<td>B402924567</td>
<td>hash</td>
<td>hash</td>
<td>hash</td>
<td>hash</td>
<td>hash</td>
<td>hash</td>
<td>hash</td>
<td>hash</td>
</tr>
<tr>
<td>C</td>
<td>C940293847</td>
<td>hash</td>
<td>hash</td>
<td>hash</td>
<td>hash</td>
<td>hash</td>
<td>hash</td>
<td>hash</td>
<td>hash</td>
</tr>
<tr>
<td>D</td>
<td>D421095241</td>
<td>hash</td>
<td>hash</td>
<td>hash</td>
<td>hash</td>
<td>hash</td>
<td>hash</td>
<td>hash</td>
<td>hash</td>
</tr>
</tbody>
</table>