CSI Redesign Data Quality Assessment

Strategic Roadmap

Prepared for

Commonwealth of Virginia
Department of Motor Vehicles
Data Quality Initiative

December 14, 2007

Revision History

Date	Version	Prepared by	Comments
12/12/2007	1.0	Pete Carr/Brian Harris/Mike Tevebaugh	1st Draft of Document
12/12/2007	1.1	Pete Carr/Brian Harris/Mike Tevebaugh	Added details around Heatmap
12/12/2007	1.2	Pete Carr/Brian Harris/Mike Tevebaugh	Updated Heatmap, added details to Heatmap section, reviewed document
12/14/2007	1.3	Pete Carr/Brian Harris/Mike Tevebaugh	Incorporated reviewed changes

Document Control

Role	Name	Email	Telephone Number
Owner/ Preparer	Peter Carr	pcarr@captechventures.com	(804) 306-8437
	Brian Harris	bharris@captechventures.com	(804) 647-3195
	Deb Kroeger	deb.kroeger@dmv.virginia.gov	(804) 367-0372
	Michael Tevebaugh	mtevebaugh@captechventures.com	(804) 647-4727

Approval

Role	Name	Signature
CSI Redesign Project Manager	Tully Welborn	
Assistant Commissioner for Driver and Vehicle Services	Karen Grim	
Data Management Services Director	Theresa Gonyo	
Data Management Services Deputy Director	Jeff Harper	
Integration and Performance Analyst	Deb Kroeger	

Table of Contents

Data Quality Initiative Wrap Up

The Data Quality Team (DQT), formed to address data quality at the DMV, is led by Theresa Gonyo, Director of Data Management Services, Jeff Harper, Deputy Director of Data Management Services, Deb Kroeger, Integration and Performance Analyst, and the CapTech Data Quality Consulting team. From September 2007 to December 2007, the DQT conducted and managed a Data Quality Initiative that involved documenting all data quality issues, creating work plans to remediate the high priority data quality problems, conducting cost benefit analysis, and serving as the "go to" work center for data quality.

During the Data Quality Initiative, the DQT delivered a total of six documents: Current State Analysis, Data Quality Findings Log, Best Practices, Future State Vision, Gap Analysis, and Strategic Roadmap. Table 1 below gives information as to the status and location of the documents.

Document	Status	Location
Current State Analysis	Published	Current State Analysis
Data Quality Findings Log	Published	Data Quality Findings Log
Best Practices	Internal	Contact DMV Data Management Services
Future State Vision	Published	Future State Vision
Gap Analysis	Published	Gap Analysis
Strategic Roadmap	Published	Strategic Roadmap

Table 1 - Document Information

The Current State Analysis and the Data Quality Findings Log documented the examination of the DMV's primary data quality issues. The documents identified and prioritized the issues through interviews with key business stakeholders. The Best Practices document discussed and compared data quality tools that could remedy the identified issues. The Future State Vision presented the DMV with recommendations on achieving continuous data quality improvement. The Gap Analysis fully described detailed business plans around installing a data quality tool, cleansing, preventing, and integrating the data quality issues. Finally, the Strategic Roadmap lists "quick wins" or immediate cost effective data quality actions that will result in substantial benefits.

Strategic Roadmap Introduction

The purpose of the Strategic Roadmap is to provide the DMV with a snapshot of the data quality issues documented during the Data Quality Initiative. The document will enable the DMV to quickly ascertain the primary data quality issues and identify "quick win" opportunities. In essence, the Roadmap aggregates the Gap Analysis and Data Quality Findings Log documents to form an executive-level summary of the Data Quality Initiative.

The Strategic Roadmap is composed of two sections: Executive Summary and Heatmap. The Executive Summary is a high level view of the issues discussed in the Gap Analysis. While the Gap Analysis document provides in-depth analysis, the Executive Summary reviews and encapsulates the crucial information about the data quality issues and includes tangible and intangible benefits.

The Heatmap provides a graphical cost and impact comparison of every issue detailed in the Data Quality Findings Log. The Heatmap also highlights "quick win" opportunities and "strategic initiatives," providing the DMV recommended data quality issues to initially address.

CSI "Executive Summary"

The CSI "executive summary" below gives a listing of all data quality improvement opportunities, followed by tangible and intangible benefits descriptions. The format is based on a template employed by the CSI redesign team to document business process re-engineering. Therefore, the DQT used it to present the following relevant data quality issues; however, only applicable sections of the template contain information while others remain blank.

Overview of Data Quality Area

The Data Quality Team (DQT), formed to address data quality at the DMV, is led by Theresa Gonyo, Director of Data Management Services, Jeff Harper, Deputy Director of Data Management Services, Deb Kroeger, Integration and Performance Analyst, and the CapTech Data Quality Consulting team. The purpose of the DQT is to identify current data quality issues, define best practices, define a future or ideal state, and develop business plans around remediating highest priority data quality issues, prior to the CSI system redesign. The DQT's primary functions include gathering and documenting all data quality issues, conducting cost benefit analysis on identified data quality issues, creating plans to remediate the data quality problems, and serving as the "go to" work center for data quality.

The DQT completed six deliverables including the Current State Analysis, Data Quality Findings Log, Best Practices, Future State Vision, Gap Analysis, and Strategic Roadmap. The Current State Analysis and Data Quality Findings Log are an examination of the DMV's primary data quality issues that were identified and prioritized after interviews with key business stakeholders. Data quality tools that could remedy the identified issues were discussed and compared in the Best Practices document. The Future State Vision presents the DMV with recommendations on achieving continuous data quality improvement. Detailed business plans around installing a data quality tool, cleansing, preventing, and integrating the data quality issues are fully described in the Gap Analysis. Finally, the Strategic Roadmap lists "quick wins" or data quality issues that can be fixed in a cost effective manner and result in large benefits.

Summary of Data Quality Improvement Opportunities

#	Priority Critical Important Desirable	CRM	Done	Description
General				
1	☐ ☐ ☐	☐	☐	**Creation of a Data Quality Manager and Team** – Given the large data quality issues around

				business customers, a data quality manager and team will oversee the cleanup of duplicate customers as well as the entry of new data quality issues into the system. This will allow a core group of DMV employees to become 'experts' in data quality.
Business Customers				
2	☐ ☐ ☐	☐	☐	**Cleanse and prevent reentry of all duplicate business records –** All duplicate records will be combined or purged from CSS by a manual combine process, cleansing all duplicate business records. Additionally, all entry points of bad business data will be eliminated by employing preventative measures with the data quality tool.
3	☐ ☐ ☐	☐	☐	**Standardize all business record names and addresses** – Use the data quality tool to generate standardized business names; use the data quality tool or manually standardize all business records according to the defined standards.
4	☐ ☐ ☐	☐	☐	**Add business contact information to business record –** This will allow the DMV to have access to the point of contact for each business if needed in the future.
5	☐ ☐ ☐	☐	☐	**Define standards for business customer name and address** – All business customer names and addresses need a standardized nomenclature or structure.
6	☐ ☐ ☐	☐	☐	**Create new business customer selection screen** – CSS would give CSC tellers an easier method for viewing multiple business customers. This screen could possibly display up to 15 business records per screen.
Individual Customers				
7	☐ ☐ ☐	☐	☐	**Cleanse and prevent reentry of all duplicate individual records –** Individual customers have duplicate records that will be combined
8	☐ ☐ ☐	☐	☐	**Auto-populate Jurisdiction Field** – Jurisdiction fields on an individual's customer record can be auto populated by linking a customer's zip code + 4 to a table that matches zip codes with jurisdictions. By creating this auto match process, incorrect jurisdiction fields can be cleansed and

				prevented from reoccurring.
9	☐ ☐ ☐	☐	☐	**Improve quality of data entering CSS from CAIS** – Approximately 20 duplicate records enter CSS from CAIS daily. Many times the data coming across is incomplete or entered incorrectly. There are edit checks to help reduce the risk of wrong information being posted to the wrong record; however, this also results in convictions not getting posted to existing CSS records.
10	☐ ☐ ☐	☐	☐	**Fix duplicate indicator issue –** All records which have an incorrectly set duplicate indicator will be removed, reducing the amount of records that are currently registered as duplicates.
Motor Carrier				
11	☐ ☐ ☐	☐	☐	**Integrate Motor Carrier Systems with CSS –** Integrating Motor Carrier systems with CSS would permit a comprehensive, single view of the customer (including carrier customers).

	Tangible Benefits (Savings in Time, Staff, Other Costs, etc.)	**Projected Savings or Cost Avoidance**
General		
1		
Business Customers		
2	Data Integrity (11 hrs/day) and SSG (6.4 hrs/day) will both save significant amounts of time and thus, money, per year by having all bad business data cleansed and prevented. Givens: Data Integrity ($25.48/hr, 260 days/yr), SSG ($40.79/hr, 260 days/yr)	$140,747 savings per year
3		
4		
5		
6		
Individual Customers		
7	CSCs, CCC, SSG, and Data Integrity will all save a significant amount of time per day by having cleansed bad data and prevented its reentry (22 hrs/day).	$162,029 savings per year

	Givens: CSCs ($25.48/hr, 260 days/yr), CCC ($25.48/hr, 260 days/yr), SSG ($40.79/hr, 260 days/yr), Data Integrity ($25.48/hr, 260 days/yr)	
8	CSCs (1.33 hrs/day) will save time by not having to type a jurisdiction in the jurisdiction field. Givens: CSCs ($25.48/hr, 260 days/yr), 320 new records entered per day	$8,833 savings per year
9		
10		
Motor Carrier		
11		

	Intangible Benefits (Communications, Efficiency, Service Alternatives, Customer Satisfaction, Improved Access, Accuracy, Security, Fraud Prevention, etc.)
General	
1	**Efficiency –** Due to specialization in data quality issues, the data quality manager and team will understand data quality issues, resulting in fewer incorrect postings and transactions.
Business Customers	
2	**Customer Satisfaction –** The DMV will be able to better fulfill business customer expectations by having zero duplicate business records. The potential for mistakes will decline as a result of fewer duplicate records, ultimately leading to an increase in customer satisfaction. Also, zero duplicate records will reduce the number of times that business customers have to reach out to the DMV due to data quality issues. **Efficiency, Accuracy -** Cleansing the business customer duplicates prior to the implementation of CSI will allow clean data to be migrated from the DMV's current systems to their new system. **Less Confusion –** Zero duplicate business records will reduce confusion CSC teller and work center staff encounter when multiple records appear for the same customer.
3	**Accuracy –** Standardizing current business records will cleanse many instances of bad business data, leading to higher consistency and accuracy.
4	**Communications, Efficiency –** Having a method of contacting businesses will allow for open communication between the DMV and the businesses. On the DMV's side, when specific questions arise pertaining to business records, it will be useful to have contact information to facilitate communication and improve the efficiency of resolving issues with business records.
5	**Accuracy –** A standardized nomenclature will result in a higher degree of consistency and accuracy among all business records. Predefined standards will also prevent future bad business data from entering the system.

	Less Confusion – The creation of a standard process for entering and storing business customer data in DMV's systems will eliminate the current confusion that surrounds the processes.
6	**Efficiency –** CSC tellers will be able to locate the relevant business records more quickly if more business records are displayed per screen. CSC tellers will save time by scrolling through fewer screens if the system displays a large number of business records per screen.
Individual Customers	
7	**Customer Satisfaction –** The DMV will be able to better fulfill individual customer expectations by having zero duplicate business records. The potential for mistakes will decline as a result of fewer duplicate records, ultimately leading to an increase in customer satisfaction. **Accuracy -** Cleansing the individual customer duplicates prior to the implementation of CSI will allow clean data to be migrated from the DMV's current systems to their new system. **Less Confusion –** Zero duplicate individual records will reduce confusion CSC teller and work center staff encounter when multiple records appear for the same customer. **Efficiency** – Correspondence sent in the mail to individual customers will decrease when there are no duplicate records (e.g. John Smith will not receive incorrect registration mailings from an address on a duplicate record). Also, SSG will not spend time undoing incorrectly combined records once there are zero duplicate records **Less Risk** – Lessened liability risk with more accurate records and zero duplicate records
8	**Accuracy** – The accuracy of the jurisdiction field will increase tremendously because all jurisdiction fields will be auto populated based on the customer's zip +4.
9	**Accuracy, Customer Satisfaction –** Correct data entering CSS from CAIS will results in higher accuracy when conducing customer transactions and ultimately, higher customer satisfaction.
10	**Accuracy –** The DMV will be able to generate more accurate reports when searching for duplicate records. A true count of duplicate affected records will enable the DMV to determine whether running programs to fix the duplicate records is worth the time and money.
Motor Carrier	
11	**Customer Satisfaction –** Having a single view of the customer would ensure that there is less redundancy and fewer mistakes, increasing the DMV's opportunity to provide improved customer service. In addition, the customer will have a more transparent view of the DMV. Not only might a customer perform functions easily, securely, and reliably from a location of their choice (e.g. home/business), but they could do so having the confidence that everything they need to accomplish (e.g. registration, tax reporting, insurance validation) can be done from a common interface. Also, the DMV could increase its ability to inform the customer of the correct fees/taxes owed, ensuring

	customers do not get charged costly fines. **Efficiency –** A tremendous amount of DMV labor is spent monitoring compliance, and researching lack of compliance by commercial carriers. If the organization can increase compliance through better communication and effective interfaces, then it would reduce non-compliance costs within DMV because there would be less of it. **Accuracy –** One of the strongest reasons to integrate Motor Carrier data is so that the organization can compare IRP registrations with IFTA reporting. It is believed that IRP registrations often underreport vehicle usage when compared to IFTA. If proper fees for registrations were collected, the DMV would see an increase in revenue.

	Added Resources (Projected added resources (non-FTE related) including space, equipment, start-up, additional cost)	**Projected Added Resources**
Business Customers		
	None determined	N/A
Individual Customers		
	None determined	N/A
Motor Carrier		
	None determined	N/A

Summary of Proposed Legislative Opportunities

	Priority Critical	Important	Desirable	**Description**
1	☐	☐	☐	
2	☐	☐	☐	
3	☐	☐	☐	

	Tangible Benefits (Savings in Time, Staff, Other Costs, etc.)	**Projected Savings or Cost Avoidance**
1		
2		
3		

	Intangible Benefits (Communications, Efficiency, Service Alternatives, Customer Satisfaction, Improved Access, Accuracy, Security, Fraud Prevention, etc.)
1	
2	
3	

Summary of Staff Impact (FTE's only)

Current Staff			Projected Staff		Changes in Required Skills
	#	Role	#	Role	
1					
2					
3					
4					

Other Information, Issues, Etc.

Artifacts from Enterprise Architect

	Unit of Work Report(s)	Date Produced	Sign-off by Business Process Re-engineering Team
1	Scenario Report		
2	Requirements Report		

Review and Validation Sign-off (Business Area)

	Name	Title	Date	Comments
		Assistant Commissioner		
		Director		

Heatmap

The purpose of the Heatmap is to provide a graphic representation of the entire catalog of data quality issues found in the Data Quality Findings Log. The Heatmap, shown below in Figure 1, displays the spectrum of data quality problems across cost and impact axes and provides a quick snapshot of the cost/impact implications, allowing the DMV to easily evaluate and prioritize the issues. In the graph below, each issue is represented by a diamond. The number within each diamond corresponds to the "Traceback ID" located in the Data Quality Findings Log. As cost decreases and impact increases, the color changes from yellow (cool) to red (hot), mirroring an increase in viability and benefit for correcting the issue. The Heatmap below also reflects the impact and cost of correcting the data quality issues after the recommendations from the Gap Analysis have been implemented.

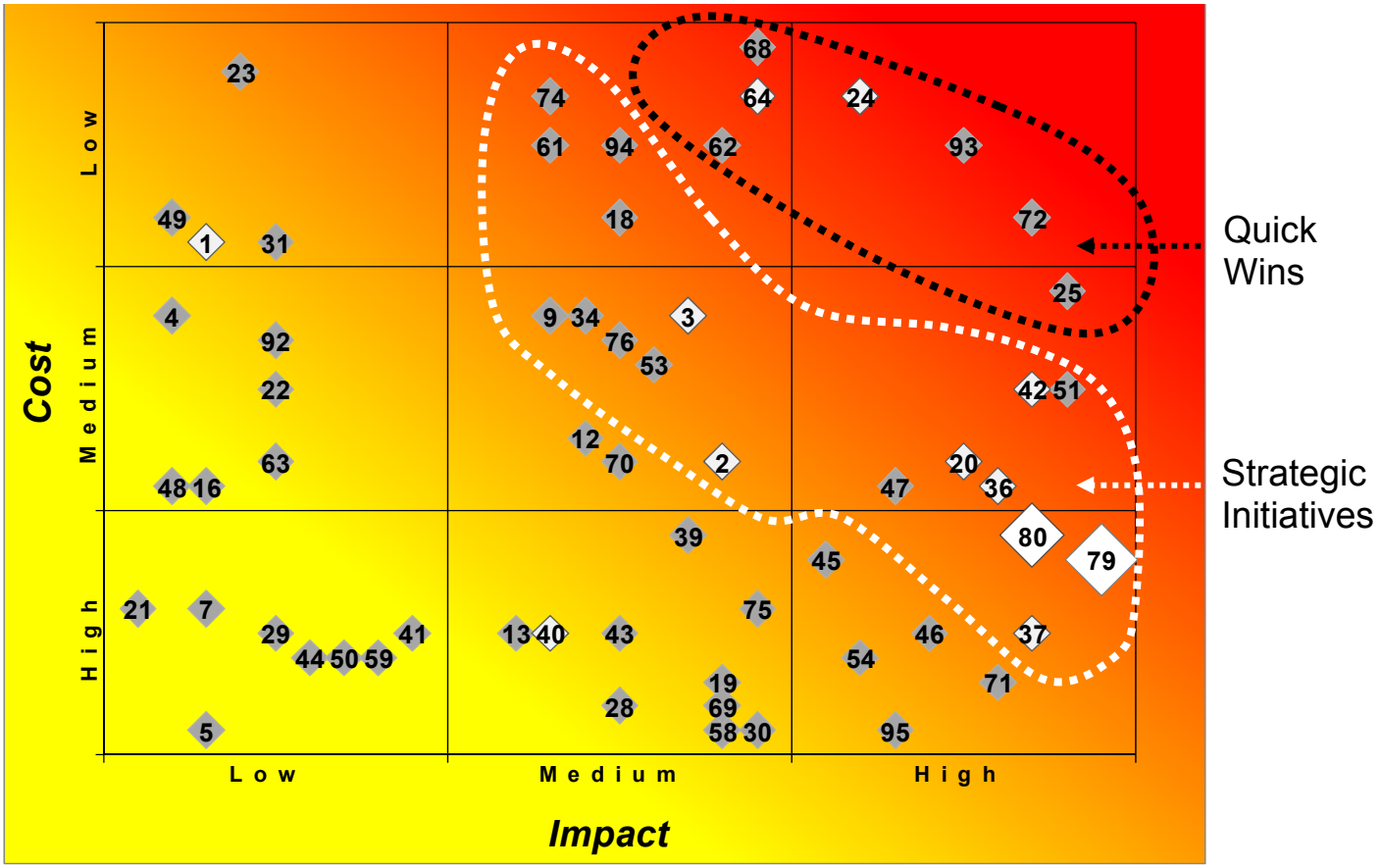

Figure 1 – Heatmap After Recommendation

Determination of Cost and Impact

The process of determining the cost to correct the data quality problem is based on a high-level assessment by the Data Quality Team. This assessment considered whether or not IT changes would need to be implemented as well as the number of work units affected. The assessment of cost was subjective, but is founded on the DQT's experience with resolving data quality problems.

The determination of business impact is based on breadth of issue (i.e. how numerous are the occurrences of the issue) and on how visible the issue is to DMV customers. In some cases the data quality problem occurs infrequently, but has a major "inconvenience factor" or "explicit cost factor" to DMV customers. These types of issues are considered highly visible to our customers. Conviction data applied twice to the same customer, issue 51, is one example of a high visibility issue.

Colors, Size, and Shading

During the Current State Analysis, extensive interviews took place across many work units within DMV resulting in a large number of identified data quality issues. In the final Current State Analysis deliverable, however, only those issues that were high impact or were thought to have a "clean up" solution were discussed in detail. The Gap Analysis outlined the business plans to remediate (i.e. cleanse) the data quality problems that were described in the earlier Current State Analysis deliverable. Thus, very little analysis was performed on the business impact of other data quality problems, nor on the costs to correct those problems. Data quality issues that were analyzed in the Gap Analysis are represented by white diamonds while the remaining issues in the Findings Log are represented by gray diamonds on the Heatmap graph.

The color of the Heatmap transforms from yellow in the lower left corner to red in the upper right corner. Issues in the yellow area are should not initially be addressed, as the issues are too costly to resolve and have little business impact. The issues in the darker red section should receive more scrutiny and attention, as they have higher impact and lower costs to resolve.

The size of the issue on the Heatmap represents the tangible benefits that were identified in the gap analysis. Only issues that were deemed clean-up and that were previously prioritized have tangible benefits associated with the resolution (eg. 79, 80). Other items might have tangible benefits associated with their correction; however, more research is needed to determine the precise amounts.

Quick Wins, Strategic Initiatives, Other Issues

Figure 1 – Heatmap After Recommendation above identifies "quick win" projects, "strategic initiatives," and other issues. A "quick win" is a solution to a data quality problem that has a high impact to the business and relatively low costs to remediate. Strategic initiatives/projects are the grouping of data quality issues that have one of the following combinations of impact/cost relationships: high impact with moderate

costs, moderate impact with moderate costs, moderate impact with low costs. Strategic initiatives will require additional planning but should be pursued as quickly as possible. The other issues on the Heatmap can be classified as long term initiatives. Long term initiatives should be addressed to achieve a zero defect state, but should be a lower priority

The quick wins identified in the Heatmap, or the issues having high impact and low cost, are listed in Table 2 below.

Quick Wins	
#	Description
24	Duplicate indicator is not removed when duplicate records are combined
25	CAIS data entry
62	Sending PIN through postal service for change of address requests on internet
64	Individual records with the same 9-digit customer control number
68	The title document print date reflects date of title transaction versus date of title print.
72	Duplicate conviction data due to manual process
93	Incomplete CAIS data causing incomplete customer records

Table 2 - Quick Wins

Two "Quick win" issues, 24 and 64, were analyzed in the Gap Analysis. Both are anticipated to incur minimal implementation costs and will both have significant impact on the DMV (see the Individual Data Cleanse Plan from the Gap Analysis). The other quick wins have not been researched in depth.

CAIS data entry, issue 25, is expected to have a very large impact on the DMV when addressed given the scope of bad data entering CSS from CAIS. Discussions have already begun with CAIS officials and it is believed that the majority of the changes will occur on the CAIS side, lessening the cost to DMV for correcting issue 25. Issue 62, sending PIN through postal service, should be a quick fix, involving sending the PIN to the DMV customer via email, and cost should be relatively low, given the minimal change to the website.

Issue 68 will involve a small change to the code that prints vehicle titles, yet the impact to the DMV could be great, as inactive and active titles will be distinguishable by the print date. The final "quick win," issue 72, will impact the DMV greatly, as convictions will not be posted multiple times to customer records, thus improving customer satisfaction. Cost will be low with the data quality tool, as rules within the data quality tool will prevent duplicate convictions from posting.

Strategic initiatives are those issues identified as having a higher cost but also medium to high impact for the DMV. The strategic initiatives are detailed below in Table 3.

Strategic Initiatives	
#	Description
2	Business customers with multiple locations
3	Business record address change
9	Dealership title incorrect
18	Duplicate information recorded in license commercial driving schools
20	End-of-Day error handling in Motor Carrier
34	Duplicate T numbers
36	Rental tax incorrect jurisdiction
37	IRP registration delay
42	IFTA/CSS address differences
47	Make tables rejections
51	Address change requests by the customer cannot always be determined or corrected
53	DMV incorrectly credits leasees' customer records for decal invoices
61	Titles archived cannot have update information posted to the record
74	3rd party vendor incorrectly keys suspended license data
76	CSC tellers can override an NDR hit
79	Multiple records for business customers
80	Individual records with different customer control numbers for the same customer
94	Duplicate lien holders

Table 3 - Strategic Initiatives

Issues 2, 3, 20, 36, 37, 42, 79, and 80 are given ample detail in the Gap Analysis document. Issues 79 and 80 in the Heatmap are the only two issues with tangible benefit, as represented in the graph.

To provide insight into the DQT's reasoning behind the strategic initiative data quality issues, detail around issue 76 and 79 is provided as follows. Issue 76 is anticipated to incur moderate cost to remediate, and have a moderate impact to the DMV. The resolution of this issue will involve a change in the Natural code to add matching criteria that will prevent a CSC teller from overriding an NDR. Issue 79 has significant labor costs to resolve, but will also provide a high impact to the DMV. The resolution of this issue will involve multiple steps, including the DMV's reaching out to their business customers to verify vehicle, address, and other data. Once resolved, the DMV will contain high quality business customer data that will result in a savings of many hours of employee time.

Detail for the additional strategic initiative data quality issues can be found in the Data Quality Findings Log.