Pennsylvania Statewide Transportation Operations Data Warehousing Business Plan

BUSINESS PLAN AND COST ANALYSIS REPORT

May 11, 2016

By Michael Baker International, Inc.

COMMONWEALTH OF PENNSYLVANIA DEPARTMENT OF TRANSPORTATION

CONTRACT # 355I01
PROJECT # 130108
## Abstract

The Pennsylvania transportation community generates a considerable amount of transportation operations data such as traffic volume data, incident data, asset information, and speed data without an effective means to share the data between one another. With a vision to provide stakeholders with access to the best available transportation operations data to support the safe, secure, and efficient movement of people and goods on Pennsylvania roads, the Pennsylvania Department of Transportation (PennDOT) along with its stakeholder community consisting of PennDOT Engineering District and transit, tolling, freight, local government, and emergency management agencies undertook the Pennsylvania Statewide Transportation Operations Data Warehouse (PA STODW) Business Plan project. The project involved determining stakeholder data needs and limitations, documenting national best practices, preparing the PA STODW Concept of Operations, and developing the Business Plan and Cost Analysis Report. The Business Plan and Cost Analysis Report is a tool for the Pennsylvania transportation operations community to develop and implement the PA STODW.

## Key Words

transportation operations, data warehouses, business plan, concept of operations, functional requirements, alternatives analysis

## Distribution Statement

No restrictions. This document is available from the National Technical Information Service, Springfield, VA 22161
Table of Contents

Executive Summary .......................................................................................................................... 1
1. The Foundation .......................................................................................................................... 2
   1.1. Project Concept .................................................................................................................... 2
   1.2. Stakeholder Data Needs Assessment ..................................................................................... 2
   1.3. External Agency Outreach ..................................................................................................... 3
   1.4. Concept of Operations .......................................................................................................... 3
2. Customer Focused Value ........................................................................................................... 4
   2.1. Vision for Value ..................................................................................................................... 4
   2.2. Goals to Achieve Value ......................................................................................................... 5
   2.3. Objectives to Provide Value .................................................................................................. 6
3. Measures of Success .................................................................................................................. 7
   3.1. Business Value ..................................................................................................................... 7
   3.2. Technical Capability ............................................................................................................ 8
4. Implementation Alternatives ..................................................................................................... 8
   4.1. Viable Alternatives ................................................................................................................. 8
   4.2. Alternatives Comparison ...................................................................................................... 11
   4.3. Alternatives Cost Analysis .................................................................................................. 12
5. Recommendation ..................................................................................................................... 14
   5.1. Recommended Implementation Alternative ............................................................................ 14
   5.2. Implementation Approach ................................................................................................... 15

Figures

Figure 1: PA STODW Deployment Alternatives Concept ................................................................... 9
Figure 2: Recommendation Decision Tree ....................................................................................... 14

Tables

Table 1: PA STODW Deployment Options ....................................................................................... 10
Table 2: PA STODW Alternatives Comparison Matrix ..................................................................... 12
Table 3: PA STODW Alternatives Cost Analysis .............................................................................. 13
Table 4: PA STODW System Usage and Sizing Requirements .......................................................... 16
Table 5: PA STODW Implementation Schedule .............................................................................. 17
Executive Summary

PennDOT’s mission is to “provide a sustainable transportation system and quality services that are embraced by our community and add value to our customers.” In support of this mission, PennDOT’s State Transportation Commission has also proclaimed that it is committed to implementing cost-saving measures and modernization of business practices that make sense and help to stretch transportation dollars and services.¹

To help PennDOT meet these ambitious objectives, the Bureaus of Maintenance and Operations and Business Solutions and Services, along with their stakeholders, recommend developing and implementing the Pennsylvania Statewide Transportation Operations Data Warehouse (PA STODW).

The Pennsylvania transportation community generates a considerable amount of transportation operations data such as traffic volume data, incident data, asset information, and speed data without an effective means to share the data between one another. The PA STODW will ultimately help to increase planning, operations, and research capabilities within the state, significantly reducing the cost of doing business, dramatically improving agency capabilities, and improving coordination with partners in transit, tolling, freight, local governments, and emergency management agencies.

Implementation of a PA STODW would “provide stakeholders with access to the best available transportation operations data in support of the safe, secure and efficient movement of people and goods on Pennsylvania roads.” In support of this vision, the PA STODW will:

- Support objectives-driven, performance-based transportation planning
- Support transportation operations performance measures
- Support data-driven operational strategy deployments
- Provide a means to integrate and interconnect operations data

The estimated cost to implement and maintain the PA STODW over a six-year horizon is $1.7M; however, there is an expected benefit of more than 10:1 for this investment. The Business Plan and Cost Analysis Report provides a detailed discussion of the implementation alternatives considered and an explanation of the recommended course of action to develop, deploy, and maintain the PA STODW based on a set of defined criteria.

1. The Foundation

The Pennsylvania Department of Transportation (PennDOT), together with its partnering transportation operations stakeholders, have set out to develop a Pennsylvania Statewide Transportation Operations Data Warehouse (PA STODW) that will in essence provide a common means to share transportation operations data between agencies so that each agency can leverage the institutional, fiscal, safety, and mobility benefits of shared information.

The following provides background information and explains the steps taken prior to developing the PA STODW Business Plan and Cost Analysis Report. Developing a business plan to justify and properly document a strategy for the PA STODW required a multi-phased approach to support sound analysis and decision making.

1.1. Project Concept

The PA STODW project was first conceptualized and gained stakeholder backing through the Pennsylvania State Transportation Innovation Council’s (STIC) Intelligent Transportation Systems (ITS) Technical Advisory Group (TAG). The ITS TAG, whose membership consisted of transportation agencies, planning organizations, consultants, academic institutions, and professional societies, developed and submitted a white paper to the STIC highlighting the background, purpose, and outcomes of pursuing a Transportation Operations Data Warehousing and Management project.

1.2. Stakeholder Data Needs Assessment

Upon selection by the STIC and subsequent funding and project award, the Pennsylvania Statewide Transportation Operations Data Warehousing Business Plan project began. The first task completed as part of the project was to identify internal (PennDOT) and external (non-PennDOT) transportation operations stakeholders within Pennsylvania and determine their data availability, data needs, and limitations for sharing data. Twenty-four (24) stakeholder entities representing numerous PennDOT bureaus, tolling agencies, emergency management agencies, planning partners, municipalities, multi-modal agencies, and commercial/freight partners were identified as the stakeholder group. Stakeholder
data needs were then assessed during a facilitated day long “visioning workshop” and a follow-up survey. Stakeholder entities attended the visioning workshop and participated in the following working sessions:

- Identify stakeholders and functional categories of data to consider as part of the PA STODW.
- Identify systems, specific data types, and data applications for inclusion in the PA STODW.
- Prioritize data and application needs and identify barriers and challenges for sharing data within the PA STODW.

Following the visioning workshop, a 30-question survey was administered to the stakeholder group and eleven (11) stakeholder entities responded. The survey focused on determining the availability of transportation operations data, which data is desirable, data sharing issues, data analysis needs, and the benefits and challenges associated with having a PA STODW. The findings of the visioning workshop and subsequent survey were summarized in the “Transportation Operations Data Assessment Report” dated May 22, 2015.

1.3. External Agency Outreach

The second task completed as part of the project was to research, document, and understand current best practices and recommendations of agencies outside of Pennsylvania who have demonstrated they are national leaders in warehousing transportation data. Investigating how these agencies developed their data warehouses and how they maintain and update their transportation data provided insights and lessons learned which assisted in developing the context for the PA STODW. The following three (3) agencies were interviewed and their data warehousing approaches were investigated:

1. Utah Department of Transportation (UDOT)
2. Maryland State Highway Administration (MDSHA)
3. United States Department of Transportation (USDOT) Joint Program Office

During the interviews, each external agency was asked about their concept of operations, business procedures, implementation challenges, and recommendations for developing and maintaining a transportation data warehouse. The findings of the external agency outreach were summarized in the “External Agencies’ Data Warehouse Systems Outreach Report” dated September 8, 2015.

1.4. Concept of Operations

Once the PA STODW needs were defined and industry leading sample deployments were evaluated, the next project task was to prepare a Concept of Operations (ConOps) for the PA STODW to document and
communicate the quantitative and qualitative characteristics of the proposed system. In simple terms, the ConOps describes “what” the PA STODW needs to be and do in order to fulfill the expectations and user-needs of the PA STODW stakeholder group. The PA STODW ConOps, current Version 1.2 dated May 2016, presents the scope of the data warehouse, the concept of the proposed data warehouse system, and the data needs of the PA STODW stakeholder group. An overview of the proposed system included a description of the current and planned operational and support environment, use-case scenarios, and a summary of impacts. The ConOps should be referenced for specific details related to the proposed PA STODW.

2. Customer Focused Value

PennDOT and its PA STODW stakeholder group partners are in the business of providing and operating a complex transportation system throughout Pennsylvania. The customer within the transportation business model is each and every patron of the transportation system whether they moving goods, taking a train or bus, behind the wheel, or walking/bicycling. These customers expect to use the transportation system in a safe, secure, and efficient manner. As stewards of Pennsylvania’s transportation system, the stakeholder group understands that in order to provide a safe, secure, and efficient system, they need information at their fingertips to measure whether or not they are meeting the needs of their customers and to identify areas for improvement and more cost effective investment. The need for transportation operations data isn’t a new concept but what has changed is the amount of data, the varying types of data available, and the ways in which data is consumed, analyzed, and shared. Transportation agencies make significant investments to acquire, store, and interpret transportation operations data but it is often difficult and time consuming to share data between agencies. In order to crowdsource each agency’s data investments and in turn better serve the collective customer base, the PA STODW will be the clearinghouse for transportation operations data.

2.1. Vision for Value

The PA STODW vision statement (above) acknowledges the core measures by which the data warehouse project will be measured to determine whether or not the project provides value to the stakeholder group. The PA STODW must first and foremost provide access to shared transportation operations data. In order to provide access, all appropriate stakeholders need to have the ability and means to interact with the data warehouse in an intuitive and cost effective manner. The PA STODW must also contain the best
available transportation operations data. A data warehouse full of data of little transactional importance or not standardized for use, will degrade the value of the PA STODW.

### 2.2. Goals to Achieve Value

The PA STODW stakeholder group established five (5) project goals that expand upon the vision statement and focus on the tangible value adding necessities of the project. These goals define what value the PA STODW needs to provide to transportation agency partners for the project to be a success.

**Goal 1**  
**Support Objectives-Driven, Performance-Based Transportation Planning**

- An objectives-driven, performance-based approach to transportation planning focuses on both short-term and long-term system performance, using system performance measures to drive decision making rather than focusing on implementation of projects as measures of success.
- The nation as a whole is moving towards objectives-driven, performance-based transportation planning whether it be in the form of a Congestion Management Process plan developed by a planning partner or a Transportation Systems Management and Operations Program undertaken by PennDOT.
- The PA STODW must support this "way in which we do business" culture shift by providing varying forms of data, from a diverse set of stakeholders, in an easily accessible manner.

**Goal 2**  
**Support Transportation Operations Performance Measures**

- The engine that supports an objectives-driven, performance-based culture are performance targets and measures. Performance measures can offer immediate answers for "real-time" decisions but they can also provide answers or insights for how to make appropriate long-term fiscal decisions to achieve defined targets.
- The PA STODW must provide insightful planning-level transportation operations data that in turn can be utilized in performance measurement.

**Goal 3**  
**Support Data-Driven Operational Strategy Deployments**

- The ultimate purpose of the PA STODW is to arm transportation agencies with the data they need to make the right decisions and in turn weigh the benefits and costs to deploy transportation operations projects.
- The PA STODW must provide data that supports the analyses of built operationally-focused deployments but also provide data that supports decisions for future deployments.

**Goal 4**  
**Provide a Means to Integrate and Interconnect Operations Data**

- The PA STODW must increase the level of data accountability and transparency within PennDOT and between PennDOT and its stakeholders.

**Goal 5**  
**Improve Transportation Operations Community Collaboration**

- The PA STODW must reduce the amount of operations data duplication by increasing the amount of data collaboration within Pennsylvania. This common information base contributes to the coordinated use of other inter-agency resources, including personnel, equipment, and funding.
2.3. Objectives to Provide Value

The following project objectives define how the PA STODW will provide value to the transportation operations community upon development and implementation.

**Objective 1**

**Goal Relationships: 1, 4, and 5**

**Open, Receptive, and Adaptable Architecture**

- The PA STODW needs to leverage existing IT infrastructure and staffing resources during the development, implementation, and maintenance of the data warehouse.
- The PA STODW architecture needs to be adaptable to remain consistent with developing national standards, industry trends, and changes in stakeholder needs.

**Objective 2**

**Goal Relationships: 1, 2, and 3**

**Integrated Planning-Level Transportation Operations Data**

- The PA STODW needs to archive planning-level transportation operations data, where archived data can then be used to support the objectives-driven decision making process. This objective excludes real-time unicast, narrowcast, multicast or broadcast messaging and other data distribution strategies used to communicate situational information between travelers, connected vehicles, ITS Centers and infrastructure attached field devices. The PA STODW is not intended to be the operational database for actual production systems. Data from these systems may be stored in the PA STODW after real-time situational analysis and related action(s) have occurred.

**Objective 3**

**Goal Relationships: 4 and 5**

**Standardized and Quality Data**

- The PA STODW needs to collect both datasets and metadata documentation from multiple data sources and store this information in a single repository.
- The PA STODW needs to support basic data fusion of multiple datasets based on common characteristics such as location and time.

**Objective 4**

**Goal Relationships: 1, 4, and 5**

**Defined Data Warehouse User Experience**

- The PA STODW needs to provide a web-based portal interface to support the exchange of datasets in their original format.
- The PA STODW needs to be accessible to stakeholders with PennDOT approved data warehouse access credentials.

**Objective 5**

**Goal Relationships: 1, 2, 3, and 5**

**Defined Data Warehouse Data Exchanges and Processes**

- The PA STODW needs to include capabilities for submitting and uploading incoming data, for administering a data repository containing data archives and a metadata catalog, and for browsing, searching and extracting the contents of the warehouse.
- The PA STODW needs to support outputs in a variety of formats including maps, business graphics, reports and structured data extracts.
3. Measures of Success

Building off of the project’s visions, goals, and objectives there are specific business and technical measures that must be evaluated during the development of the PA STODW to ensure that the end product provides the intended value to the stakeholder group and ultimately our customers. These measures have been categorized into two (2) general areas:

1. Business Value: PA STODW business requirements
2. Technical Capability: PA STODW technical requirements

Considered collectively, these measures define a performance framework against which various potential PA STODW project implementation alternatives can be evaluated. Alternatives that meet or exceed the measures of success at lower deployment and operating costs are considered to provide better value to the stakeholder group. The measures of success are intended to provide guide posts for PA STODW funding sources and managers to use in the system cost analysis rather than be absolute and comprehensive specifications for systems development purposes. The following subsections describe the measures and the subsequent chapter presents PA STODW project implementation alternatives contrasted against the measures.

3.1. Business Value

The PA STODW, as defined in the Concept of Operations, is meant to be a common resource used by a wide range of stakeholders including PennDOT and its stakeholder partners. The following parameters represent business value performance measures. These measures and those included in the next subsection are discriminators used to evaluate and rank various project implementation alternatives in the Implementation Alternatives section.

- **Leverage Existing Investments**: PennDOT has made significant investments in its PennDOT Data Integration Facility (PDIF). This Data Warehouse / Business Intelligence (DW/BI) platform contains many components that could potentially be reused for the PA STODW. Other data feeds, platforms, architectures, and processes have also been developed by PennDOT. To the extent possible, these investments should be leveraged to reduce costs and minimize redundancies.

- **Minimize Disruption**: The PA STODW will provide a data resource that stakeholders can rely on for business decisions, research, and otherwise implement new processes based on access to the data found within. The development, operations, and maintenance of the PA STODW should minimize unnecessary disruption to existing business processes and work flows, legacy systems, and personnel.

- **Minimize Total Cost of Ownership (6-year)**: Assuming that different implementation alternatives will offer commensurate benefits, the PA STODW with the lowest total cost of ownership over the planning horizon is preferred.
3.2. Technical Capability

In addition to the business value measures of success, the PA STODW’s technical characteristics also provide discriminators between potential project implementation alternatives. Some of these parameters have been derived from PennDOT’s current technology environment; others were adopted from the peer benchmarking exercise.

- **Scalability**: The PA STODW must support uploading, downloading and archiving of large data sets from a large and diverse set of partner organizations. Transportation operations data is composed of large structured and unstructured datasets, including data obtained from various roadside equipment, recorded surveillance video, connected vehicle data and other “big data” sets. The PA STODW is not intended to be the operational database for actual production systems (Advanced Traveler Information, Roadside Device Systems, Traffic Management Systems, etc.). Data from these systems can be stored in the PA STODW after real-time situational analysis and related action(s) have occurred. The PA STODW must be able to scale appropriately, when new users and new data are added to the system. Greater scalability is more desirable.

- **Reliability**: Although the PA STODW is not considered to be a mission-critical operations system whose failure would endanger people, property or the environment, the PA STODW must be able to consistently perform its required functions to meet business needs and deliver full value. Greater reliability is more desirable.

- **Risk**: In this context, risk is defined to include any factor that would prevent or impede the successful development, deployment or operation of the PA STODW. These risks may arise from many sources, including technological (unproven technologies or new implementations), personnel-related (low staff levels, lack of capacity and capabilities, or lack of succession planning), financial (high cost or dependent on unreliable sources of funding), contractual, etc. Lower risk is more desirable.

- **Level of Support**: The PA STODW will require a certain amount of support services from multiple PennDOT units including legal, purchasing, IT, BOMO and/or outside contractors. Lower support requirements are more desirable.

4. Implementation Alternatives

4.1. Viable Alternatives

The proposed PA STODW functionality and capabilities determined during the early visioning tasks of this project found that PennDOT had previously invested in and maintains the majority of the IT infrastructure required to deploy the PA STODW. This includes network servers and storage capabilities that could be extended to support the early and future demands of the PA STODW stakeholder community.

The Concept of Operations for the PA STODW defines the system and functional requirements and the relationship between the PA STODW and the ITS National Architecture. Based on current industry
knowledge and research performed, a commercial off-the-shelf (COTS) solution does not exist currently that can satisfy the defined requirements of the PA STODW.

PennDOT has several options on how to build, deploy, and maintain the PA STODW. The most fundamental decision is to determine if the system will be developed in-house or by hired consultant support. This decision depends on PennDOT having the skills, staff availability, funding, and patience to wait to develop on their own, or if consultants may be capable of delivering a solution cheaper, quicker, or more feature-rich as described previously.

![Figure 1: PA STODW Deployment Alternatives Concept](image)

Viable alternatives to design, build, deploy, and maintain the PA STODW are defined as follows:

1. **In-house Deployment:**
   A. PennDOT builds and deploys the fully customized PA STODW with internal staff and resources
   B. Consultants build and deploy a fully customized PA STODW

2. **Hosted/Cloud Deployment:**
   A. PennDOT builds and deploys the PA STODW to a cloud environment
   B. Consultant builds and deploys the PA STODW to a cloud environment
      i. Brand new, fully custom deployment
      ii. Customization of a COTS product
   C. Software as a service (SAAS) provider builds and deploys the PA STODW and charges a monthly or annual fee

Options 1A, 1B, 2A, 2Bi provide PennDOT with the ability to fully “own” the application source code; however, option 2Bii, or C, would result in only partial or no ownership of the application source code. **Table 1** on the following page presents development and deployment criteria, provides a comparison of the alternatives described above, and sets the foundation for the PA STODW Alternatives Comparison Matrix (**Table 2**).
Table 1: PA STODW Deployment Options

<table>
<thead>
<tr>
<th>Deployment Alternatives</th>
<th>In-house</th>
<th>Hosted/Cloud</th>
<th>Software as a Service (SAAS)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Who does the Work?</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Server &amp; Storage</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Infrastructure</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PennDOT</td>
<td>Leverages existing PennDOT pooled servers and networking resources.</td>
<td>Utilizes virtual/cloud services that rely on PennDOT's underlying physical servers.</td>
<td>A pool of cloud infrastructure resources designed specifically for the PA STODW that could include automated provisioning to simplify administration and maximize operations efficiency.</td>
</tr>
<tr>
<td>Consultant</td>
<td>Leverages existing PennDOT pooled servers and networking resources.</td>
<td>Build the application to the specification as defined by PennDOT. This option provides PennDOT full control throughout the build process.</td>
<td>Consultant would integrate data from PennDOT and customize the existing COTS to meet the needs as documented by PennDOT. This would be a software licensing and delivery model that would be paid through a subscription services method over the six (6) year planning horizon. As such, PennDOT would have limited involvement in the hosted architecture and technology decisions so long as business needs are being met. PennDOT would still have control over defining new functionality, user interface design, etc.</td>
</tr>
<tr>
<td><strong>Delivery</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Method</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Build the application to</td>
<td>Build the application to the specification as defined by PennDOT.</td>
<td>Consultant would build and host the PA STODW to meet the need as documented by PennDOT. PennDOT could still have the option of owning and/or maintaining the solution depending on their agreement with the consultant. Options to have the consultant maintain on PennDOT's behalf would also be available.</td>
<td></td>
</tr>
</tbody>
</table>
| the specifications as    | Build the application to the specification as defined by PennDOT. | Consult
| defined by PennDOT. This | Build the application to the specification as defined by PennDOT. |                             |
| option provides PennDOT  |          |              |                             |
| full control throughout  |          |              |                             |
| the build process.       |          |              |                             |
| **Operating**            |          |              |                             |
| Method                   |          |              |                             |
| The PA STODW is          | The PA STODW is housed on 3rd party infrastructure, and supported by in-house staff and/or consultant services. | The PA STODW is housed on 3rd party infrastructure, and supported by in-house staff. PennDOT would be responsible for platform administration, maintenance, system administration, system upgrades, etc. | The PA STODW would be deployed and running on an externally hosted environment. IT specific functions such as infrastructure, failover, hardware resource allocation, etc. would be the responsibility of the hosted provider. |
| housed on PennDOT         |          |              |                             |
| infrastructure, and      |          |              |                             |
| supported by in-          |          |              |                             |
| house staff and/or        |          |              |                             |
| consultant services.     |          |              |                             |
4.2. Alternatives Comparison

As there are a few defined optional approaches to deploying the PA STODW, a comparison and evaluation of each alternative is required. The performance indicators described below and derived from the measures of success are the means by which each alternative will be scored.

- **Ability to Leverage Existing Investments**: In the design, development, and deployment of the PA STODW, existing PennDOT infrastructure would be utilized to their full potential to minimize additional overlapping investments.

- **Minimize Disruption of Existing Business Functions**:
  - **Systems and IT Infrastructure**: The PA STODW’s impact on the existing hardware and IT infrastructure. This means that it has the ability to co-exist with other systems without negatively impacting the performance of other systems.
  - **Staff**: The PA STODW’s impact on system support staff.
  - **Business Processes**: The PA STODW’s impact on other existing business processes. This is separate from the infrastructure impact, and relates more to business data workflows.

- **Scalability**: Ability of the PA STODW to quickly scale-up storage, resources, and bandwidth if/when the need arises with minimal disruption to end users.

- **Reliability**: Overall system stability and redundancy, including system uptime, response time, etc.

- **Level of Risk**:
  - **Initial Investment/Deployment**: Project-related risk during the system design and deployment phase of the project. Risks include, but are not limited to, cost overruns, schedule overruns and not aligning to the defined scope of work.
  - **Obsolescence (end of life)**: Risk associated with functional obsolescence before, during, or after production implementation. Post-production implementation risk is a function of how agile the PA STODW can be in regards to becoming functionally or technologically obsolete.

- **Level of Support**: The level of support and training required to operate and maintain the PA STODW.

Table 2 represents the alternative comparison matrix that provides a rating for each of the performance indicators that have a direct correlation with the infrastructure and/or the technical resources. The scoring is based on a 1 to 5 scale, with 1 being the most desirable and 5 being the least desirable. Scores denoted as N/A implies that the performance indicator is either not applicable to the alternative or the performance indicator is irrelevant since all alternatives account for the indicator equally. Additionally, each performance indicator has been weighted equally. For example, leverage existing investment is no more or less important than minimize disruption.
### Table 2: PA STODW Alternatives Comparison Matrix

<table>
<thead>
<tr>
<th>Performance Indicators</th>
<th>Server Infrastructure</th>
<th>Storage Infrastructure</th>
<th>Delivery Method</th>
<th>Operating Method</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>In-house</td>
<td>Hosted</td>
<td>Hybrid</td>
<td>In-house</td>
</tr>
<tr>
<td>Leverage Existing Investment</td>
<td>1</td>
<td>5</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>Minimize Disruption</td>
<td>5</td>
<td>1</td>
<td>3</td>
<td>5</td>
</tr>
<tr>
<td>Scalability</td>
<td>5</td>
<td>1</td>
<td>3</td>
<td>5</td>
</tr>
<tr>
<td>Reliability *</td>
<td>2</td>
<td>4</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>Level of Risk **</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Level of Support ***</td>
<td>5</td>
<td>1</td>
<td>3</td>
<td>5</td>
</tr>
<tr>
<td>Average</td>
<td>3.6</td>
<td>2.4</td>
<td>3</td>
<td>3.6</td>
</tr>
</tbody>
</table>

* Considers infrastructure and connectivity, relating to relative reliability

** Excludes risk associated with lack of server and/or storage resources

*** Includes functional group involvement (legal, procurement, etc.)

In general, the scoring shows that the combination of a hosted infrastructure solution that is built in-house and maintained by in-house staff provides the most desirable alternative. Critical factors that are part of this scoring include hosted system vulnerability due to connectivity dependency, and the exclusion of the risk associated with the lack of available in-house resources.

### 4.3. Alternatives Cost Analysis

Table 3 represents the cost analysis for each of the alternatives identified. Costs provided are for planning purposes only, and may or may not be accurate to the precision provided. These alternatives assume a one-year system build-out and a five-year system hosting and maintenance period, totaling a 6-year total cost of ownership. Additionally, the costs provided assumes a minimal or zero infrastructure capital investment. It is assumed that licensing, hardware and connectivity needs are purchased separately or already available.

The following are the categories used to assist in developing the alternatives cost analysis.

- **Data Governance**: Accounts for costs to develop data sharing and other applicable policies related to standardization, metadata, update procedures, ownership responsibility, etc. A data governance program is typically administered by a governing body or council that is established to set policies and foster internal adoption and compliance.
• **Design Specification**: Accounts for defining the functional requirements of the PA STODW. The Concept of Operations set the foundation for the system, whereas the design specifications define precise capabilities of each component, and further define the workflow and “look and feel” of the application. The formal document developed will set criteria and application characteristics that the development team will need in order to build the system.

• **System Build**: Accounts for the process to actually write source code and build the application. Programmers develop code in accordance with the design specification, develop test plans/cases, perform quality control, perform customer acceptance testing, identify and resolve bugs, etc.

• **Training**: Accounts for developing and implementing a training program for end users and it also includes marketing outreach elements to promote the PA STODW and educate future stakeholders.

• **Delivery**: Accounts for deployment and testing procedures to ensure successful operation of the application on PennDOT infrastructure or cloud services. This is only applicable to a PennDOT in-house or a consultant built solution.

• **Hosting & Maintenance**: Accounts for application maintenance and hosting services provided after the initial deployment and throughout the defined planning horizon. Includes maintenance escalator costs over the planning horizon.

### Table 3: PA STODW Alternatives Cost Analysis

<table>
<thead>
<tr>
<th>Alternative</th>
<th>Data Governance</th>
<th>Design Specification</th>
<th>System Build</th>
<th>Training</th>
<th>Delivery</th>
<th>Hosting &amp; Maintenance (5 years)</th>
<th>Total *</th>
</tr>
</thead>
<tbody>
<tr>
<td>In-house</td>
<td>$ 70K</td>
<td>$ 110K</td>
<td>$ 800K</td>
<td>$ 90K</td>
<td>$ 40K</td>
<td>$ 600K</td>
<td>$ 1.710M</td>
</tr>
<tr>
<td>Consultant</td>
<td>$ 130K</td>
<td>$ 190K</td>
<td>$ 1.3M</td>
<td>$ 90K</td>
<td>$ 70K</td>
<td>$ 900K</td>
<td>$ 2.680M</td>
</tr>
<tr>
<td>SAAS **</td>
<td>$ 130K</td>
<td>$ 190K</td>
<td>$ 250K</td>
<td>$ 87K</td>
<td>$ -</td>
<td>$ 1.0M</td>
<td>$ 1.657M</td>
</tr>
</tbody>
</table>

* Represents a 6-year total cost of ownership

** Time to deploy a SAAS system is 6 months
5. Recommendation

5.1. Recommended Implementation Alternative

Although the performance indicators and cost comparisons among the three implementation alternatives both point to the same implementation recommendation, the Team realizes that there may be other factors influencing a go/no-go decision. Therefore, the recommended implementation alternative is presented as a set of interdependent decisions described below and shown in Figure 2.

1. If PennDOT has sufficient staff resources to build and maintain the PA STODW, then a build and operate in-house solution is preferred;
2. If not, then PennDOT should assess the market place for a mature provider of a suitable SAAS solution and procure such a service;
3. If the SAAS solution is infeasible, then PennDOT should contract with a work-for-hire systems integrator to build the PA STODW. The system could be maintained by either a contractor or by in-house staff.

![Figure 2: Recommendation Decision Tree](figure2)

If security, scalability, or capacity issues become significant design concerns, it is recommended that a cloud-based hosting solution be considered. While the industry experience with cloud-based computing is that there are no significant cost savings using this strategy, it does offer some advantages when used as a platform for data warehousing applications. The decision to host the PA STODW in the cloud will not affect either the in-house or contractor build implementation decision and is irrelevant if PADOT decides to pursue a SAAS strategy.
5.2. Implementation Approach

The ability to maintain flexibility in terms of how the PA STODW functional requirements are met and implemented is important. A progressive approach, understanding that the PA STODW will adapt and mature over time, will provide a method to gradually deploy the PA STODW in a manner that will test the proposed physical and logical methodologies. This iterative process allows for early identification and resolution of problems and will ultimately save time and effort during subsequent and larger deployments. Implementation should consist of the following maturity milestones:

- Roll-Out
- Base Year
- Planning Horizon

Roll-Out will include the initial design and deployment of the PA STODW to a small select group of stakeholders to test the validity of the system before expanding the user base to a larger number of stakeholders. This milestone would function as a “pilot”, where there is a release to a limited audience to confirm the functionality, ease of use, and be the mechanism for users to provide initial feedback. Feedback received needs to be reviewed and addressed accordingly.

The pilot would consist of a small PennDOT user base with a mixed skillset to both build and test PA STODW system functionality. The Concept of Operations has provided a breakdown and description of the five (5) subsystems, or components, of the PA STODW. The components are as follows:

1. Portal User Interface
2. Load and Update
3. PA STODW Administration
4. Browse, Search, Extract
5. Partially Integrated Repository – which includes the “Native” Data Archive, Staging Area, Production Area, and Metadata Catalog

Each component will have a list of functional requirements that could be initially designed, built, and tested in parallel. As part of the overall PA STODW design, component dependencies and relationships would be defined. As components are independently built, tested, and approved, connections to other components will be established and further validated for conformance until the entire PA STODW is operational in a test environment.

Roll-Out will have a one year duration for planning purposes and it will be followed by the Base Year maturity milestone which includes the release of the PA STODW to a substantially larger group of stakeholders and include an incrementally larger number of datasets. The Base Year milestone will also have a one year planning duration. During this timeframe, a focus on more proactive marketing and education to a large group of potential stakeholders will occur. Similar to Roll-Out, end user feedback received will be reviewed and addressed accordingly.
The Planning Horizon maturity milestone commences concurrently with the Base Year and concludes five-years upon completion of Roll-Out. The purpose of this period of time is to account for user feedback and enhancements identified during the Base Year and throughout the planning horizon along with concurrently conducting system maintenance and operating the system.

Table 4 provides an example set of system usage and sizing requirements that can be used for project planning and budgeting purposes for each maturity milestone.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Roll-Out</th>
<th>Base Year (RO+1 year)</th>
<th>Planning Horizon (RO+5 years)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of Data Sets</td>
<td>25</td>
<td>50</td>
<td>250</td>
</tr>
<tr>
<td>Number of Business Attributes</td>
<td>250</td>
<td>500</td>
<td>2500</td>
</tr>
<tr>
<td>Storage (GB)</td>
<td>10</td>
<td>25</td>
<td>100</td>
</tr>
<tr>
<td>Number of Participating Agencies</td>
<td>5</td>
<td>20</td>
<td>50</td>
</tr>
<tr>
<td>Number of Users</td>
<td>30</td>
<td>75</td>
<td>250</td>
</tr>
<tr>
<td>Average Browse and Search Sessions per Day</td>
<td>8</td>
<td>20</td>
<td>64</td>
</tr>
<tr>
<td>Average Uploads and Downloads per Day</td>
<td>1</td>
<td>4</td>
<td>8</td>
</tr>
</tbody>
</table>

Note that these milestones represent user adoption and usage thresholds although the Roll-Out and Base Year milestones also coincide with major software releases.

The following high-level implementation activities are provided for planning purposes and a planning-level implementation schedule is provided in Table 5. The schedule assumes two development phases resulting in at least two releases of the core software. The Roll-Out will provide core system functionality. The Base Year, system release, is estimated to occur about a year after Roll-Out, and will likely contain performance and functional enhancements. A more detailed Work Breakdown Structure should be developed during the project procurement process.

1. **System Definition and Design**: This activity defines detailed functional requirements of the system. This activity includes the development of the PA STODW system and subsystem requirements, verification plans, the detailed project architecture, and the detailed design specifications.
2. **System Development and Implementation**: This activity focuses on developing, verifying, and deploying subsystem and system components into the technical environment.
3. **System Enhancements, Changes and Upgrades**: This activity is responsible for assessing the system against the requirements, performing maintenance, developing, and implementing an upgraded version.
4. **System Operations and Administration**: This activity encompasses the normal day-to-day use of the production system.
5. **Data Governance**: This activity includes the development of data sharing agreements, data cleaning, data documentation, and other data-related tasks.
6. **Training:** This activity encompasses technical training for all intended users of the system including system administrators, data providers, and data consumers.

### Table 5: PA STODW Implementation Schedule

<table>
<thead>
<tr>
<th>Implementation Activities</th>
<th>Years From Authorization to Proceed</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.5</td>
</tr>
<tr>
<td>Roll-Out</td>
<td></td>
</tr>
<tr>
<td>System Definition and Design</td>
<td></td>
</tr>
<tr>
<td>System Development and Implementation</td>
<td></td>
</tr>
<tr>
<td>Base Year</td>
<td></td>
</tr>
<tr>
<td>System Enhancements, Changes and Upgrades</td>
<td></td>
</tr>
<tr>
<td>Operations Phase</td>
<td></td>
</tr>
<tr>
<td>System Operations and Administration</td>
<td></td>
</tr>
<tr>
<td>Data Governance</td>
<td></td>
</tr>
<tr>
<td>Training</td>
<td></td>
</tr>
</tbody>
</table>

This schedule assumes either the build in-house or contractor build option is selected
A SAAS based approach would have a different set of task and durations