FUELING THE PUV MODERNIZATION PROGRAM WITH DATA

Department of Transportation: PUV Modernization PMO

A Study on the Data Science and Strategy

Employed by the Department of Transportation's

PUV Modernization PMO

Rex Gregor Laylo
Paula Joy Martinez
John Cris Orenday
Jeonne Joseph Ramoso

In partial fulfillment of Data Science and Ethical AI (DASCI 213)
Of Master of Science in Data Science (MSDS) Program

Submitted to:

Erika Fille T. Legara, PhD
Asian Institute of Management

September 13, 2023

Executive Summary

The Department of Transportation (DOTr) stands on the precipice of effecting change that will redefine its policymaking capabilities. A look into the Project Management Office (PMO) of the Public Utility Vehicle Modernization Program (PUVMP) shed light on the nascent data practices within the Department and the gaps that a good data strategy can address, if properly implemented. This paper centers on a three-pronged data strategy encompassing aspects of people and culture, process improvements, and technological investments.

A diagnostic survey was conducted to align the organization's data strategy with its objectives and identify the key priorities and areas for collaboration within the PUVMP PMO. An in-depth interview then ensued to obtain granular insights into the existing data-driven initiatives, organizational objectives, and the roles of internal and external stakeholders. This elaborated the technological capabilities, overall sentiment towards data science and analytics, ongoing and planned data initiatives, business cases, and organizational structure.

A Data Science Maturity Model was then designed and utilized to comprehensively evaluate the PUVMP PMO's data capabilities. This model, drawing from established frameworks, assessed technical and organizational aspects, cultural acceptance, conceptual knowledge, and strategic planning. The PMO obtained a 'Foundational' rating (2.74) in the model, indicative of its existing foundation for data-driven initiatives. The key findings revealed strengths in organizational structure and cultural acceptance, but identified necessary improvements in data infrastructure, data management, data governance, data analytics, and strategic planning.

Guided by three-faceted data strategy, optimizing processes, fostering a data-centric culture, and maximizing the right technology has the potential bring forth better policy decisions and operational efficiencies within the Department of Transportation.

1 Introduction

What does it mean to be truly alive? As human beings, we have our body systems performing specific functions and one of them is the circulatory system. Among its roles is to ensure our blood's continuous flow to deliver oxygen and nutrients through our veins. In much the same way that our body relies on a robust circulatory system, a thriving society is equally contingent on an efficient transportation network. This fundamental requirement transcends individual backgrounds and occupations, and the persistent, and often vexing, absence of this efficient system is what society has to grapple with on an ongoing basis.

The 2022 Public Transit sub-index
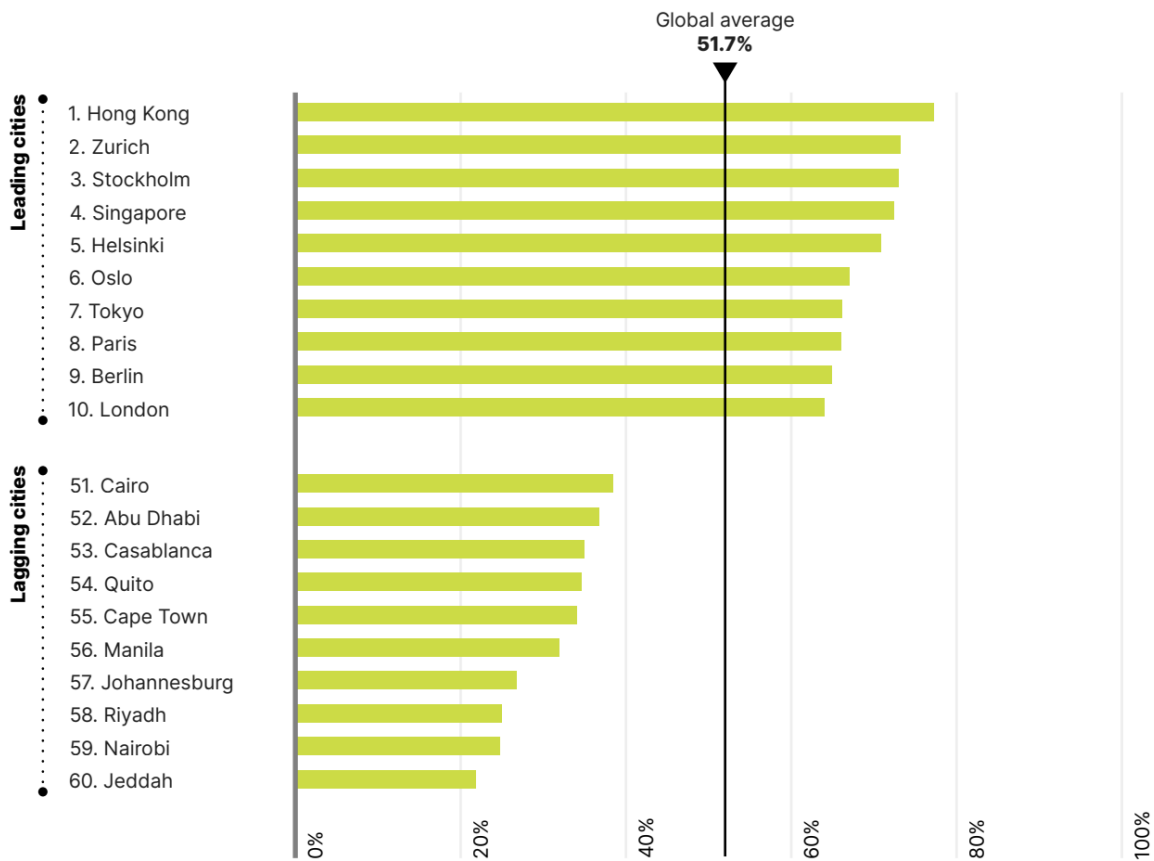
Source: Oliver Wyman Forum and University of California, Berkeley analysis

Figure 1. Ranking of 60 Cities on the Public Transit sub-index of the Urban Mobility Readiness Index

However, we do not live in a utopian world, and we are confronted by the realities of the issues of our transportation system. In 2022, the Oliver Wyman Forum and the University of California published a report about the Urban Mobility Readiness Index. It assesses various factors influencing the quality of mobility in urban areas. In the public transport component of the study, Manila ranked 56th out of the 60 cities included in the research. This underscores the need to improve Manila's public transport system when compared to its counterparts in the study, and it

also reflects the challenges faced by the National Capital Region and other urban cities of the country. However, it is important to acknowledge that the study's evaluation does not cut across an entire country, which can obscure the disparities in transportation infrastructure if there are remote rural areas where public transport is virtually non-existent.

As future data scientists driven by a collective passion for social responsibility, we pose a fundamental question: How can we contribute meaningfully to the improvement of the transportation sector? There is no more powerful means of effecting change within this sector than by collaborating directly with the policymakers responsible for its oversight, the Department of Transportation (DOTr)

1.1 About the Organization

Established in 1899, the **Department of Transportation** serves as the central authority within the executive branch of the Philippine government in crafting policies, creating plans, and overseeing the implementation and administration of an integrated and reliable transport system. In pursuit of this objective, the DOTr has outlined its organization's vision and mission.

Vision
By 2030, the DOTr is a world-class organization, providing integrated transport, connecting people, islands, families, communities, and the nation with the rest of the world, and constantly responding for environmentally sustainable and global competitive transport.

Mission
To provide the country with efficient, effective, and secure transportation systems that are globally competitive, compliant with international standards, and responsive to the changing times.

There are four sectors under the DOTr: Aviation, Railways, Maritime, and Road. Under the Road sector are five subsectors: the Land Transportation Franchising & Regulatory Board (LTFRB), the Land Transport Organization (LTO), the Office of Transportation Cooperatives (OTC), the Toll Regulatory Board (TRB), and the Program Management Office (PMO), which is the focus of this paper.

1.2 PMO: Driving Progress and Innovation

The Program Management Office is a specialized unit under an agency established for the implementation, coordination, and monitoring of project activities. It complements the existing roles and responsibilities of a government agency by creating projects aimed at driving growth and innovation. According to the PMO's National Budget Circular Rationalization of 2003, it has the following key functions that encompass the project lifecycle:

- Project Operations Planning
- Project Monitoring and Evaluation
- Project Coordination
- Project Operation and Management

- Troubleshooting
- Financial Management
- Coordination with Various Stakeholders
- Post Project Completion / Mainstreaming of Project Concerns in Regular Activities

The core structure of a PMO comprises of administration, finance, project operations and management, and project monitoring and evaluation. What holds these structures together in terms of delivery and effectiveness is the influence of data, which has a vital role in informing decision-making processes across all facets of a PMO's operations. In essence, the influence of data is present in every aspect of a PMO's structure, enhancing its ability to deliver results, drive efficiency, and produce tangible value for the organization.

As our research progressed, our focus gravitated towards a specific PMO within the DOTr: the PUV Modernization Program.

1.3 PMO Spotlight: The PUV Modernization Program

Launched by the DOTr in 2017, the Public Utility Vehicle Modernization Program (PUVMP) is a transformational large-scale initiative that aims to holistically transform the road transportation system by addressing the following:
- Vehicle Safety and Quality
- Route Network Efficiency
- Fleet Management

The PUVMP aims to modernize our jeepneys in adherence to safety, environmental, and transport standards defined by the organization. To do this, the PUVMP has 10 components:

1. **Initial Implementation:** The Program is inaugurated, and its foundational components, processes, and structures are put in place to support project management activities within an organization.
2. **Regulatory Reform:** For the review of existing policies and introduction of new ones in line with the vision of the program.
3. **Local Public Transport Route Plan (LPTRP):** Route rationalization by the LGUs to address local mobility requirements.
4. **Rationalization:** To rationalize existing routes and propose new routes based on passenger demand, road hierarchy, and road capacity.
5. **Industry Consolidation:** Strategic merging of operators into a consolidated entity; Engagement of drivers and other transport workers as regular employees.
6. **Fleet Modernization:** For the replacement of old, dilapidated PUVs with Omnibus Franchising Guidelines (OFG)-compliant PUVs; formulation of standards for PUVs.
7. **Financing:** Partnership with financing institutions and provision of subsidies to enable the transport sector to comply with the program.
8. **Vehicle Useful Life:** Ensures that old PUVs are taken off the road, surrendered and its components are properly and responsibly disposed of.

9. **Social Support:** To provide a social safety net to the drivers and operators who may be displaced because of the Route Rationalization and LPTRP.
10. **Communications:** To Link stakeholders to the necessary information about the PUVMP.

1.4 The Core Strategic Imperatives and Objectives of the PUVMP

In the context of policymaking, the PMO acknowledges the integral role of data in the successful implementation of the PUV Modernization Program, which necessitates the pursuit of the following best practices:

- **To establish a clear connection between the components and stakeholders of the PUV Modernization Program**, which centers on the institutionalization of a consolidated database that would act as the single source of truth;
- **To effectively monitor the compliance, progress, and performance of all the PUVMP components**, which involves crafting effective policies to safeguard the integrity of its decision-making processes;
- **To optimize existing processes,** primarily through a digitalized and real-time system for data collection monitoring, and analysis, and;
- **To outline a clear hierarchy of accountability in all aspects and processes of the PUV Modernization Program**, which involves delineating roles and responsibilities governing data-related activities

However, similar to many organizations, the PUVMP PMO grapples with challenges that hinder the materialization of its data-driven objectives and strategic imperatives.

1.5 PUVMP Data Management Problems at a Glance

1.5.1 Manual Data Collection

The PUVMP PMO receives data in a standardized form from various LGUs via e-mail. The team manually encodes the data in a Google Sheet. For LGUs with no internet, they resort to accomplishing a physical form and sending it via snail mail to the program's head office. These physical forms are manually entered by the PUVMP team upon receipt.

1.5.2 Data in Silos

Other sources of the PUVMP's data are databases from other government agencies such as the LTFRB. There is no interface that would allow the required data to automatically flow to the database of PUVMP. Since there is no immediate connection, the data captured from LTFRB may soon be outdated causing discrepancies in the data.

1.5.3 No Defined Frequency on Data Update

There is no cadence for regular database maintenance, cleansing, and updates. There are cases where the team performs the updates on an ad-hoc basis preventing them from doing real-time data analysis.

1.5.4 Budget and Resource Limitations

Budgetary constraints and limited personnel diminish the ability of the PMO to fully leverage on the enterprise software and other tools that would improve its data management and analytics capabilities.

2 Methodology

In this section, we discuss the strategies and techniques we employed to gather relevant data about the organization. This involves quantitative and qualitative methods.

2.1 Survey on Data Strategy Position

The diagnostic questions serve to assess the alignment between the organization's existing data strategy and objectives. This insight will help us pinpoint the team's key priorities and enable us to enhance the effectiveness of our collaboration.

DATA DEFENSE	Score
1. Reduce general operating expenses	
2. Meet industry regulatory requirements	
3. Prevent cyberattacks and data breaches	
4. Mitigate operational risks such as poor access controls and data losses	
5. Improve IT infrastructure and reduce data-related costs	
6. Streamline back-office systems and processes	
7. Improve data quality (completeness, accuracy, timeliness)	
8. Rationalize multiple sources of data and information (consolidate and eliminate redundancy)	
Total	
	Score
1. Improve revenue by adjusting pricing strategies and executing initiatives within the Department of Transportation	
2. Create new products and/or services	
3. Respond rapidly to shifts, changes, and challenges in the transportation landscape	
4. Use sophisticated customer analytics to drive operational outcomes and strategic decisions within the Department of Transportation	

5. Leverage new sources of internal and external data		
6. Monetize organizational data assets (potentially offer data-driven products and/or services)		
7. Optimize existing strong bench of analysts and data scientists		
8. Generate return on investments in big data and analytics infrastructure		
Total		
Overall data strategy:		

Table 1. Data Strategy Position Survey

2.2 Interview

While the initial Data Strategy Position Survey immensely helped ground our understanding of the PMO's inclinations, the more granular information about the existing data-driven initiatives, organizational objectives, and roles and responsibilities of the internal and external stakeholders were obtained through an in-depth interview with project leaders.

2.3 Probing into the Data Science Maturity of the PUVMP PMO

The valuable insights collected during the interview served as our framework for evaluating the Department's progress in the field of data science and analytics. Some of these information included the following:
- Current technological and analytical capabilities
- Sentiment of the entire PUVMP PMO towards data science
- Data initiatives and strategies completed, or in the pipeline
    - Number of deployed initiatives
    - Ongoing development and testing
    - Plans
- Presence and clarity of business cases on data-driven projects
- Knowledge and understanding of data science concepts and innovations
- Organizational structure and ways of working

3 DATA SCIENCE MATURITY ASSESSMENT

3.1 DATA SCIENCE MATURITY MODEL

The team created its own data science maturity model that combines the components and methods of existing ones. The objective of the model is to establish comprehensive criteria and classification levels to evaluate our chosen PMO more holistically.

Our model is a modified version of the *Consensus Composite Framework* developed by Arcalea, a marketing consulting firm specializing in data science and AI utilization. Arcalea examines the applications of various models in diverse domains and identifies shared sub-components to develop a universal assessment criterion. The sub-components include: (1) Organization, (2)

Infrastructure, (3) Data Management, (4) Analytics, (5) Governance, and (6) Best practices (Arcalea, n.d.).

While some elements from Arcalea's model are present in our framework, our model distinguishes itself through a nuanced categorization and grouping of these components. Furthermore, our model incorporates additional elements that Arcalea's model does not encompass. These supplementary elements are derived from the definition of AI maturity outlined in the study entitled: "Reshaping Business with Artificial Intelligence," conducted by the MIT Management Review in collaboration with The Boston Consulting Group. One of the measures considered in the study was the level of understanding of businesses on artificial intelligence - concepts, implementation, and impact. We aimed to incorporate a similar measure into our criteria, but within the broader context of data science instead (Massachusetts Institute of Technology, 2017).

3.1.1 MATURITY ASSESSMENT CRITERIA

The team looked at four categories in evaluating the maturity of the PUVMP PMO:

TECHNICAL & ORGANIZATIONAL

Looks at the institution's **data infrastructure, management, governance, and analytics** along side it's **organizational structure**

CULTURAL ACCEPTANCE

Evaluates how much the entire organization's, including management's, **belief in the value of data science and analytics**

CONCEPTUAL KNOWLEDGE

Assess the organization's **grasp of implementation effort, current trends, and potential impact** in the context of data science

STRATEGIC PLANNING

Check if they have a **defined purpose for utilizing data science** and if there are **established roadmaps and business cases**

Figure 2. Four Major Categories for Assessing Data Science Maturity

3.1.1.1 Technical and Organizational

This category assesses the PMO's data infrastructure, data management practices, data governance procedures, analytics capabilities, and organizational structure (Arcalea, n.d.):

- **Data infrastructure** refers to the systems and tools used in the company's operations, any software and hardware they may have to collect, store, and transform data.
- **Data management** looks at the scale of the data stored in the databases of the PMO. It asks whether the size, complexity, and granularity are enough to create value. Furthermore, the component also highlights the existing data architecture at the office. Are their data collection, scraping, and storage activities efficient and complete enough to ensure accuracy?
- **Data governance** is an amalgamation of the following: (1) Data access and lineage - the reach of data to various stakeholders, (2) Data catalog - clarity in how each data point is defined, (3) Data security and privacy - protection from potential risks in data, (4) Data

control - level of influence on data processes, and (5) Data integrity - accuracy and completeness of data.

- **Data analytics** refers to how well the PMO utilizes data to produce insights and analysis to boost decision-making. In the context of the assessment, it refers to the capability of the Office to transform raw data into useful information.
- **Organizational structure** probes into the coherence of teams within the PMO, the alignment of its personnel with their roles, and the fluidity of communication and decision-making processes. It looks into the presence of roles within the PMO that are ideally positioned to take on the crucial responsibilities surrounding data-related activities, and it scrutinizes the clarity with which these roles and responsibilities are delineated, especially in the management of both current and prospective data processes.

3.1.1.2 Cultural Acceptance

The team evaluated cultural acceptance based on the sentiments of the PUVMP PMO team toward data science, and their readiness to go through change management to shift to a more data-driven approach. Moreover, cultural acceptance extends to the leadership team's (PUVMP PMO head and road sector undersecretary) commitment to data science initiatives - do they consider data projects as one of their priorities? In summary, here are the key considerations under cultural acceptance:

- **Organizational Culture:** The PUVMP PMO's belief in data science, and ease of Change management
- **Leadership:** Upper management's (PUVMP PMO head and road sector undersecretary) commitment and faith in employing data science in their day-to-day operations

3.1.1.3 Conceptual Knowledge

This category mainly focuses on the Office's level of understanding in the context of the implementation efforts involved in implementing data science, the dynamic landscape of data-related technologies and strategies, and the potential impact of embracing a data-driven approach:

- **Implementation effort:** Refers to the PMO's understanding of the technological and cultural challenges associated with incorporating data science into its operational and decision-making processes.
- **Innovation and Trends**: Refers to the awareness of the PMO about new data practices and technologies. Are they considering implementing tools, strategies, and best practices being used by first-world transportation sectors?
- **Impact and Value**: This investigates whether the PMO has a clear idea of the potential positive outcomes that could stem from the successful integration of data science. Has the PMO considered the effect on multiple aspects such as people, processes, and technology?

3.1.1.4 Strategic Planning

In terms of strategic planning, the team assessed whether there already exists a roadmap or business case that lays out the PMO's purpose in having data science, as well as the clarity of the PMO's short-term and long-term direction within its context.

- **Roadmap and business case:** Framework or documentation that formalizes the plan for data science in the PMO.
- **Clarity of purpose for data science:** The PMO's mission and vision for data science.

The team assessed each category using a rating system similar to a Likert scale, where scores range from 0 to 5. To provide a clearer understanding of how the ratings function, please refer to the accompanying visual below:

Note: The rest of the ratings are relative to 0, 1, and 5.

Figure 3. Rating Scale for Major Categories

3.1.2 MATURITY CLASSIFICATIONS

Upon rating each of the four major categories within our model, the team calculated their average. This average will serve as the basis for determining the PMO's placement within our maturity classification groups. The use of the average underscores that the team assigned equal significance to all categories when evaluating maturity.
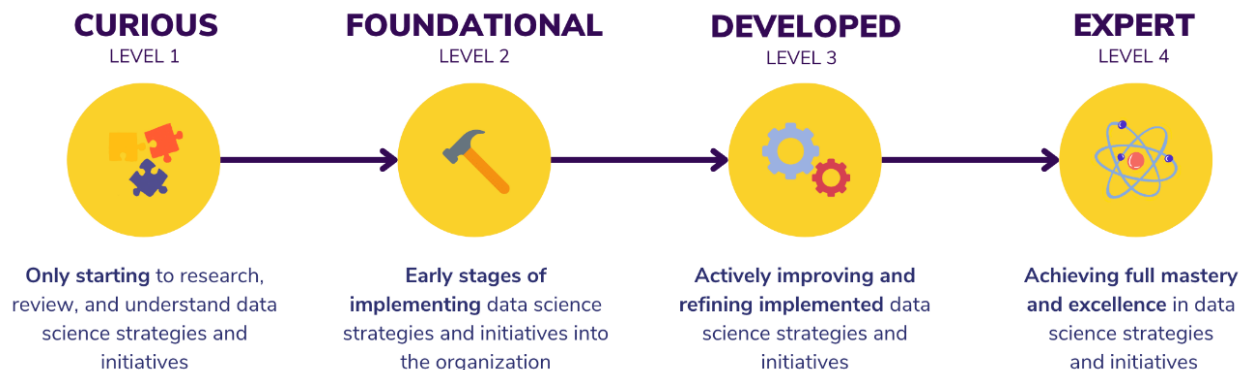
Figure 4: Data Science Maturity Classification Levels

Curious or level 1 (average rating of 0.1 to less than 2): This is an organization that is just beginning to explore the potential of data science as a valuable asset for its operations. Furthermore, these organizations are in the process of gathering information about potential applications and the impact of data science on their business.

Foundational or level 2 (average rating of 2 to less than 4): This organization is in the initial phases of putting data science strategies and initiatives into practice. Additionally, they have acquired sufficient resources and tools to construct and integrate data-driven approaches.

Developed or level 3 (average rating of 4 to less than 4.5): This organization has already integrated data science into its operations. However, they still need to actively improve and refine some elements to become one of the leaders in data science implementation in their own domain.

Expert or level 4 (average rating of 4.5 to 5): This organization has attained a state of complete excellence and mastery in the field of data science. They can be regarded as industry leaders for their exceptional implementation of data-driven approaches in their business.

3.2 FINDINGS

2.74

OVERALL CLASSIFICATION

FOUNDATIONAL

LEVEL 2

1.49

TECHNICAL AND ORGANIZATIONAL

4.88

CULTURAL ACCEPTANCE

2.83

CONCEPTUAL KNOWLEDGE

1.75

STRATEGIC PLANNING

Figure 5. PUVMP PMO's Data Science Maturity Rating

The PUVMP PMO received a rating of **2.74** which classifies them as **'Foundational'** based on our maturity model. The PMO is not starting from scratch - they have existing data tools and processes

at their disposal. These resources can be gradually utilized to introduce more and more data science strategies and projects into the day-to-day operations of the PMO.

3.2.1 Technical and Organizational (Rating: 1.49)

3.2.1.1 Data Infrastructure (Rating: 1)

The PUVMP PMO solely relies on G Suite applications and MS Excel to execute data-related processes. While these tools have their merits, their use may present potential risks to data security and could also restrict the full capabilities of their current data processes. Moreover, the infrastructure surrounding their main data tools limits them. They have no internet connection - their staff relies heavily on personal data consumption to access software. Also, they address data management and reporting aspects but lack the necessary tools for data collection and data analytics.

3.2.1.2 Data management (Rating: 2)

When it comes to the volume of data employed by the PMO, it is quite substantial - they gather various data points from multiple external organizations and over 1,500 LGUs. Based on their Excel database, the collected data is granular and complex enough to enable some analytics, particularly descriptive analytics, but it may not support predictive and prescriptive analytics.

A positive side to their data architecture is that they have a standardized template that they give out to LGUs to have some uniformity in the data that they store in their database. Although this is a commendable best practice, they still have a long way to go. There are still LGUs that opt to send their data through physical copy, which might be subject to human error. Furthermore, data obtained from LGUs represents just a portion of their database. They also gather data by scraping from the databases of partner organizations like the LTFRB and LTO. However, the challenge lies in the fact that despite the PMO's commendable database organization, data accuracy is compromised by the inaccuracies present in their partner organizations' data.

Overall, the PMO has foundational processes that characterize a bare minimum data architecture. Regarding data flow, efficiency remains a challenge due to their reliance on external data sources and manual processes. These limitations are most pronounced when it comes to their data collection and data analytics activities.

3.2.1.3 Data Governance (Rating: 1.2)

The PMO faces constraints in terms of access and control over their source databases, which are external and owned by their stakeholders. Their access is predominantly limited to data extraction purposes. In case issues arise, they possess minimal authority to drive changes in data cleaning and updating efforts within their partner organizations. While they have a greater degree of

influence over data from LGUs, the challenge lies in the fact that exerting this influence and control is a time-consuming process, primarily because LGUs often have restricted internet access and rely on manual processes.

The stakeholder databases that the PMO extracts from are updated sporadically, potentially attributed to the continued use of legacy systems that their personnel are no longer familiar with. During our interview with the PUVMP team, they disclosed that one of their stakeholders' databases had not been updated for the past two months. This situation underscores a significant likelihood that a substantial portion of the data they acquire is inaccurate and of poor quality.

Despite working with flawed data points, the KPIs and metrics that the PMO have developed are impressive and sufficient for generating analyses that provide a snapshot of their progress within the PUVMP PMO. The team recognizes that the only missing element in their data points, KPIs, and metrics is the absence of formal documentation. Without a structured data catalog, there is a possibility of integrity issues arising as the database transitions from one owner to another. The standards set by the original owner may become distorted as new employees take charge of the database.

Regarding data security and privacy, the PMO heavily depends on G Suite tools for their data-related activities, and these tools provide minimal safeguards against potential risks and cyberattacks. An article by Kate C from MSP360 highlights that "Google G Suite is only as secure as the accounts that your users rely on to access the platform. If their accounts are compromised, there is little you can do to prevent their sensitive documents from being stolen or deleted."

In summary, the team acknowledges that the PMO has implemented certain best practices for maintaining data quality. They utilize a standard template during data collection and maintain a well-organized database with established KPIs and metrics. However, these practices alone are insufficient to earn a high score, primarily due to their limitations in data access, lineage, control, and security.

3.2.1.4 Data Analytics (Rating: 2)

The team can construct descriptive analytics based on the available data, but they are unable to progress further. The lack of reliable data integrity and irregular updates from external sources restricts them from engaging in predictive and prescriptive analytics. Consequently, their output is limited to generating reports with minimal value for politicians and senior staff and creating informative primers.

Aside from their limitations in data, the team also identifies skill as a possible restriction to building effective data analytics. Upon reviewing their database, it becomes evident that they offer numerous metrics and indicators; however, there is little to no data visualizations to facilitate a more comprehensive analysis of trends and relationships within the data.

The team acknowledges that the primary focus of the PMO is policy development, but it is evident that data analytics should also play a significant role in their operations. It can be said that they have the necessary tools and metrics readily available; the key challenge is to overcome their existing limitations. Skill gaps can be mitigated through training, while data constraints may find resolution through process agreements with external stakeholders.

3.2.1.5 Organizational Structure (Rating: 1)

The PUVMP PMO is well structured, with designated leads overseeing each component and multiple supporting staff members dedicated to facilitating specific activities within each component. Furthermore, in addition to their expertise in their respective domains, these team members have adequate exposure to data through their current processes, such as data collection, management, and reporting. This indicates the presence of potential roles where data science responsibilities can be seamlessly integrated.

Even though they have a very capable team, the team believes that the PMO falls short of defining clear roles and responsibilities for data-related processes. This lack of clarity can lead to confusion regarding accountability when issues arise in their data processes. Clearly defined roles and responsibilities are essential for achieving efficiency in executing data-driven strategies and activities.

To summarize, while there are already candidates within DOTR who can fill roles as data stewards and domain experts for each PUVMP component, the missing piece is the establishment of explicit roles and responsibilities regarding ownership of the data processes.

3.2.2 Cultural Acceptance (Rating: 4.88)

Disclaimer: The provided rating is specific to the PUVMP PMO and does not represent the entirety of the DOTR organization. Additionally, it is important to clarify that this rating is an approximation of the PMO's sentiment, as it is derived solely from feedback provided by those who attended the interview only. To obtain an accurate assessment of the PMO's sentiment, the team would ideally need to conduct a comprehensive office-wide survey. However, this was not carried out due to constraints related to timing.

During our interview, it became evident that the entire team is enthusiastic about initiating and advancing data-driven projects. They clearly recognize the significance of integrating data science into their tools, strategies, and processes. Notably, this sentiment is also shared by the management, particularly the PUVMP head and the road sector undersecretary. They have even expressed interest in engaging MSDS (Master of Data Science) students to develop a capstone project based on their existing processes.

While this conclusion may not represent the sentiments of the entire department and PMO, it can be inferred that, based on their sentiments, they may experience smoother change management when integrating data science into their processes. The PMO recognizes the value of data science, leading the team to believe they will be more open and receptive to any digital changes.

3.2.3 Conceptual Knowledge (Rating: 2.83)

As a review, conceptual knowledge refers to the grasp of the PMO on the following items related to data science: (1) implementation effort, (2) innovations and trends, and (3) potential impact or value.

3.2.3.1 Implementation Effort (Rating: 3)

In terms of understanding the implementation effort, they recognize the complexities associated with the PMO's extensive scale and scope in relation to integrating data science tools and strategies. However, the team believes they may not be giving due consideration to other factors, such as the effort required to harmonize human and automation elements and the effort involved in change management, along with shaping the overall office culture and ways of working.

3.2.3.2 Innovations and Trends (Rating: 2.5)

While they do possess a reasonable awareness of current trends in data science, their immediate needs tend to narrow their view. Their primary attention is presently directed towards database management and analytics, which means their industry knowledge primarily revolves around these domains. This limited focus has potentially led to a lack of awareness regarding other tools that could be implemented in the future, once they have progressed beyond their current requirements. For example, they are advocating for the adoption of an enterprise system, yet they appear to overlook the subsequent tools that should follow, particularly those aimed at extracting value from the raw data stored in their systems. It is essential for the PMO to start considering what analytical tools can complement and enhance the capabilities of the enterprise system.

3.2.3.3 Impact and Value (Rating: 3)

The PMO has some understanding of the significance and worth of data initiatives in bringing positive enhancements and transformations to their operations. However, the team suspects that their concept of value and impact may not be comprehensive, potentially constrained by their current level of exposure to the field of data science, particularly due to the absence of formal training in this domain.

3.2.4 Strategic Planning (Rating: 1.75)

The PUVMP PMO has a well-defined idea of where they want to take data science in the context of their operations. This is apparent in the business objectives for data that they have communicated with the team. Mainly, they want data science solutions and strategies: (1) to monitor the effectiveness of policies safeguarding the welfare of all stakeholders under the PUV Modernization Program and the compliance of PUVs to sustainability standards as defined by the Department, (2) to digitalize data collection, monitoring, and analytics processes, and (3) To establish clear roles and responsibilities to the team on data processes.

What significantly contributed to their low score was the absence of a formalized purpose for incorporating data science. They lack roadmaps and business cases, essential tools that would provide guidance and direction when they encounter obstacles at various decision points during the implementation of data science processes.

It is important to note that all of the ratings provided stem exclusively from the team's high-level interview with the PUVMP PMO. This assessment only represents an initial evaluation of the PUVMP PMO and may undergo changes as the team gains a comprehensive understanding of the office and its processes.

4 PROPOSED DATA STRATEGY

_____

The main strategy position of the PUV Modernization PMO is to improve the integrity of its data within its existing repositories, particularly those employed in generating reports. Its primary objectives are centered on standardizing, consolidating, and governing data assets and sources.
_____

Figure 6. The Desired Data Strategy Position of the PUVMP PMO

Taking into account the current state of the PUV Modernization Program in the context of data maturity, the PMO's overall desired strategy position (Figure 6), and its vision to embed data science and analytics into its processes while acknowledging the associated operational constraints, we anchor our proposed data strategy on three pillars: process, people and culture, and technology.

Figure 7. Three Pillars of the Proposed Data Strategy

The overarching goal of the PUVMP PMO is to be able to use *data as an asset*, i.e., to use data to create value consistently. While the PMO has an existing practice of utilizing data to inform decisions and policies, not to mention the strong cultural acceptance and enthusiasm from its members, the bulk of the challenge is organizational in nature and can be summarized by this optimization problem: "Given its limitations, what investment(s) should the PMO prioritize to employ data more effectively into its operations?" Using the three building blocks, the paper attempts to provide a comprehensive and practical approach to the optimization problem presented.

4.1 PROCESS: Defining Structures and Foundations

Nearly every decision made in life relies on information in some way. Because even erroneous data can be used to support decisions, albeit with potentially catastrophic consequences, what takes precedence in decision-making is the validity of data to justify certain choices, more than anything else. The policymaking process within the Department works in a similar vein but has greater costs if not done correctly. Thus, it is crucial to establish standards and guidelines within the PUV PMO to uphold the integrity of its data, and in doing so, the outcomes grounded on it. Keeping this in mind, the backbone of data governance within the PMO must be anchored on rules tackling data quality and security.

The standards on data quality should encompass the completeness, accuracy, and consistency of data with the goal of minimizing wasted resources allocated on data cleanup that could otherwise be utilized to maximize data-driven opportunities. This may already be considered a scalable practice within the PMO given that it had laid the groundwork of setting a uniform data format in one of the ten components in the PUV Modernization Program. However, it is worth noting that a quality-assuring governance can vary depending on the type of data being managed—because the Program deals with a wide range of datapoints across many aspects, the definition of a "good data" may differ depending on the specific project component involved.

Nevertheless, there must be a unifying set of guidelines that standardizes all datapoints and makes them suitable for analysis. A useful monitoring tool for this is a data catalog (Figure 8) that summarizes the characteristics of all data being observed and how it relates to other data.

Variable	Description	Data Type	Source	Number of Rows	Number of Missing Values	Minimum Value	Maximum Value
region	Lorem ipsum dolor sit amet	str	LGU	1000	1	CAR	Region XIII
lgu	Lorem ipsum dolor sit amet	str	LGU	1000	100	Ana	Paula
classification	Lorem ipsum dolor sit amet	str	LGU	1000	50	Alonzo	Vasquez
draft_number	Lorem ipsum dolor sit amet	int	LGU	1000	75	25	40
status	Lorem ipsum dolor sit amet	str	LGU	1000	5	approved	pending for review

Figure 8. A Sample Data Catalog

On the other hand, guidelines on data security should be focused on protecting data privacy and defining clear accessibility rules to dictate who can access specific data and under what circumstances. One practical application of this is in the context of collecting the route plans of provincial LGUs, where the PMO wants to prohibit the access of LGUs to route plans other than their own. We further extend this and recommend control measures that would identify the specific personnel and ensure that only those with authorization can access certain information. For example, data access may only be granted to email addresses with a certain domain, or there is a master list of authorized users who are permitted to view and make changes to certain datasets.

To fully utilize data as an asset, all principles concerning data quality and data security must be reinforced with measurable and reasonable key performance indicators (KPIs), which are pivotal to iteratively monitor and steer a successful implementation of data governance across the PMO. For instance, measuring data completeness may involve looking into the percentage of missing values and delving into strategies to reduce this deficiency, whereas data consistency could be in terms of formatting and structure within and across datasets. On the other hand, a practical method to measure data security may involve investigating the number of times that datasets were accessed by unauthorized entities.

Recommended Next Steps for the Department of Transportation

1. *Define what "good data" means across all the ten components of the PUV Modernization Program.*
2. *Craft rules and guidelines based on this definition. Consider the expected format of the data.*
3. *Consider the access restrictions that the PMO would like to impose on its data assets, and how these restrictions would be imposed. As an example, you may start with restricting access based on email domains. You may list down the names of authorized users.*

4.2 PEOPLE & CULTURE: Embedding Best Practices

While having robust process to uphold data quality and security is a solid foundation for establishing data governance, their effectiveness could be undermined without strong leadership to initiate and sustain these efforts – this is where the People & Culture building block becomes critical. People & Culture comprises of two key components: data ownership and data culture.

4.2.1 Data Ownership: It Takes the Right Leadership

Data ownership encompasses the structure of governance over the PMO's data assets. Currently, multiple sources produce the datapoints being used by the PUV PMO, including 1500+ LGUs, the LTFRB, private and commercial banks, and manufacturers. Knowing that the PUV Modernization Program tracks 10 components, each having its own set of datapoints, what is the most practical strategy to handle these datapoints with limited personnel?

Our vision entails that every component will be led by a *data steward* responsible for establishing data quality and security standards tailored to the specific data points observed within each component. Given that the Department of Transportation serves as the policymaking authority of the PUV Modernization Program, the role of data stewardship is instrumental in ensuring that downstream decisions are accurate and reliable.

When executed properly, data stewardship would promote confidence around the reliability of the data being used to craft policies. Data stewards are also expected to uphold the security of data. They must be able to classify which sets of data are most vulnerable to breaches, damage, theft, and other external attacks, then identify the appropriate responses based on the nature of the exposed data and its relationship with other data (this is where having a comprehensive data catalog pays off). Simultaneously, each data source must be under the supervision of a *data owner*, who will ensure the integrity and consistency of their data while adhering to the guidelines established by the data stewards.

> **Recommended Next Steps for the Department of Transportation**
>
> 1. *Define distinct roles and responsibilities for a data champion, data steward, and data owner.*
> 2. *In consideration of the limited personnel, appoint data stewards and one data champion that would oversee them. Collaborate with external sources (e.g., LGUs and the LTFRB) to designate data owners.*

4.2.1 Data Culture: The Democratization of Data Capabilities

By this point, the PMO would already have the foundations of good data governance: a set of data management standards and the right people to enforce it. However, these may not be enough for the PMO to harness the true value of its data without the democratization of the right management and analytics capabilities that align with the overall objectives of the Department. In the similar way that effective leadership within the PMO assures the dependability of its data as basis of policy decisions, there is still a need to solidify the belief and deepen the understanding of people in the ability of data to deliver solutions.

Our suggested approach revolves around improving data literacy within the PMO, beginning with an assessment of the employees' current skillset in the context of working with data. Based on an individual's expertise, or lack thereof, they may be classified as one of the following:

- **Data Explorer**: Individuals who are in the early stages of developing their data skills and only has basic understanding of data concepts
- **Data Practitioner**: Individuals who are proficient in data management, analysis, and visualization, and who can effectively work with data but may not have expertise in advanced analytics.
- **Data Expert**: Individuals who are highly knowledgeable and experienced and have advanced skills in data management, analysis, visualization.

Based on this assessment, the PMO should establish a continuous process of upskilling and re-skilling that prioritizes competencies aligned with its objectives, such as data management, analysis, and visualization. It's crucial to emphasize that this assessment should be conducted on a regular basis to proactively fine-tune the approach in accordance with the PMO's evolving goals and priorities.

Recommended Next Steps for the Department of Transportation

1. *Design an assessment tool to evaluate the skills of employees against the PMO's specific operational requirements.*
2. *Conduct an assessment using this tool for all employees.*
3. *Initiate upskilling and/or re-skilling sessions based on the results of the assessment and the priority skills.*
   - *The PMO may consider potential collaborations with external organizations, such as the Asian Institute of Management, to support this endeavor.*
   - *Internally, the PMO may also conduct brownbag sessions to foster knowledge sharing and skill development among its team members, focusing on areas where expertise can be shared and cultivated.*

5 TECHNOLOGY: Tooling-up the Department of Transportation

The conducted evaluation of the extent to which the PUVMP PMO is able to utilize its data exposed the insufficiency of its system and tools to collect, store, and transform data effectively and efficiently. Taking into consideration both this limitation and the budgetary constraints faced by the PMO, we designed a prototype for a central repository that can be adopted to improve the PMO's operational efficiency, whilst maximizing the tools at its disposal. The successful implementation of this initiative relies heavily on the use of Microsoft Excel and the Visual Basic (VB) programming language to automate the transfer, monitoring, and formatting of incoming data (Figure 9).

The primary advantage of these tools stems from their cost-effectiveness and seamless integration with the PMO's current operations. Because they are already well-acquainted with MS Excel, the learning curve for VB is minimal compared to other programming languages. By harnessing MS Excel and VB, we can streamline the data consolidation process across the PUVMP group, which culminates in the creation of a single source of truth that unifies data sources, formats, and structures (Figure 10).
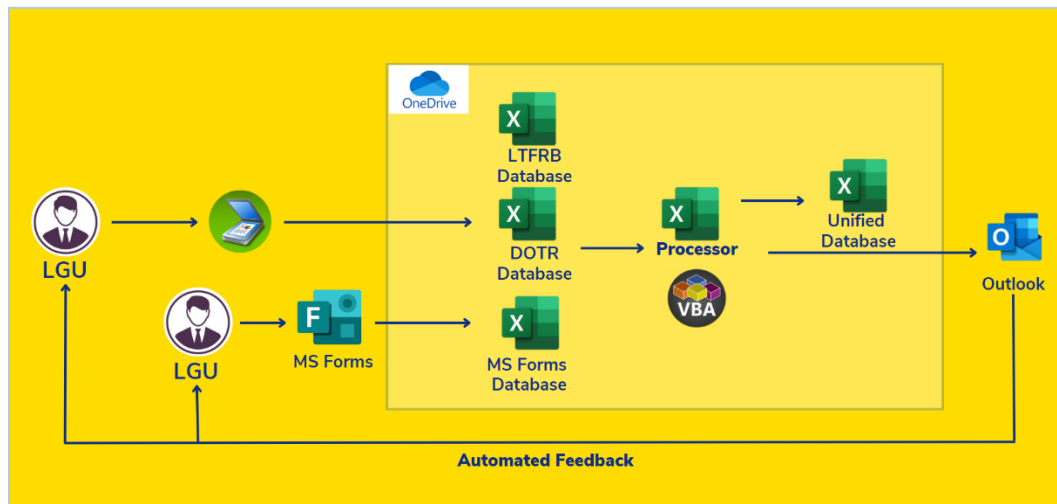
Figure 9. Data Tools and Lineage Summary

Figure 10. Processor

5.1 Data Collection

The process of consolidation begins with data collection. As of writing, there are multiple data sources within the PUVMP framework, but our focus will only be on the PMO's primary sources: the local government units (LGUs) and the LTFRB. While there is an existing shared database between the LTFRB and the DOTR, we see the merit of redesigning its structure for better consolidation and data cleansing and monitoring capabilities, which will be facilitated using Microsoft 365 Applications. To safeguard the central database from the operational constraints confronting its data sources, we categorize the LGUs into two groups: those with internet access and those without. Those with stable internet connectivity will be asked to transfer their data using Microsoft Forms (Figure 11). A notable benefit of utilizing MS Forms is its real-time synchronization with the corresponding MS Excel database, ensuring 24/7 updates.

Figure 11. PUVMP LGU Application Updates MS Form

Take note, however, that only authorized LGU representatives with granted access codes would be allowed to transmit their data to the unified database through the online form. If a submission failure occurs, an automated email notification will be generated and sent to the sender (Figure 12). An additional feature may also be incorporated to only consider responses entered the system using a specific email domain (Figure 13), e.g., *.dotr@gmail.com*. All successful responses will then automatically enter the database (Figure 14), which will be housed in a Microsoft OneDrive folder with limited access permissions. To maintain the consistency of all incoming data, we also recommend including an anomaly detection feature that would prohibit responses from entering the system if it does not conform to the data quality rules embedded in the system.
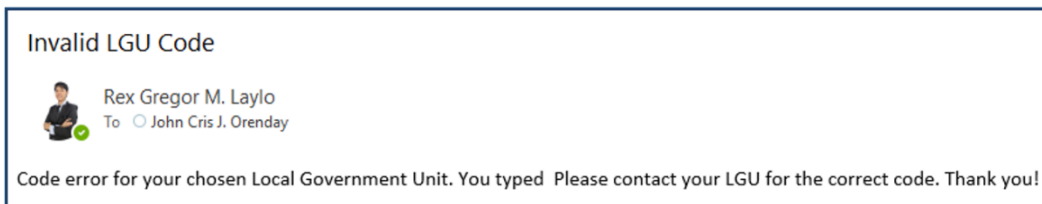
Figure 12. Automated System Response
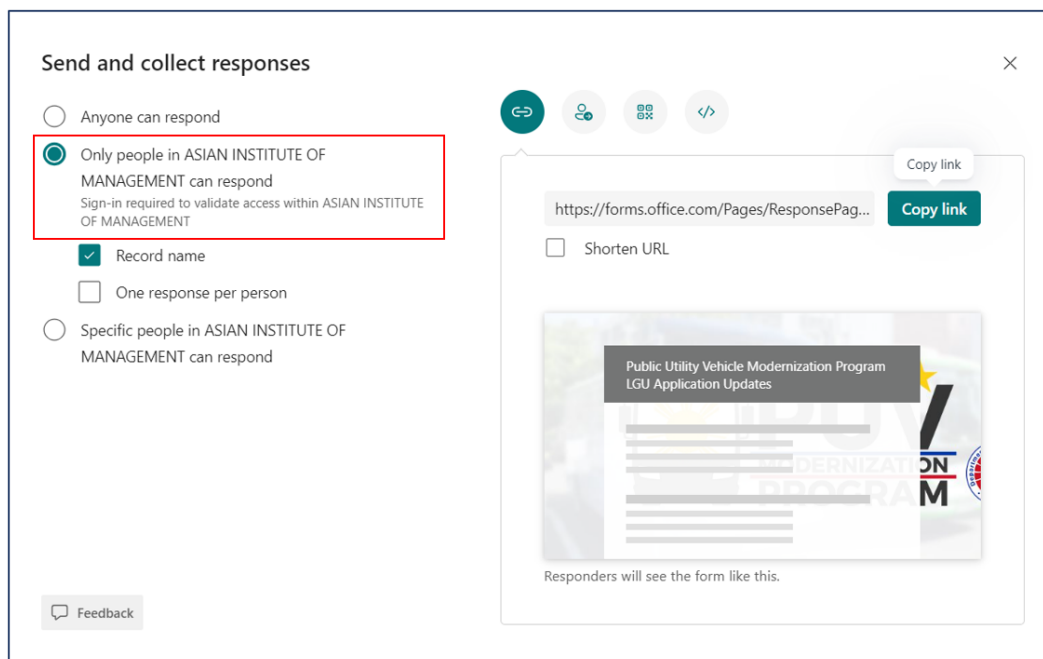
Figure 13. Response Collection Restriction

Figure 14. MS Forms Database

On the other hand, for LGUs facing internet connectivity issues, they may opt to use the *ClearScanner* software application to scan their hardcopy files and convert them into text files before uploading to their respective OneDrive folder that is separate from the database. The same level of security measures will be applied – only those given access to the LGU folder will be able to upload files and it is also possible to send an email notification if the formatting of the uploaded files does not adhere to the PMO's data quality standards as established by data stewards.

5.2 Automated Data Analysis

More than the automated data collection system, the VB-powered repository also has the capability to automatically generate dashboards and reports at predefined intervals. It can operate around the clock, providing the opportunity for real-time data processing (Figure 15).
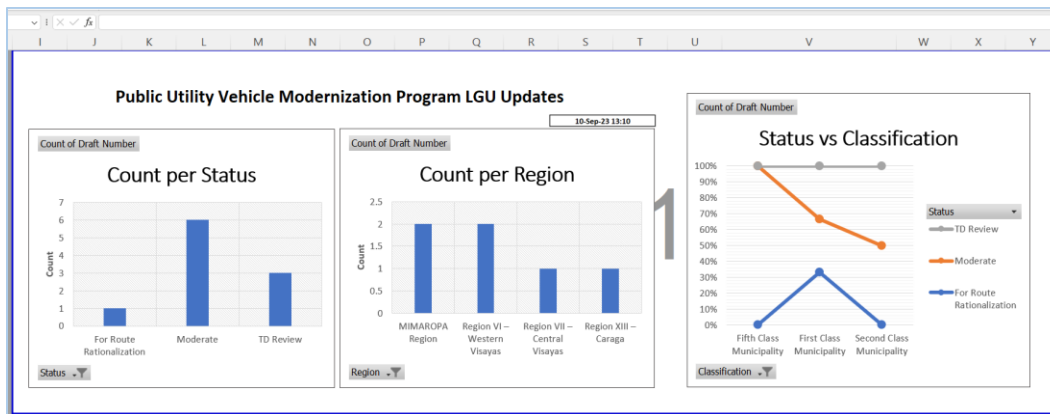
Figure 15: Automated Report Generation

Table 3 summarizes the features that the consolidated database offers and the corresponding prerequisites of each. To prevent storage capacity issues in the future, we recommend instituting a policy that dictates the periodic removal of data that had been integrated already into the database, every month for instance.

Features	Description	Prerequisites
Extract Data from Hardcopies	Extract data from LGU Forms in hardcopy, convert it to text format, and transfer it to DOTr database via Processor.	• Smartphone with minimum 1080p camera which has resolution of 1920 x 1080 (2.07 MP) • Wifi / Mobile Data • Visual Basics for Application
Paperless LGU Application Forms	Facilitates paperless LGU transactions for PUVMP via MS Forms and store data in MS Forms Database.	• Wifi / Mobile Data • Authorized e-mail account • Valid code
Database Integrator (Processor)	An Excel file, powered by VBA Macro, designed to consolidate data from LTFRB, DOTr, and MS Forms Databases to single database. This also maintains data quality through the implementation of migration restrictions.	• Laptop (at least 8GB RAM, i5 Processor) • Visual Basics for Application • OneDrive • MS Excel • Microsoft Outlook
Automated System Response	Processor automatically sends feedback to the data source via Outlook when inputs fail to meet formatting and data type criteria.	
Automated Data Analysis and Report Generation	The processor auto-generates data visualizations via Pivot Tables and Charts, exports them to PDF at specified cadence, and stores it to Onedrive.	

Table 2. The Features of and Prerequisites of the Consolidated Database

6 CONCLUSION

In the context of the DOTr's data strategy, while the adoption of the aforementioned tools could significantly improve the operational efficiency within the PMO, it is the proper management and deployment of these tools that could better dictate the trajectory of its contributions to the decision-making process of the Department. In essence, the success of our proposed data strategy hinges on the development of these three core competencies:

- **Data Accountability:** This competency serves to establish clear lines of responsibility within the PMO for its data-related activities, starting with the institutionalization of robust standards for effective data management. It further ensures that data-related processes and decisions are entrusted to individuals best suited for the task.
- **Data Fluency:** This competency is meant to enhance the Department's understanding of its own data and the value that it contributes to its projects and operations. Prioritizing data literacy enables the PMO to maximize the advantages of using data to inform and shape its policies.
- **Data curation:** Establishing a single source of truth helps the PMO to exercise better control over its data assets. This not only streamlines data management processes and enhances decision-making, but also helps facilitate comprehensive audits for data integrity and reliability across all entities involved in the implementation of the PUV Modernization Program.
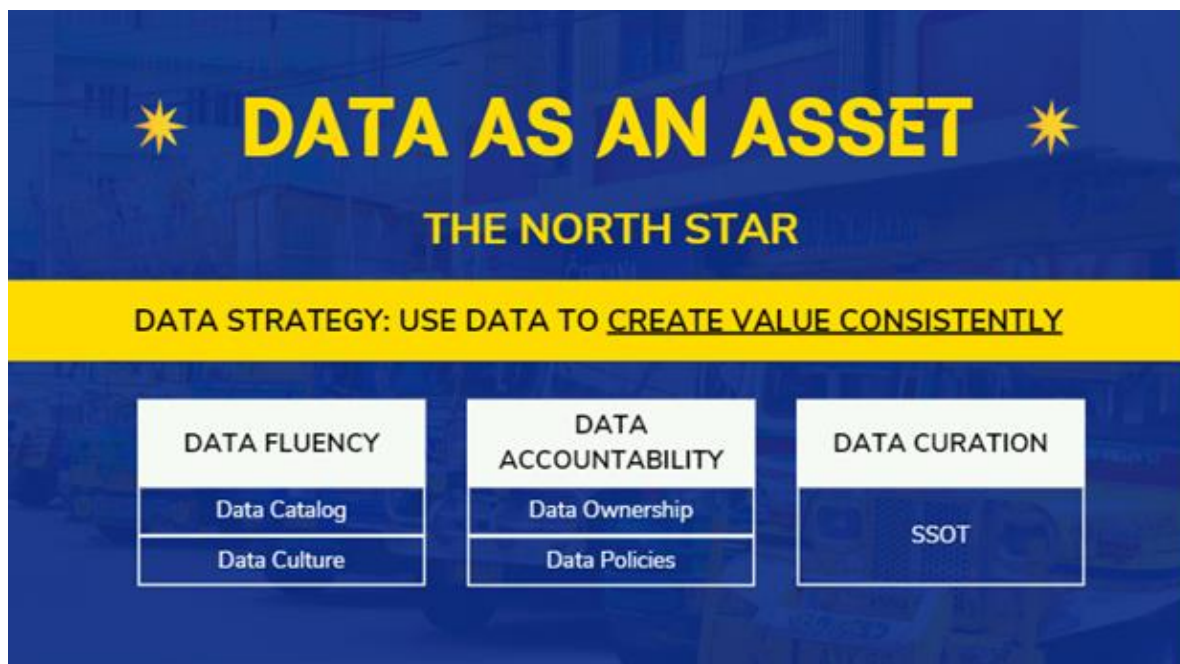
7 REFERENCES

Almasi, B. (Mar. 15, 2021). How to Measure Data Quality. Towards Data Science. https://towardsdatascience.com/how-to-measure-data-quality-815076010b37

Arcalea. (n.d.). Data Science Maturity Models (DSMMs): A Broad Consensus Composite. https://www.arcalea.com/blog/data-science-maturity-models.

Department of Budget and Management. National Budget Circular. "Rationalization of Project Management Offices (PMOs)." https://www.dbm.gov.ph/wp-content/uploads/2012/03/National-Budget-Circular-Rationalization-of-PMOs.pdf

Informatica. (May 1, 2019). Managing Data as an Asset: An Interview with the CEO of Informatica. McKinsey & Company. https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/managing-data-as-an-asset-an-interview-with-the-ceo-of-informatica

Kate C. (2020, October 28). An MSP's Guide to Google G Suite Security. MSP 360. https://www.msp360.com/resources/blog/google-g-suite-securityguide/#:~:text=Secure%20G%20Suite%20User%20Accounts,from%20being%20stolen%20or%20deleted.

KPMG. (2019, October). Data as an Asset. KPMG.

Massachusetts Institute of Technology. (2017). Reshaping Business with Artificial Intelligence. https://sloanreview.mit.edu/projects/reshaping-business-with-artificial-intelligence/

McKinsey & Company. (Sept. 2018). Why Data Culture Matters. McKinsey & Company. https://www.mckinsey.com/~/media/McKinsey/Business Functions/McKinsey Analytics/Our Insights/Why data culture matters/Why-data-culture-matters.ashx

McKinsey & Company. (June 26, 2020). Designing Data Governance That Delivers Value. McKinsey & Company. https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/designing-data-governance-that-delivers-value

McKinsey & Company. (Jan. 5, 2023). Real-World Data Quality: What Are the Opportunities and Challenges. McKinsey & Company. https://www.mckinsey.com/industries/life-sciences/our-insights/real-world-data-quality-what-are-the-opportunities-and-challenges

Oliver Wyman Forum, & University of California, Berkeley. (n.d.). Urban Mobility Readiness Index. https://www.oliverwymanforum.com/content/dam/oliver-wyman/ow-forum/template-scripts/urban-mobility-index/PDF/Mobility-Index-Executive-Summary.pdf

8 APPENDICES

8.1 Data Strategy Survey Response

This is the result of the survey answered by the PUVMP PMO Team that highlights their priorities in terms of Data Strategy.

DATA DEFENSE	Score
1. Reduce general operating expenses	ü
2. Meet industry regulatory requirements	ü
3. Prevent cyberattacks and data breaches	
4. Mitigate operational risks such as poor access controls and data losses	ü
5. Improve IT infrastructure and reduce data-related costs	
6. Streamline back-office systems and processes	
7. Improve data quality (completeness, accuracy, timeliness)	ü
8. Rationalize multiple sources of data and information (consolidate and eliminate redundancy)	ü
Total	**5**
DATA OFFENSE	
1. Improve revenue by adjusting pricing strategies and executing initiatives within the Department of Transportation	
2. Create new products and/or services	
3. Respond rapidly to shifts, changes, and challenges in the transportation landscape	ü
4. Use sophisticated customer analytics to drive operational outcomes and strategic decisions within the Department of Transportation	
5. Leverage new sources of internal and external data	ü
6. Monetize organizational data assets (potentially offer data-driven products and/or services)	
7. Optimize existing strong bench of analysts and data scientists	ü
8. Generate return on investments in big data and analytics infrastructure	
Total	**3**
Overall data strategy: Data Defense	

What is your envisioned data strategy for your team and/or the Department?
The Department envisions to establish clear data governance policies and procedures and ensure that the data collected is accurate, consistent and reliable. The Department intends to develop a process to clean, validate, and maintain data quality over time. Furthermore, we would like to explore different ways on how to share the data to the public.

What challenges and/or limitations are impeding your ability to realize this vision?
The team currently experiences lack of resources in data management leading to a limited access to different data management tools and storage solutions. The team also lack experience and training in managing the data collected.

8.2 Photo Op with PUVMP PMO Team

Our learning team had the opportunity to conduct a face-to-face interview with the members of PUVMP PMO team last August 23, 2023.