

IDAHO TRANSPORTATION DEPARTMENT

RESEARCH REPORT

Guidance on Evaluating Quality and Determining Return on Investment for Internal and Third-Party Data Research Project

RP 307

Prepared By:

Rebecca Van Dyke, Donner Kahl, Mark Egge, & Isaac Gonzalez

High Street Consulting Group

Prepared for

Idaho Transportation Department

[ITD Research Program, Planning and Development Services](#)

Highways Division

December 2024



YOUR Safety •••▶ **YOUR Mobility** •••▶ **YOUR Economic Opportunity**

Disclaimer

This document is disseminated under the sponsorship of the Idaho Transportation Department and the United States Department of Transportation in the interest of information exchange. The State of Idaho and the United States Government assume no liability of its contents or use thereof.

The contents of this report reflect the view of the authors, who are responsible for the facts and accuracy of the data presented herein. The contents do not necessarily reflect the official policies of the Idaho Transportation Department or the United States Department of Transportation.

The State of Idaho and the United States Government do not endorse products or manufacturers. Trademarks or manufacturers' names appear herein only because they are considered essential to the object of this document.

This report does not constitute a standard, specification or regulation.

Technical Report Documentation Page

1. Report No. FHWA-ID-24-307	2. Government Accession No.	3. Recipient's Catalog No.	
4. Title and Subtitle Guidance on Evaluating Quality and Determining Return on Investment for Internal and Third-Party Data Research Project		5. Report Date December 2024	
		6. Performing Organization Code	
7. Author(s) Rebecca Van Dyke Donner Kahl Mark Egge Isaac Gonzalez		8. Performing Organization Report No.	
9. Performing Organization Name and Address High Street Consulting Group 6937 Blenheim Ct. Pittsburgh, PA 15208		10. Work Unit No. (TRAIS)	
		11. Contract or Grant No.	
12. Sponsoring Agency Name and Address Idaho Transportation Department (SPR) Highways Division, Planning and Development Services, Research Program PO Box 7129 Boise, ID 83707-7129		13. Type of Report and Period Covered Final Report 01/02/2024 – 12/31/2024	
		14. Sponsoring Agency Code RP-307	
15. Supplementary Notes Project performed in cooperation with the Idaho Transportation Department and Federal Highway Administration.			
16. Abstract ITD recognizes the challenge of selecting the best approach to making decisions around which data products are worth investing in and which ones may appear promising on the surface, but actually add little value to the agency's business processes. In light of these challenges, ITD launched a research project to explore methods for evaluating the return on investment (ROI) of different data investment opportunities to aid in their decision-making processes. The goal of this research was to create both processes and tools that will enable ITD to make more standardized, objective decisions around data investments. This report covers the research findings that informed the development of two new data asset evaluation tools that ITD can use to vet data investment opportunities, avoid the potential data purchasing pitfalls, and present comprehensive documentation that thoroughly substantiates a decision to move forward with (or rule out) what can be significant resource outlays.			
17. Key Words Return on investment, benefit cost analysis, data assets, data quality, data investment		18. Distribution Statement Copies available from the ITD Research Program	
19. Security Classification (of this report) Unclassified	20. Security Classification (of this page) Unclassified	21. No. of Pages 123	22. Price None

Technical Advisory Committee

Each research project is overseen by a Technical Advisory Committee (TAC), which is led by an ITD project sponsor and project manager. The TAC is responsible for monitoring project progress, reviewing deliverables, ensuring that study objectives are met, and facilitating implementation of research recommendations, as appropriate. ITD's Research Program Manager appreciates the work of the following TAC members in guiding this research study.

- Project Sponsor: Margaret Pridmore
- Project Manager: Margaret Pridmore
- TAC Members: Ned Parrish, Amanda Laib, Josh Roll, Lance Johnson, Carin Pluto, Lou Anne Daugherty, Will Thoman, David Coladner
- FHWA-Idaho Advisor: Lance Johnson

Table of Contents

Executive Summary.....	12
1. Introduction.....	15
Research Overview	15
2. Literature Review	17
Task Purpose.....	17
Resource Details By Type	17
Idaho Transportation Department resources	17
State and local government documents	20
FHWA and USDOT	24
NCHRP	28
Academia and industry	29
Other resources	31
Synthesis of Data Evaluation Criteria and Methods.....	33
Value Creation.....	34
Data and Provider Quality.....	34
Data Integration	35
Cost	35
Decision-Making Methods	36
3. Surveys, Meetings, and Interviews.....	37
Task Purpose.....	37
Internal Survey	37
Key takeaways.....	37
Respondent information	38
Current data asset procurement processes.....	38
Making data investment decisions	41
Third-party data experiences	44
Final comments.....	46
External Survey.....	46
Key Takeaways	46

Respondent information	46
Current data asset procurement processes.....	47
Making data investment decisions	50
Third-party data experiences	53
Final comments.....	54
ITD Townhall and Technical Advisory Committee Meetings.....	55
Peer State Interviews	55
New Mexico Department of Transportation	55
Minnesota Department of Transportation	56
4. Data Asset Evaluation Methodologies	58
Task Purpose.....	58
Comprehensive Methodology List	58
Recommended Approaches	61
Data Purchase/Pilot Abstract Proposal – Short Form Tool	61
When to use the DPAP	61
Components of the DPAP.....	62
Data Investment Needs and Evaluation – Long Form Tool	64
Instructions and summary	64
Descriptive information and management.....	65
Business use cases and benefit cost analysis.....	66
Quality, integration, and risk evaluations	68
5. Test Case Evaluations	71
Task Purpose.....	71
Vetting Tool Evaluation Summary	71
Test Case Details.....	73
Michelin Driving Data to Intelligence data.....	73
INRIX Travel Time data.....	75
Continuous Count data	78
6. Conclusions and Next Steps	82
7. Cited Works	83
Appendix A. Surveys.....	85

Appendix B. Data Proposal/Pilot Abstract Proposal (DPAP)	101
Appendix C. Data Investment and Needs Evaluation Long Form	103

List of Tables

Table 2.1 Literature review sources summary.....	17
Table 2.2 Literature review synthesis studies.....	34
Table 5.1 ITD DINE feedback and project team response	71
Table 5.2 Additional DINE formatting and formula checks.....	72

List of Figures

Figure 2.1 ITD Contracting Procurement Decision Tree	18
Figure 2.2 Screenshot of ITD GIS Data Prioritization Table (Idaho Transportation Department 2022)	19
Figure 2.3 Screenshot of COG's big data evaluation table (Metropolitan Washington Council of Governments 2021)	21
Figure 2.4 IDOT data source summary for pavement data (Iowa Department of Transportation 2020) ..	23
Figure 2.5 Caltrans data quality objectives table (California Department of Transportation 2019)	24
Figure 2.6 Data.gov publishing guidelines dimensions (Data.gov Working Group 2010)	27
Figure 2.7 Data.gov data scoring criteria (Data.gov Working Group 2010).....	28
Figure 2.8 Gartner Magic Quadrant (Gartner n.d.).....	32
Figure 3.1 Survey of unmet data needs among respondents.....	39
Figure 3.2 Respondents' approaches to addressing data needs	39
Figure 3.3 How respondents evaluate a dataset's value	40
Figure 3.4 Respondents' most important considerations when making data-related decisions	42
Figure 3.5 Respondents' most important considerations when evaluating data quality	43
Figure 3.6 Respondents' most important data integration considerations.....	44
Figure 3.7 How respondents identify use cases for commercial data	45
Figure 3.8 Survey of unmet data needs among external respondents.....	48
Figure 3.9 External respondents' approaches to addressing data needs	49
Figure 3.10 How external respondents evaluate a dataset's value	49
Figure 3.11 External respondents' most important considerations when making data-related decisions	51
Figure 3.12 External respondents' most important considerations when evaluating data quality	52
Figure 3.13 External respondents' most important data integration considerations	53
Figure 5.1: Description and ownership details from Summary Page (MichelinDDi)	74
Figure 5.2 Use cases and BCA ratio details from Summary Page (MichelinDDi)	74
Figure 5.3 Data quality, integration, and risk scores (MichelinDDi)	75
Figure 5.4: Description and ownership details from Summary Page (INRIX)	76
Figure 5.5: Use cases and BCA ratio details from Summary Page (INRIX)	77
Figure 5.6: Data quality, integration, and risk scores (INRIX)	77
Figure 5.7: Description and ownership details from Summary Page (Continuous Counters)	79
Figure 5.8: Use cases and BCA ratio details from Summary Page (Continuous Counters)	80
Figure 5.9 Data quality, integration, and risk scores (Continuous Counters)	80

List of Abbreviations and Acronyms

AASHTO	American Association of State Highway and Transportation Officials
AICPA	American Institute of Certified Public Accountants
ADA	Americans with Disabilities Act
APTA	American Public Transportation Association
BCA	Benefit Cost Analysis
COG	Metropolitan Washington Council of Governments
DDi	Michelin Driving Data to Intelligence
DINE	Data Investment Needs and Evaluation
DoEd	United States Department of Education
DOT	Department of Transportation
DPAP	Data Purchase/Pilot Abstract Proposal
EBT	Executive Briefing Tool
ETS	Enterprise Technology Services
FAST	Fixing America's Surface Transportation Act
FHWA	Federal Highway Administration
GIS	Geographic Information System
HPMS	Highway Performance Monitoring System
HSIP	Highway Safety Improvement
IDOT	Iowa Department of Transportation
IET	Institution of Engineering and Technology
ITD	Idaho Transportation Department
IT	Information Technology
LiDAR	Light Detecting and Ranging
MAP-21	Moving Ahead for Progress in the 21st Century Act
MIRE	Model Inventory Roadway Element
MnDOT	Minnesota Department of Transportation
NCHRP	National Cooperative Highway Research Program

NMDOT New Mexico Department of Transportation
NPV Net Present Value
OHS Office of Highway Safety
ROI Return on Investment
SHSP Strategic Highway Safety Plan
SWAAI Statewide Asset Attribute Inventory
TAC Technical Advisory Committee
TBP Transportation Planning Board
US United States
USDOT United States Department of Transportation
WRAP Work Request Abstract Proposal
YYYY Formatting shorthand for “Year” value

Executive Summary

In recent years, transportation agencies and planners have experienced an influx of new data sources and services. While agencies and planners have utilized many of these products to meet regulatory requirements, justifying a dataset's worth when trying to meet non-regulatory goals, such as strategic initiatives, remains difficult. The Idaho Transportation Department (ITD), in recognizing the potential in these data sources and services, partnered with High Street Consulting Group (collectively, the "project team") to explore the ways in which transportation agencies can determine the return on investment (ROI) for a procurement as nebulous as data. The goal of this research was to create both processes and tools that will enable ITD to make more standardized, objective decisions around data investments. The final report that follows covers the research findings that informed the development of two new data asset evaluation tools that ITD can use to vet data investment opportunities, avoid the potential data purchasing pitfalls, and present comprehensive documentation that thoroughly substantiates a decision to move forward with (or rule out) what can be significant resource outlays.

This research project was comprised of four main tasks:

Task 1: Literature Review

The project team assessed existing literature on strategies and criteria for data investment decisions. Team members reviewed 16 resources, which included state department of transportation (DOT) data management plans, National Cooperative Highway Research Program (NCHRP) research, United States Department of Transportation (USDOT) and Federal Highway Administration (FHWA) reports, academic publications, and other resources.

After an extensive review of resources, the project team organized common data asset evaluation criteria into the following four (4) classifications:

- **Value Creation** – Value creation includes criteria that evaluate the dataset's abilities to satisfy the agency's need, save time or money, or provide benefit for multiple departments and staff. It could require an internal needs assessment and a clear record describing the features of the data options.
- **Quality** – Quality classification covers the various metrics for evaluating the data's reliability, accuracy, precision, etc. In addition, it can refer to the quality of the provider, such as training and assistance provided, transparency in methodology, and user reviews.
- **Integration** – integration classification covers usability. It seeks to answer the question of how easily ITD will be able to incorporate and analyze the different data options.
- **Cost** – Cost acts as the tradeoff factor and it bounds any decision and provides an objective metric for comparison across data investment decisions.

Task 2: Surveys, Meetings, and Interviews

After conducting the literature review, the project team synthesized and presented the findings to ITD's internal and external partners for further discussion. Stakeholder input was provided through internal

(ITD-staff only) and external (other state DOT) survey, townhall-style discussion with ITD data experts, several meetings with the eight-member Technical Advisory Committee (TAC), and two interviews with peer state DOTs (New Mexico and Minnesota). This research revealed additional priorities and helped the project team select the evaluation methods that best suited ITD.

The 41 Internal Survey responses revealed common themes and practices, which are outlined below:

- Most respondents reported data gaps, such as insufficient variable coverage.
- Most respondents believe that existing data and software are underutilized by ITD staff.
- Quality, specifically accuracy, was reported to be the most important consideration, but it is difficult to evaluate.
- Responses were split for primary decision-making methods between evaluating business use cases and leveraging professional judgement.
- Most respondents agreed that department wide data prioritization and governance would address many challenges.

The following themes and practices emerged in the 30 external survey responses:

- Agencies focused on business use cases when evaluating data-needs.
- Agencies noted challenges in capturing the full value of data.
- Data accuracy is the highest priority for agencies, but it is difficult to assess.
- Agencies highlighted the importance of governance, data warehousing, and ownership.

The “town hall” with ITD staff that have data-related responsibilities and two meetings with the research project Technical Advisory Committee (TAC), which was made up of professionals both internal to and outside of ITD. The goals of these meetings were to discover ways in which ITD and other TAC members’ agencies handle data decisions, discuss successes and challenges, and gather ideas for the most important considerations for the Tool. The takeaways spanned three themes: ROI Calculation, Communication with Leadership, and Data Investment Considerations.

Two state DOTs were selected for follow-up interviews to discuss the specifics of their responses: New Mexico Department of Transportation and Minnesota Department of Transportation. New Mexico Department of Transportation (NMDOT) was selected because they indicated using a BCA process when making data investment decisions. The project team had a follow-up discussion with Minnesota DOT (MnDOT) because they reported using data pilots for evaluating data investment decisions. These interviews provided the project team with practical insights that informed the final tool design and components.

Task 3: Data Asset Evaluation Methodologies

Next, the project team developed a comprehensive of methodologies and subsequently selected a dual form approach. The forms include a short form for smaller (i.e., less costly) investments or data pilots and a longer evaluation tool for larger (i.e., more costly) investments. The two tools are as follows:

- **Data Purchase/Pilot Abstract Proposal (DPAP)** – A document-based short form to evaluate data pilots and small, one-off data purchase requests. DPAP provides decision-makers with a preliminary summary of business use cases, strategic alignments, cost, and management considerations. Then, a decision can be made whether to make the investment or pursue the pilot, request a longer evaluation, or defer investment.
- **Data Investment Needs and Evaluation (DINE)** – A spreadsheet-based long form tool to help inform larger, longer-term investments. DINE is a comprehensive evaluation form that requires collaboration across teams and departments to complete.

Task 4: Test Case Evaluation

For the final task in the research project, ITD staff completed the DINE Tool to evaluate three ITD data investment opportunities. The project team used the feedback from the test case exercise to gauge the usability of the DINE Tool and make any final adjustments. ITD completed the DINE for the following three datasets:

- Michelin Driving Data to Intelligence (DDi)
- INRIX Travel Time Data
- Continuous Count Data (produced in house)

The feedback led to DINE Tool adjustments. Overall, ITD was satisfied with the tool functionality and felt that the DINE achieved its intended purpose.

Conclusions

This project sets the path for ITD to develop a more comprehensive data management plan. As ITD continues to employ the DPAP and DINE Tool, ITD may benefit from assessing its approach to data management and move beyond standardizing new data acquisitions. By adapting the method for valuing data to current datasets, ITD can better understand data as an asset. Moreover, the research and cross-department collaboration encouraged by the DINE Tool can further help ITD staff improve their own understanding, along with that of leadership, of the data used across the agency, creating opportunities for additional efficiencies and reinforcing the view of data as a valuable asset, just like roads, bridges and other infrastructure that the state maintains.

1. Introduction

A proliferation of data sources, providers, and third-party vendors have created numerous opportunities for the Idaho Transportation Department (ITD) to procure data that fits their regulatory and strategic business requirements. New datasets consistently enter the world of transportation planning and, while they may appear promising at first, they can also lead an agency astray (this is often referred to as the “shiny object” phenomenon). Facing this increase of opportunities, some data owners and users may struggle to articulate the value of a data investment while others may be lured in by an exciting new product and encourage data purchases without thoroughly exploring their real value. These “data investments” could either be datasets purchased from a third-party vendor or internal resources that have been dedicated to collecting and maintaining data in-house. With the increasing availability of data and the third-party vendors that produce, package, and sell data solutions, how and when ITD invests in datasets are critical considerations.

ITD recognizes the challenge of selecting the best approach to making decisions around which data products are worth investing in and which ones may appear promising on the surface, but actually add little value to the agency’s business processes. In light of these challenges, ITD launched a research project to explore methods for evaluating the return on investment (ROI) of different data investment opportunities to aid in their decision-making processes. The goal of this research was to create both processes and tools that will enable ITD to make more standardized, objective decisions around data investments. This report covers the research findings that informed the development of two new data asset evaluation tools that ITD can use to vet data investment opportunities, avoid the potential data purchasing pitfalls, and present comprehensive documentation that thoroughly substantiates a decision to move forward with (or rule out) what can be significant resource outlays.

Research Overview

This research project was comprised of four main tasks, each of which is briefly described below.

Task 1: Literature Review

The literature review involved a thorough examination of national, state and local government, ITD, academic, and industry resources. The material covered data quality evaluations criteria and approaches, return on investment calculation methods, and other data investment considerations. The topics and methods were ultimately distilled into a summary catalogue, which was provided to ITD as a supplemental resource.

Task 2: Surveys, Meetings, and Interviews

The project team (ITD and High Street Consulting Group) researched existing ITD and peer state data investment decision making methods using an internal (within ITD) and external (other state DOT) survey, peer state DOT interviews, and by consulting with an eight-member Technical Advisory

Committee (TAC). The various stakeholder engagement methods helped identify additional priorities and home in on the methods that would best fit ITD's existing processes.

Task 3: Recommended Methodologies

From the extensive background research of the previous two tasks, the project team developed a complete list of the methods that ITD could incorporate into their data asset evaluation process and tools. Using this list, the project team selected the methods that best fit ITD needs and developed two separate data asset vetting tools. The first, called the Data Purchase/Pilot Abstract Proposal (DPAP), would help evaluate small investments and data pilots, while a second longer and more comprehensive approach was designed for larger investments. This second tool, called the Data Investment Needs and Evaluation (DINE) Tool, is a spreadsheet-based form and is the primary focus of this report.

Task 4: Test Case Evaluations

The final main step was to test the DINE using real world examples. The project team selected three datasets and an ITD testing team that would use the DINE to evaluate each one. The results of the test cases provided the project team with valuable feedback that was used to make final adjustments to each of the two tools.

2. Literature Review

Task Purpose

The purpose of the literature review is to cover a broad cross section of sources to identify existing strategies and considerations for data investment decisions. The project team reviewed state Department of Transportation (DOT) data management plans, National Cooperative Highway Research Program (NCHRP) research, United States Department of Transportation (USDOT) and Federal Highway Administration (FHWA) reports, academic publications, and other resources. **Table 2.1** provides a summary of the resources identified and reviewed by type. The project team made a concerted effort to review data investment decision-making methodologies employed by other State DOTs. Each of the sources' methodologies and criteria are summarized in subsections that follow.

Table 2.1 Literature review sources summary

Resource Type	Count
ITD Resources	3
State and Local Government Documents	3
USDOT and FHWA Materials	3
NCHRP Reports	1
Academic and Industry Research	4
Other Resources	2
Total	16

Resource Details By Type

Idaho Transportation Department resources

The project team reviewed three sources from within ITD itself. Each is briefly summarized below.

Contracting Procurement Decision Tree

ITD produced the Contracting Procurement Decision Tree to help agency staff determine whether projects fall under Purchasing, Highways or Facilities. While the purpose of the document doesn't apply directly to data investments, the use of a decision tree could assist with data investment decisions or determine the detail of the analysis needed to inform the decision. For instance, smaller investment decisions may have simpler criteria compared to the effort required for larger decisions. **Figure 2.1** shows the Contracting Procurement Decision Tree.

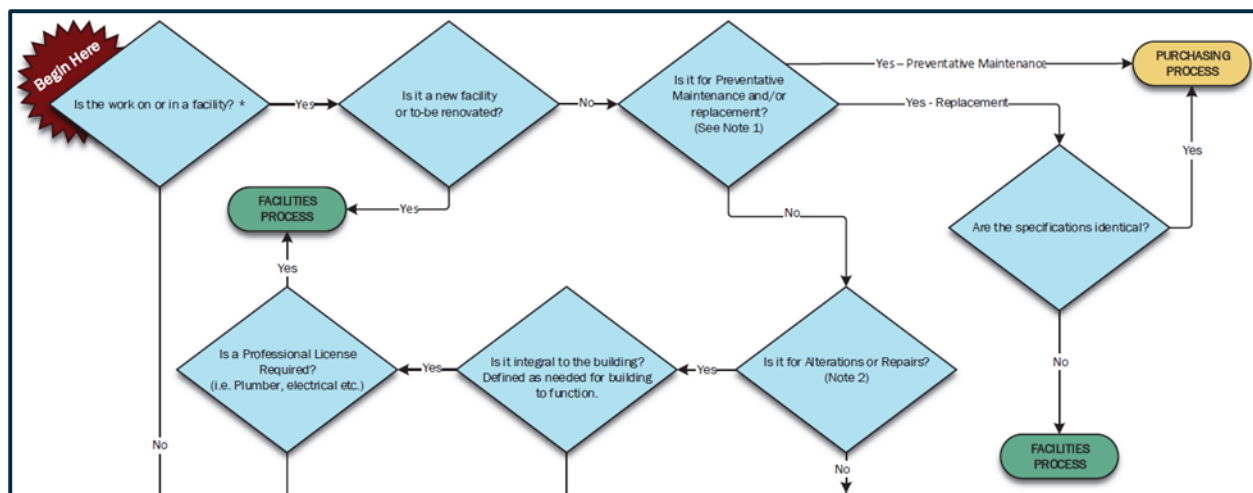


Figure 2.1 ITD Contracting Procurement Decision Tree

Statewide Asset Attribute Inventory

In May 2020, ITD selected Cyclomedia Technology to develop a statewide asset by driving a vehicle equipped with Light Detecting and Ranging (LiDAR) equipment across the state. The resulting document, the *Statewide Asset Attribute Inventory* (SWAAI), reports on the success of the project, citing 28 geographic information system (GIS) asset inventory types that include over 300,000 point and line features and an additional 6,000 polygon features (Idaho Transportation Department 2022). Examples of asset inventory types included signs, billboards, lights, lanes, sidewalks, utilities, and bridges. The report also covers different projects that the Cyclomedia data enabled and improved, such as the ITD *Americans with Disabilities Act (ADA) Transition Plan* and the *GIS Pavement Preservation Planning Tool* with J-U-B Engineers, Inc. This successful project points to criteria that ITD seems to value:

- **Coverage** – The broad coverage of the data both in terms of data types (e.g., signs, lanes, pavement condition, etc.) and geography (the vehicle covered 7,200 miles).
- **Usability** – The project cites the geodatabase as a useful output, referencing the importance of a data output and software platform that agency staff are familiar with. Numerous departments were able to employ the data, demonstrated by the various project examples. The report also provides statistics on the number of users.
- **Value** – The projects examples, ranging from ADA improvements to speed limit analysis, also provide concrete examples of the value that the Cyclomedia data investment provided. The investment saved time and money and contributed to a number of projects. Benefits could be converted to dollar values creating a direct comparison to cost. For instance, ITD could either employ a “time-consuming boots-on-the-ground inventory” or assume that every sidewalk needed attention, both of which would be expensive endeavors. They were replaced with a categorization of sidewalks.

When deciding on which assets (e.g., signs, barriers, pavement, bridges, and curbs and ramps) to prioritize for data collection, ITD employed a three-rank scale – High, Medium, Low – which were converted to point values. The asset data types with the most points were prioritized. ITD also identified a few data types that were known priorities, or “Top 5s.” **Figure 2.2** from the SWAAI report shows various data types and their respective scores.

Data Type	Score	D1	D2	D3	D4	D5	D6	HQ Safety	HQ
Signs & Supports (inc. Milepost/Equation Markers)	Top 5								
Barriers (Guardrail)	Top 5								
Pavement Striping & Messages	Top 5								
Rumble Strips	Top 5								
Vertical Clearance	Top 5								
Approaches	32 (High)	H	H	H	H	H	H	H	H
Curb (ADA) Ramps	21 (High)	H	M	M	M	H	M		M
Luminaires & Lighting	18 (High)	M	L		H		H	H	M

Figure 2.2 Screenshot of ITD GIS Data Prioritization Table (Idaho Transportation Department 2022)

ETS Technology Request Intake Process and Intake Council overview

ITD and its Information Technology (IT) service group, Enterprise Technology Services (ETS), developed a presentation articulating the process for filling out and reviewing technology requests. The goals of the intake process may be applicable to data investment decisions (Idaho Transportation Department n.d.):

- **Avoid Duplication** – New data investments should avoid duplicating existing data.
- **Ensure Compatibility** – New data should integrate into existing data products and infrastructure.
- **Identify Relationships** – When possible, data investments should satisfy multiple data needs, increasing the value.

Based on the presentation, the Technology Request Intake Process requires the requestor to fill out a Work Request Abstract Proposal, or WRAP, which outlines the technology need. Then, an Intake Council reviews the WRAP and considers the request. The WRAP sections may be useful as ITD considers criteria for data investment decisions:

- **General Details** – Defining the need will help identify the value of data investment options.
- **Fiscal Requirements** – The cost of the technology investment is an important consideration across options.
- **Strategic Alignment** – Strategic alignment implies how well the data need fits in with the larger agency goals, indicating importance.

- **Technology Requirements** – The IT time and training needed to utilize data can contribute to the overall cost of a data investment option.

ITD employs an Intake Council to make the final decisions on whether to make the investment. When evaluating or weighing criteria, a degree of judgement is needed. The Intake Council provides a good example of a comprehensive decision-making board that represents various priorities and perspectives, which helps ensure the decision-making accounts for relevant considerations from across the agency (Idaho Transportation Department n.d.).

State and local government documents

The project team reviewed three state and local government documents. Each is briefly summarized in the sections that follow.

Metropolitan Washington Council of Governments Big Data for Regional Travel and Mobility Analyses

A team of consultants developed the *Big Data for Regional Travel and Mobility Analyses* report for the National Capital Region Transportation Planning Board (TBP)—Metropolitan Washington Council of Governments (COG) is the administrative agent for the TBP—to help them improve big data investment decision-making processes (Metropolitan Washington Council of Governments 2021). The report authors first performed a literature review and interviewed local government and peers to identify big data use cases and needs. Then, they interviewed the data providers, such as INRIX, to understand pricing and their product offerings. Finally, they performed an independent evaluation of the products with technical experts. In addition to providing use cases of big data, the report includes an evaluation of an extensive list of big data providers. The culmination of the research was a discussion of recommended big data vendors.

When reviewing the data providers, the researchers considered the following topics:

- Product overview
- Applicable research areas
- Agencies using the provider
- Cost and pricing structures
- Data licensing and sharing
- Data storage, processing, and analysis

The technical experts analyzed the data products using seven evaluation criteria:

- **Applicability for research area** – Determines the data’s ability to meet TBP needs, answer research questions, and address concerns that arise over time.
- **Data reliability and validity** – Verifies the quality of documentation and validation of the methodology, the maturity of the data source, and the data’s integrity over time.
- **Data coverage** – Evaluates the coverage of the data, whether measured as population, time periods, or data frequency, accounting for generalizability (results applying everywhere) and transferability (results applying to other locations or times).
- **Technical staff competencies and training required** – Considers whether the data is user-friendly, whether the vendor provides software or an interface for analysis, how well the data fits into easy-to-use analytic software (Excel or ESRI), and how much staff training or augmentation is needed.
- **IT requirements** - Determines how well the data fits into the existing IT infrastructure and the amount of effort needed for data management.
- **Data sharing restrictions** – Ease with which data can be shared with stakeholders and partners, including sharing raw data or accessing a web-interface.
- **Data cost** – The relative cost of a subscription or download.

The experts gave each source a binary positive-negative score. (If there was insufficient information, the data provider received an “N/A”.) They also provided comments. The result of the evaluation, or the “Overall Recommendation”, was a Yes or No, indicating whether TBP should consider the big data provider. **Figure 2.3** is an example of the binary evaluation for three of the vendors’ ability to provide travel demand management data. Not shown are the comments on each of the criteria. Note that it isn’t clear how the final recommendation designation was determined based on the seven criteria.

Product	Applicability to Research Area	Data Reliability and Validity	Data Coverage	Resource Requirements (Technical Staff and Training Requirements)	Resource Requirements (Technology, IT)	Data Sharing Restrictions	Cost	Overall Recommendation
InfoUSA/Dun & BradStreet Business Listings	+	+	-	+	+	+	-	Yes
CoStar Data	+	+	-	+	+	+	-	Yes
Google Places	+	+	-	+	+	+	-	Yes

Figure 2.3 Screenshot of COG’s big data evaluation table (Metropolitan Washington Council of Governments 2021)

Iowa DOT Data Management Strategic Plan

The Iowa Department of Transportation (IDOT) put together the *Data Management Strategic Plan* to outline the strategic, operational, and tactical approaches to managing data (Iowa Department of Transportation 2020). The plan does not include a process for making data investment decisions, but it does provide a review of data importance based on various criteria. For instance, the management plan identifies data concerns, which may be valuable considerations for data investment. These include the following:

- **Data utilization** –How, by who, and how often would the data be used?
- **Data duplication** – Do similar data already exist, and what makes this new data better?
- **Inefficient integration** –How well would the new data integrate into the current data system?
- **Unaddressed data needs** – How well do the data meet the actual need?
- **Underappreciated data value** –Are there other analyses that the data can be used for?
- **Inefficient use of resources** – Are data difficult to access, can they only be analyzed by a few people, or does poor communication around data use cause replication of analyses?
- **Inconsistent results** – If two or more datasets with similar data exist, are the values similar or reasonable?
- **Gap in skillset** – Do staff have the technical capability to use the data?

Additionally, an IDOT Data Management Committee performed a qualitative assessment of current data sources with five criteria on a three-point scale either from low to high importance or poor to excellent. While the Management Plan does not explicitly define the criteria, they can be inferred as follows:

- **Importance** – How critical is the data source or the analyses they feed?
- **Quality**, which includes:
 - **Currency** – How recent are the data?
 - **Accuracy** – How reliable are the data?
 - **Completeness** – How many missing or estimated values are in the data?
- **Usability** – How user-friendly are the data?

An example of the assessment done on pavement data is shown in **Figure 2.4**.

Data Source	Importance	Quality			Usability
		Currency	Accuracy	Completeness	
Pavement Data Condition - Roughness, Rutting, Faulting, Cracking	High Importance	Good	Good	Good	Good
Drainage/Roadside Assets (culverts, pipes, shoulders, vegetation)	Medium Importance	Poor	Poor	Poor	Poor
Geospatial Foundation Data - base map, road centerlines, district/region boundaries, jurisdiction boundaries, urban boundaries	Medium Importance	Good	Good	Good	Good
Route designations, lengths, linear referencing system (LRS)	High Importance	Good	Good	Good	Excellent

Figure 2.4 IDOT data source summary for pavement data (Iowa Department of Transportation 2020)

Caltrans Data Quality Management Plan

The California Department of Transportation (Caltrans) produced a *Data Quality Management Plan* template to document the quality of data that is collected, maintained, and exported (California Department of Transportation 2019). The Plan provided nine (9) data quality dimensions to evaluate the quality of the data:

- **Accuracy** –The degree of bias and closeness to truth.
- **Precision** – Exactness, or closeness of multiple observations to each other.
- **Validity** – Conforms to established formats, types, and ranges and matches believed values.
- **Completeness** – Lacks gaps in the data, both in terms of scope coverage and individual records (missing values).
- **Consistency** – Data from different points of time are in alignment based on methodology, format, and meaning.
- **Currency and timeliness** – The frequency and recency of data collection and time required to process and publish the data.
- **Granularity** – The ability of the data to reach the desired level of detail.

- **Uniqueness/non-repudiation** – The degree to which data comes from a single authoritative source and if there are similar data sources; the data are not contradictory.
- **Accessibility** – The extent that the data is in a usable form.
- **Reputation** – The trust level of the data and the source.

Caltrans does not employ a ranking scale for the dimensions, but the *Data Quality Management Plan* user describes the assessment performed for each criterion and comments on the result. **Figure 2.5** depicts an example template that would be used for the data quality assessment.

Data Quality Dimension	Objective	Measurement/ Tracking Method	Comments
<i>Validity</i>	<i>All numeric values should fall within their specified min-max values</i>	<i>Data validation script</i>	
<i>Completeness</i>	<i>The data set should cover no less than 95% of the state highway system's lane mileage, as reported in the most recent annual HPMS report.</i>	<i>Data validation script</i>	

Figure 2.5 Caltrans data quality objectives table (California Department of Transportation 2019)

FHWA and USDOT

The project team reviewed three resources published by FHWA and USDOT. Each is briefly summarized in the sections that follow.

FHWA Data Quality White Paper

The FHWA produced a white paper that provided an analysis on data quality measures and their applications (Federal Highway Administration 2008). The white paper included a review on various quality measures and recommended the following:

- **Accuracy** – The measure or degree of agreement between a data value or set of values and a source assumed to be correct. It is also defined as a qualitative assessment of freedom from error, with a high assessment corresponding to a small error.

- **Completeness** (also referred to as availability) – The degree to which data values are present in the attributes (e.g., volume and speed are attributes of traffic) that require them. Typically described in terms of percentages or number of data values.
- **Validity** – The degree to which data values satisfy acceptance requirements based on defined validation criteria or fall within the respective domain of acceptable values. Data validity can be expressed in numerous ways. One common way is to indicate the percentage of data values that either pass or fail data validity checks.
- **Timeliness** – The degree to which data values or a set of values are provided at the time required or specified. Timeliness can be expressed in absolute or relative terms.
- **Coverage** – The degree to which data values in a sample accurately represent the whole of that which is to be measured. As with other measures, coverage can be expressed in absolute or relative units.
- **Accessibility** (also referred to as usability) – The relative ease with which data can be retrieved and manipulated by data consumers to meet their needs. Accessibility can be expressed in qualitative or quantitative terms.

FHWA Decision-Making Guidebook: Benefit-Cost Analysis of Investing in Data Systems and Processes for Data Driven Safety Programs

FHWA produced a Decision-Making Guidebook for performing a Benefit-Cost Analysis (BCA) for safety data investments (Federal Highway Administration 2012). The analysis steps can be summarized as follows:

1. **Define the proposed investment** – The first step is to define the objectives of the data investment along with the status quo and one or more alternatives. The objective should be clear to assess benefits and should include constraints and assumptions.
2. **Establish framework for the analysis** – The framework involves defining the project life and identifying a discount rate for benefits and costs. The project life allows the investment and benefits to be distributed across the project life cycle.
3. **Quantify the benefits** – The benefits can include direct monetary benefits or non-monetary benefits that should be monetized based on clearly defined assumptions. The Guidebook includes safety specific example categories, which can be generalized:
 - a. **Improved efficiency** – Saving time required to perform the analyses that would be replaced by the new data.
 - b. **Faster project benefits** – The data should decrease the time required to complete a project and the decreased time to project results is a benefit. In a safety example, if a data investment decreases the time and effort required to identify dangerous

intersections, speeding up the department's ability to make adjustments, the benefit is reducing crashes quicker.

- c. **Improved efficacy** – Improved project performance, such as identification, provides a benefit compared to the would-be previous project selection.
4. **Quantify the costs** – The costs can include the following categories:
 - a. **Investments costs** – Upfront cost for the data investment; for example, a one-time fee, pay per download, or a subscription.
 - b. **Operations and maintenance cost** – Cost to incorporate and maintain the data; for example, the hours the IT team spends incorporating the data or the time spent to update the data as it changes.
 - c. **Data storage** – If the data is kept on an agency's server, there is a cost attached; for example, the cost per month for the quantity of data stored.
 5. **Evaluating the benefits and costs** – Once the benefits and costs are identified, the values are calculated in common units in present currency values. A risk, or sensitivity, analysis can be performed to determine how assumption changes impact the final results. Finally, the decision rules can be applied:
 - a. If Net Present Value (NPV) ≥ 0 , a project is economically justified.
 - b. If Benefit Cost Ratio (BCA ratio) ≥ 1 , a project is economically justified.
 - c. NPV and the BCA ratio can also be compared with the NPV or BCA ratios of alternative projects in order to select the most beneficial option. The highest NPV or BCA ratio for a given level of constraint will be the best alternative.

USDOT Data.gov Interim Identification & Prioritization Process and Guidelines

Data.gov is an initiative to improve transparency and an open government by publishing data for public use. USDOT released guidelines for how it prioritizes the data it publishes on Data.gov (Data.gov Working Group 2010). USDOT defines three dimensions—Value, Quality, and Manageability—as categories of criteria. The dimensions and the criteria are weighted to determine whether data is a sufficient priority to publish. The dimensions can be visualized as a notional plot with Value and Quality at either axis and the size of the project dot indicating the Manageability dimension (**Figure 2.6**).

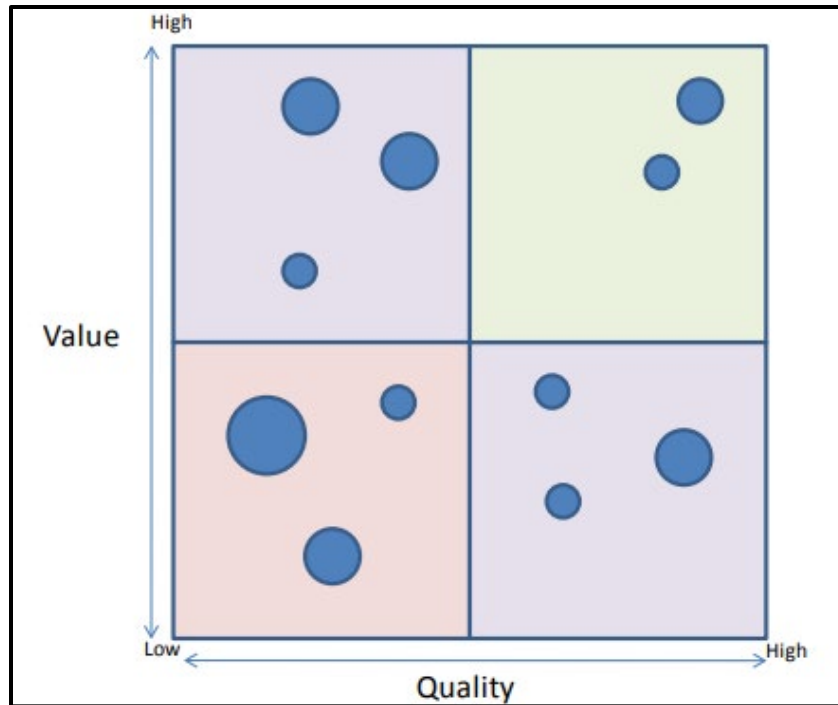


Figure 2.6 Data.gov publishing guidelines dimensions (Data.gov Working Group 2010)

The three (3) dimensions with accompanying weights are:

- **Value (40%)** – Valuable data is determined by applying three criteria:
 - **Relevance** – Relevant data is useful to a variety of users, such as different departments within an organization.
 - **Usability** – Usable data is easily interpretable and in a format that is ready to be employed.
 - **Availability** – Available data is regularly requested, needed, or used.
- **Quality (35%)** – The Guidelines admonishes that quality can be subjective, but defines it with two criteria:
 - **Completeness** – A complete dataset has few missing values and clear methodology for derived values.
 - **Consistency** – Consistent data standardization that adheres to industry or government norms. Consistency means easier data cleaning and joining, especially with other data sources.
- **Manageability (25%)** – Manageability measures the effort required to publish the data, which can also be applied to working with the data. This includes two criteria:

- **Sensitivity** – Sensitivity refers to sensitive information contained in the data, which can incur privacy concerns. It can also be interpreted as the ability to share or publish data, which can depend on data provider licensing agreements.
- **Level of Effort** – While the Guidelines refer to the effort required to prepare the data for publication, this criterion can also account for the level of effort required to manipulate and analyze the data.

Each dimension has a data scoring table, where each criterion is evaluated based on a point scale of either 0 to 2.5 or 0 to 3.5. The criteria and dimensions are then aggregated to calculate a single value that is compared to a threshold that determines whether to publish. **Figure 2.7** provides a snapshot of the table with a description of the dimensions and questions used to evaluate the data.

	Dimensions			Yes	No
40 = Highest Value	Relevance & Accountability	1	Does the data support the Secretary of DOT's current strategic priorities? <i>(That is, improve safety, protect the environment, support national security preparedness and response, reduce congestion for all Americans and increase global transportation connectivity in support of Nation's economy.)</i>	2.5	0
		2	Does the dataset have the potential to support multi-modal or cross-modal analytics and decision-making? <i>(that is, does the data support the mission of more than one operating administration)</i>	2	0
		3	Could the data enhance potential contribution to the creation of novel and useful third-party applications and services?	2	0
		4	Is the data of statutory reporting requirements?	2	0
		5	Will the release of this data contribute to improve accountability?	2	0
		6	Could the data support decision making or emergency response activities at the individual, state, local, DOT or other external agency's level?	2	0
	Usability	1	Are the data entries clearly defined and described?	2.5	0
		2	Is the data consumer ready?	2.5	0
		3	Does the data have breadth of coverage <i>(for example, national versus local)</i> ?	2.5	0
		4	Is the frequency of data usage monitored?	2.5	0

Figure 2.7 Data.gov data scoring criteria (Data.gov Working Group 2010)

NCHRP

The project team reviewed one NCHRP resource, which is briefly summarized below.

NCHRP Improving Employment Data for Transportation Planning

An NCHRP guidebook, *Improving Employment Data for Transportation Planning*, compares three publicly available employment data sources to identify which may be most useful for transportation planning (National Academies of Sciences, Engineering, and Medicine 2011). The report provides an in-depth review of each data source before comparing them. While some of the categories are employment data specific, the following considerations could be applied to broader data investment evaluations:

- **Geography** – Compares the geographic granularity of each data source.
- **Update frequency and timelines** – Reviews the frequency with which the data sources are updated.
- **Stratification** – Compares the variables available that can be used to split the data for various analyses. In the case of employment data, this includes employer size, type of job, and age of workers. The source with more and more relevant variables that stratify the data, the better.

Academia and industry

The project team reviewed four academic and industry resources. Each is briefly summarized in the sections that follow.

AICPA criteria for describing a set of data and evaluating its integrity

The American Institute of Certified Public Accountants (AICPA) and the Chartered Institute of Management Accountants developed a white paper to provide its members with a framework for describing then evaluating the integrity of data (AICPA Assurance Services Executive Committee 2020). The white paper describes the following criteria:

- **Purpose** – Datasets should be collected for a specific reason, which informs what data are included.
- **Population of the dataset** – Defining the population of the data collected provides insight into the strengths and weaknesses and how data can and should be employed.
- **Nature of the elements** – The definition of each observation and variable should be clear to fully understand what is being measured. For example, if a dataset includes an economic output measure, it should also define whether the output is in dollars or a relative measure.
- **Sources of data** – Defined sources of the data help avoid misinterpretation. The source description should include specificity and methodology such that the user could theoretically collect and replicate the data on their own.
- **Units of measurement** – Each variable's units should be clear.
- **Accuracy, correctness, or precision of measurement** – Understanding the accuracy, correctness, and precision measurements ensures the user understands any shortcomings of the data and employs the data properly.
- **Uncertainty in the data and population** – Every estimate will have uncertainty, so when estimates exist in the data, the uncertainty should be published. This can include:
 - Standard deviation

- Historical variations
- Margin of error
- **Date of measurement or period of occurrence** – The date when the data was collected and when it might be updated is critical for determining how to employ the data.

US Department of Education Decision Criteria and Recommendations

The United States Department of Education (DoEd) allocates federal funds for English language learners based on attendance. The DoEd can either use state reported data or the American Community Survey. The DoEd engaged the National Research Council to produce a report, *Allocating Federal Funds for State Programs for English Language Learners*, to evaluate which dataset is more accurate (National Research Council 2011). Chapter 7, “Decision Criteria and Recommendations,” outlines the criteria that can help the DoEd make the decision. The criteria that can be readily applied to data investment decisions are as follows:

- **Conceptual fit** - How well does the data fit the research question?
- **Level of geographic detail** – Do the data reach the desired geographic granularity?
- **Timeliness** – What is the time between when the data was produced and when it is used?
- **Quality** – Quality includes:
 - **Utility** – Usefulness to the user
 - **Objectivity** – Whether the data is accurate, reliable, unbiased, and clear and presented in an accurate, clear, unbiased manner
 - **Interpretability** – The availability and completeness of the documentation
 - **Integrity** – Protection from unauthorized access or revision
 - **Accuracy** – Difference between actual values and estimates
 - **Comparability** – Similarity with other sources across geographic and demographic dimensions
- **Stability** – Do the estimates change much over time and across updates?
- **Comparability** – Does the method does not change across regions?
- **Transparency** – Are assumptions and methods clear?

The report evaluates each dataset based on the criteria with a “+” or “++” determination, but eventually recommends employing both datasets to balance out each of their shortcomings.

Institution of Engineering and Technology Big Data in Transport

The Institution of Engineering and Technology provides *Sector Insights* reports that cover various industry topics. The *Big Data in Transport* provides an overview of how big data can and should be used in the transportation industry (Institution of Engineering and Technology n.d.). The report provides some data quality requirements, but does not define them in-depth:

- Spatial granularity (line, station, urban road network, link, junction, etc.)
- Temporal granularity (minutes, hours days, annual, etc.)
- Direction discrimination
- Modal discrimination
- Sample size within provided spatial and temporal quantities
- Bias (free or bias)

APTA Procuring Software to Support Transit Asset Management

The American Public Transportation Association (APTA) put together the report, *Procuring Software to Support Transit Asset Management*, to help transit agencies make software procurement decisions (APTA Standards Development Program 2020). The strategies recommended in the report, while for software, may be applicable to data investment decision. The APTA recommends a fit-gap analysis, which includes the following steps:

1. **Define the need** – Define the analyses that require or can benefit from new data.
2. **Data requirements** – List the data requirements that the analyses require.
3. **Prioritize requirements** – Prioritize the requirements as (1) mandatory; (2) highly desirable; or (3) desirable.
4. **Existing data evaluation** – Evaluate how existing data meets or could meet the requirements.
5. **Determine the gap** – Identify the areas that existing data cannot meet the need and then evaluate options in the context of the gap.

Other resources

The project team reviewed two additional resources that did not fit into the aforementioned categories. Each is briefly summarized in the sections that follow.

Gartner Magic Quadrant

Gartner is a nationally recognized management consulting firm that has coined a “Magic Quadrant” to help with decision making (Gartner n.d.). The “Magic Quadrant” is an evaluation matrix that maps options based on dimensions, typically two. It is especially helpful to make decisions with tradeoffs as the decisions furthest between the two dimensions are optimal.

Gartner names the quadrants, which can help provide a conceptual understanding of how the two dimensions can interact (**Figure 2.8**).

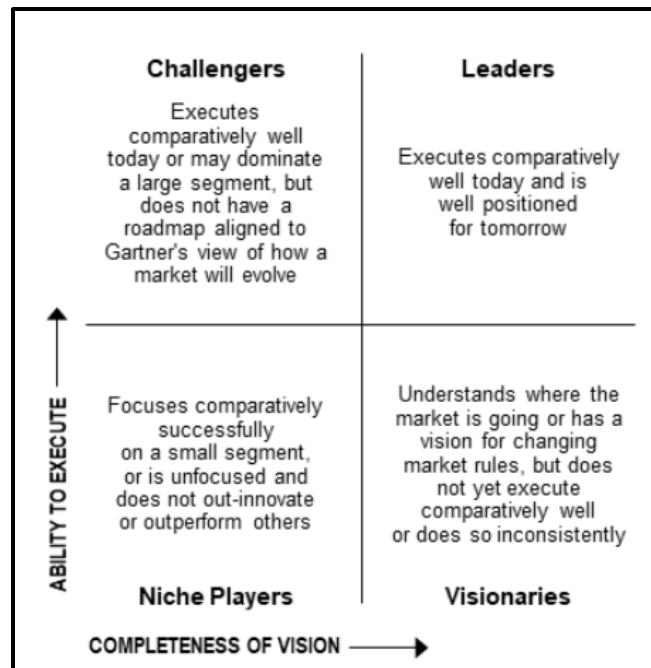


Figure 2.8 Gartner Magic Quadrant (Gartner n.d.)

Forbes How to Choose the Right Data Vendor

Forbes Technology Council produces an article, *How to Choose the Right Data Vendor*, to help inform readers how to make data investment decisions (Gladys Kong 2019). The article provides the following considerations:

- **Compare the vendor's data to whatever data you already have** – Perform a validity assessment by comparing the vendor's data to existing data.
- **Get client referrals** – Vendors that are confident in their product will allow you to speak with other customers.
- **Make sure they are transparent about any potential biases** – Bias exists in most datasets, so a reliable data vendor will be transparent about the biases in their data.

- **Ask the vendor if they're willing to work with you on a test case** – Good data vendors will provide data samples to allow customers to perform their own evaluations.
- **Determine whether data are customizable and scalable** – Customizable data improves the data's ability to meet customers' specific needs. Similarly, considering the breadth of the data helps ensure that all desired geographies or stratifications are available.
- **Find out whether the vendor offers training**– Not all data is easy to use or intuitive, especially if a data portal is included. Vendor training or support improves the user experience and confidence that the data is being employed properly.

Synthesis of Data Evaluation Criteria and Methods

Considering the evaluation criteria identified through the literature review, they can be separated into four classifications: Value Creation, Quality, Integration, and Cost. Using these broader criteria divisions is supported by ITD's *SWAAI* considering Coverage, Usability, and Valuable, IDOT's *Data Management Strategic Plan* using a similar division with Important, Quality, and Useability and the *Data.gov Interim Identification & Prioritization Process and Guidelines* employing the three dimensions Value, Usability, and Availability. The four classifications – Value Creation, Quality, Integration, and Cost –encompass the criteria categories employed by the other sources. These criteria classifications can be defined as such:

- **Value Creation** – Value creation includes criteria that evaluate the dataset's abilities to satisfy the agency's need, save time or money, or provide benefit for multiple departments and staff. It could require an internal needs assessment and a clear record describing the features of the data options.
- **Quality** – Quality classification covers the various metrics for evaluating the data's reliability, accuracy, precision, etc. In addition, it can refer to the quality of the provider, such as training and assistance provided, transparency in methodology, and user reviews.
- **Integration** – integration classification covers usability. It seeks to answer the question of how easily ITD will be able to incorporate and analyze the different data options.
- **Cost** – Cost acts as the tradeoff factor and it bounds any decision and provides an objective metric for comparison across data investment decisions.

Some sources that instead of, or in addition to, providing specific criteria, discussed methodologies for decision making. For instance, the ITD's Contracting Procurement Decision Tree and the Gartner Magic Quadrant did not mention relevant criteria but offer two strategies that can be leveraged with criteria to inform decisions. These Decision-Making Methods provide frameworks that employ criteria to inform or reach a final decision.

Table 2.2 provides a summary of the information covered in the literature review categorized into one of the criteria classifications or as a decision-making method. The entire list of criteria and methods

discussed in the literature review can also be found in a table format within a supplemental deliverable prepared by the project team. The classifications do not provide absolute categorizations as some criteria could fit in different classifications depending on how it is defined. They do offer a method for considering the different types of criteria and how they can be applied.

Table 2.2 Literature review synthesis studies

Resource Type	Count
Value Creation	42
Quality	47
Integration	13
Cost	7
Decision-Making Method	11

Value Creation

An identified need should initiate a data investment decision. The Value Creation classification aims to capture the criteria that evaluate the dataset's ability to meet the need, or potentially go beyond it. Some materials employed a single metric, such as COG's *Big Data for Regional Travel and Mobility Analyses* report with "Applicability to research area" or Iowa DOT *Data Management Strategic Plan's* "Unaddressed data needs". The sources that provide multiple metrics demonstrate how a comprehensive evaluation of value relies on a detailed understanding of the need and the data. For instance, ITD's *SWAAI* include three criteria than be classified as Value Creation: 1) need assessment, the understanding of the reasoning behind the data investment can be equated to value; 2) coverage, the understanding of the data type and geography will help determine how and who can employ the data; and 3) value, the calculation of benefits based on different use cases. The following are a few examples of considerations and criteria based on the literature review:

- **Meets the need** – Determining how well the data applies to the need, which requires a clear definition of the need. How clear is the data need and could it evolve?
- **Coverage** – Does the granularity, feature breadth, and geography of the data meet the data need?
- **Usefulness** - How often will the data be used? Who will and can use it?
- **Data considerations** – What is the temporal resolution? Is the data static or dynamic? Is it updated regularly?

Data and Provider Quality

There are many criteria to measure quality, especially criteria for measuring the quality of the provider in addition to the data. USDOT's Data.gov Interim Identification & Prioritization Process and Guidelines

report defines quality with two criteria, completeness and consistency, while the Caltrans document focuses almost entirely on quality. This demonstrates the variety of methods for evaluating quality for data. Considering the providers, the Forbes article includes questions that can help evaluate the quality of providers, such as interviewing client referrals and assessing the transparency of their methodology. A few examples of ways to evaluate quality of data and provider include:

- **Data quality**
 - **Accuracy** – Comparing data to known values.
 - **Completeness** – Considering the number of missing and inferred values.
 - **Timeliness** – Recency between estimation and the desired application.
- **Provider quality**
 - **Training and responsiveness** – Does the vendor provide training and a helpdesk to answer questions that may arise?
 - **Reputation** – Is the data provider known for quality or excellent service?
 - **Data test** – Will the vendor share a sample of the data for evaluation?

Data Integration

Data does not exist in a silo, especially if the product is to be used across departments, in various analyses, and over time. The Data Integration classification covers criteria that measure the degree of effort to incorporate and analyze data. For example, ITD's *ETS Technology Request Intake Process* includes Usability and Ensure Compatibility and the FHWA *Data Quality White Paper* mentions Accessibility/Usability. Examples of Data Integration criteria include:

- **IT requirements** – How much time is required by IT staff to integrate the data?
- **Usability** – Is training or different software needed to properly analyze the data?
- **Licensing and sharing** – Are there limits on who can review and analyze the data?

Cost

The options with the best Value, Quality, and Integration will always be the right decision before considering cost. The Cost bounds decisions to ensure that the benefits, caliber of product, and ease of use outweigh the investment. Cost is considered in the COG *Big Data for Regional Travel and Mobility*, both Cost and Pricing Structures and the Data Cost. Examples of Cost metrics include:

- **Licensing cost** – The upfront cost for accessing and downloading the data. Includes any one-time fee, a subscription, and pay-per download.

- **Storage cost** – The cost to store new data on agency servers.
- **Operations and maintenance cost** – The cost to collect, upkeep, and integrate the data, typically measured in hours spent

Decision-Making Methods

The criteria provide metrics for evaluation, but the Decision-Making Method provides the framework that, when followed properly, should lead to a thoughtful decision. This can be as straight forward as a binary evaluation, as done in the Institution of Engineering and Technology (IET) *Big Data in Transport* report, or a more complex weighting system, as used by the USDOT's Data.gov *Interim Identification & Prioritization Process and Guidelines* decision making. Alternatively, criteria can be quantified and evaluated through a Benefit Cost Analysis, as was done in the FHWA *Decision-Making Guidebook*.

A second consideration is scoring. For example, in the COG *Big Data for Regional Travel and Mobility* report, technical experts evaluate the options, while ITD employs a decision tree for contract procurement and an Intake Council for technology requests.

3. Surveys, Meetings, and Interviews

Task Purpose

This task included an internal and external survey (collectively the “Surveys”), an internal townhall style discussion with ITD data experts, several meetings with the eight-member Technical Advisory Committee, and two interviews with peer state DOTs. The information collected during this task helped prioritize data evaluation methods, identify ITD priorities, and learn about actual data asset decision-making practices from peer states. The chapter is divided into the following subsections:

- **Internal Survey**
- **External Survey**
- **ITD Townhall and Technical Advisory Committee**
- **Peer State Interviews**

Internal Survey

The internal survey aimed to gather information from a broad cross section of ITD staff on current processes around data investment decision making. This information helped the project team to identify the development methodologies that would best suit ITD. This section outlines the key takeaways from the responses and provides a review of each question by survey section. The complete internal survey form is provided in **Appendix A. Surveys**.

Key takeaways

Survey responses revealed common themes and practices, which are outlined below:

- Most respondents reported data gaps, such as insufficient variable coverage, when asked about unmet data needs.
- Most respondents believe that existing data and software are currently underutilized by ITD staff.
- Quality, specifically accuracy, was reported to be the most important consideration, but there was recognition that is difficult to evaluate.
- Responses were split for primary decision-making methods between evaluating business use cases and leveraging professional judgement.
- Most respondents agreed that department wide data prioritization and governance would address many challenges.

Respondent information

The first five questions in the survey collected the respondents' names, emails, department, position, and data-related responsibilities. Name and email enabled the project team to reach out to respondents with follow-up questions and the latter three questions allowed the project team to categorize responses by department and ensure that an appropriate cross section of staff was represented.

During the one month that the survey was live, 41 members of ITD staff submitted responses. The respondents' data responsibilities included data creators, analysts, and managers. The following departments were represented:

- Planning & Scoping
- ETS Service Center
- Pavement Management
- Highway Safety
- Materials & Planning
- Traffic Operations
- Construction Materials
- Government Affairs
- Data Analytics
- District Staff

Current data asset procurement processes

This section of the survey focused on how ITD departments address data needs and make data investment decisions. It included three selection questions and one open-ended question. Summarized responses and an analysis of each question follows.

Q6. Do you or your department currently have any of the following unmet data needs? [Select all that apply]

The identification of data gaps as a primary unmet data need (**Figure 3.1**) indicates that a feature of the tool that facilitates or includes an evaluation of these gaps could be useful for ITD staff. It could also help staff present their needs and efficient solutions in an approachable format. The need for more recent and better-quality data as the second and third most common response speaks to the importance of recency and quality when evaluating data options.

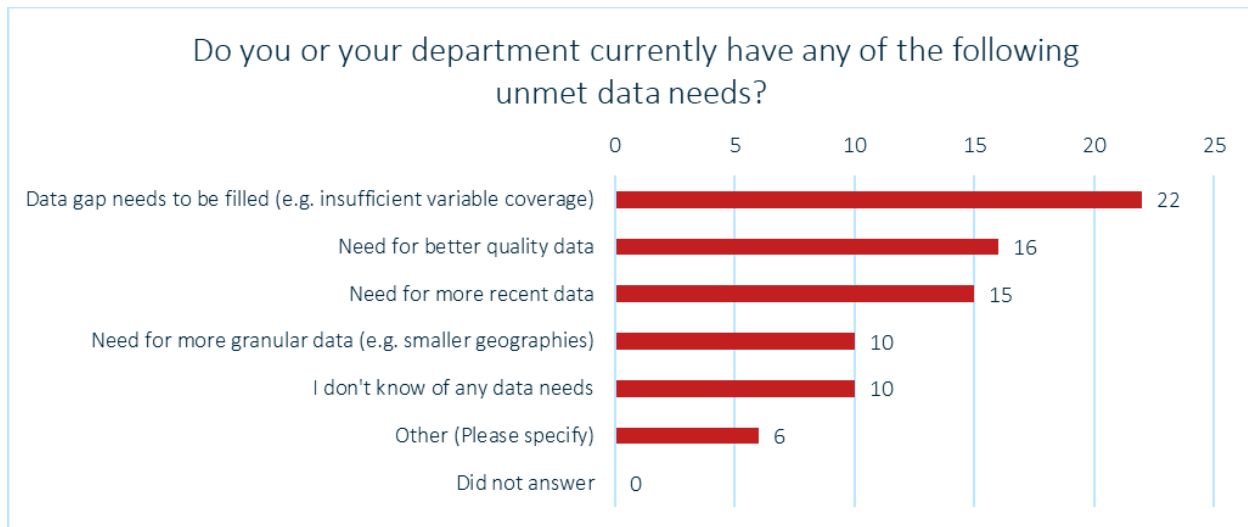


Figure 3.1 Survey of unmet data needs among respondents

Q7. Which of the following approaches do you or your department employ for addressing data needs? [Select all that apply]

Approaches to addressing data needs vary among staff and includes data collection, public resources, and purchasing third-party data (**Figure 3.2**). The fairly even response split for invest in data collection (19), identify public resources (16) and purchase third-party data (15) further supports the need for this research as there is no identifiable standard across the Department.

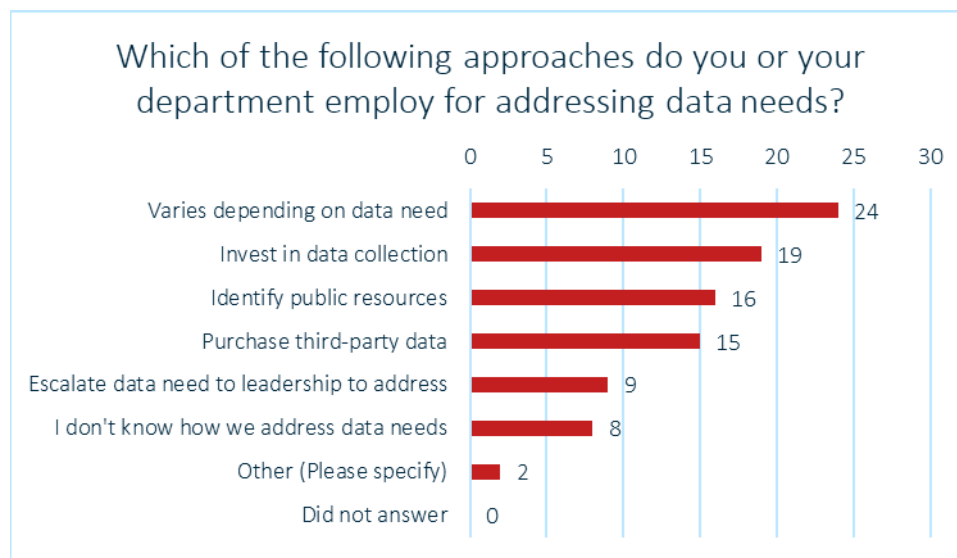


Figure 3.2 Respondents' approaches to addressing data needs

Q8. How do you or your department primarily evaluate whether a dataset is important enough to invest in? [Select one]

When evaluating the importance of whether a dataset is worth the investment, there is no consensus across ITD (**Figure 3.3**). While identifying business use cases was a common selection, there are almost as many “no formal process” and “I don’t know” selections. Since business use cases and professional judgement are two common methods, they were highlighted for incorporation into the evaluation tool. This could include identifying as many business use cases as possible and incorporating professional judgement for some qualitative aspects.

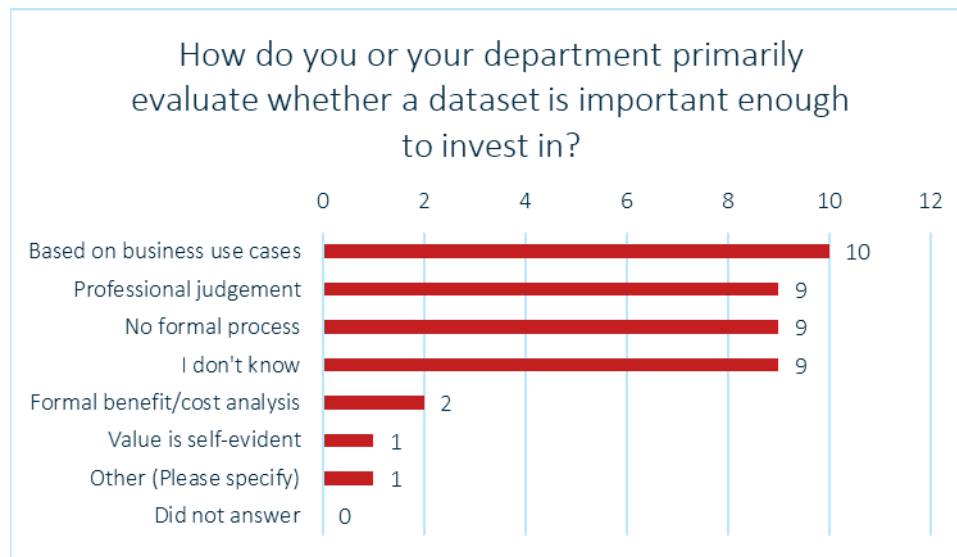


Figure 3.3 How respondents evaluate a dataset's value

Q9. What specific challenges or limitations do you or your department encounter when attempting to determine the value of a dataset? [Open-ended]

The final question was open-ended and asked for feedback about challenges. Many comments were related to a general lack of agency-wide prioritization and governance, which would help with decisions around maintenance, recollection, integration, and other considerations. Prioritization and data-governance processes may also improve communication around dataset utilization and value. Secondly, respondents commented on the desire to assign value to datasets, involving business use case identification and clear communication with leadership. Specific comments are summarized below.

- Respondents cited data quality and consistency challenges, such as data age and missing or incomplete data, especially when aggregating across sources.
- The data incompleteness speaks to doubts about whether the data presents an accurate, overall picture of existing conditions. For instance, does the data truly represent the entire state or only populated regions?
- Comments included the need to identify data management roles and responsibilities within the agency, such as data owner or maintainer and integration lead.

- ITD staff responded that ITD could do better at identifying, sharing, and articulating use cases, which would allow for better data utilization. This requires working across departmental silos.
- Respondents suggested that improving business case identification and developing a work flow integration plan would help increase dataset utilization.

Given these responses, the tools could incorporate business use identification as a means for determining the full value of a dataset.

Making data investment decisions

The questions in this section target processes, important considerations, and challenges in making data investment decisions. The purpose was to give the project team guidance on how to address or avoid pitfalls and account for ITD priorities. Summarized responses and an analysis of each question follow.

Q10. What do you think are the three most important considerations when making data-related decisions (e.g., investing in internal data collection, purchasing third-party data, using a public resource, etc.)? [Select three]

When evaluating the most important considerations for making data related decisions, data quality rose above the others (**Figure 3.4**). Next were the ability for the data to satisfy a previously specified need and the ease at which the data can be integrated, utilized, and maintained. These priorities speak to the staff's desire for quality data that fits their needs above all else, which could be an important purpose for the tool.

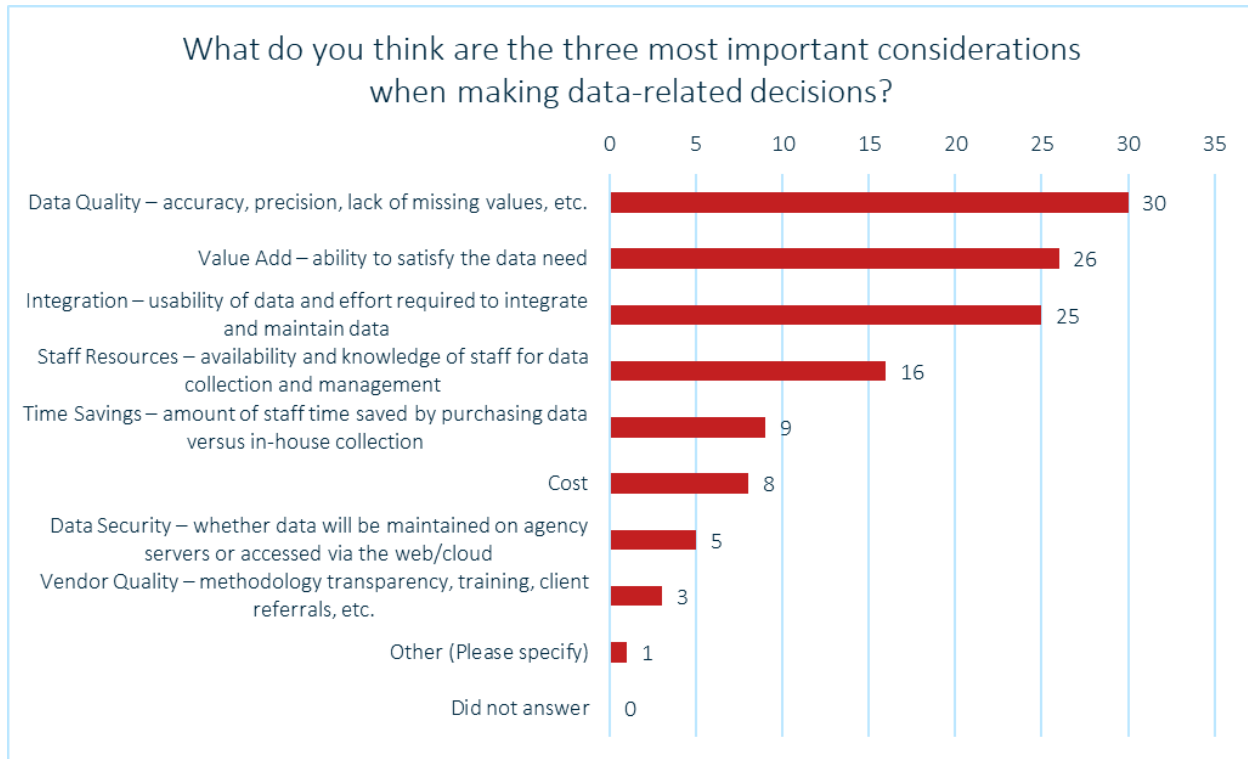


Figure 3.4 Respondents' most important considerations when making data-related decisions

Q11. What specific challenges or limitations have you or your department encountered when attempting to determine whether and how to invest in a dataset and, if any, what strategies have you or your department successfully used to overcome these challenges and limitations? [Open-ended]

The comments on the challenges that staff face for determining whether to invest in a dataset mention limited staff and funding, differences in priorities across the agency, and communication. The tool could, in part, address the communication challenges by improving coordination in dataset evaluation and utilization, as well as help ITD staff better articulate dataset investment value. Another response highlighted the importance of a data integration plan, both in terms of use case and data infrastructure to ensure that an investment can be utilized effectively and efficiently.

Q12. What do you believe are the two most important considerations when evaluating data quality? [Select two]

Considering data quality, accuracy is most important, followed by completeness, and then timeliness (**Figure 3.5**). These rankings could influence the weighting of data quality elements within the tool. The difference in responses could also speak to differing data needs. For instance, recency could be more important for some rather than completeness, because missing values could be imputed.

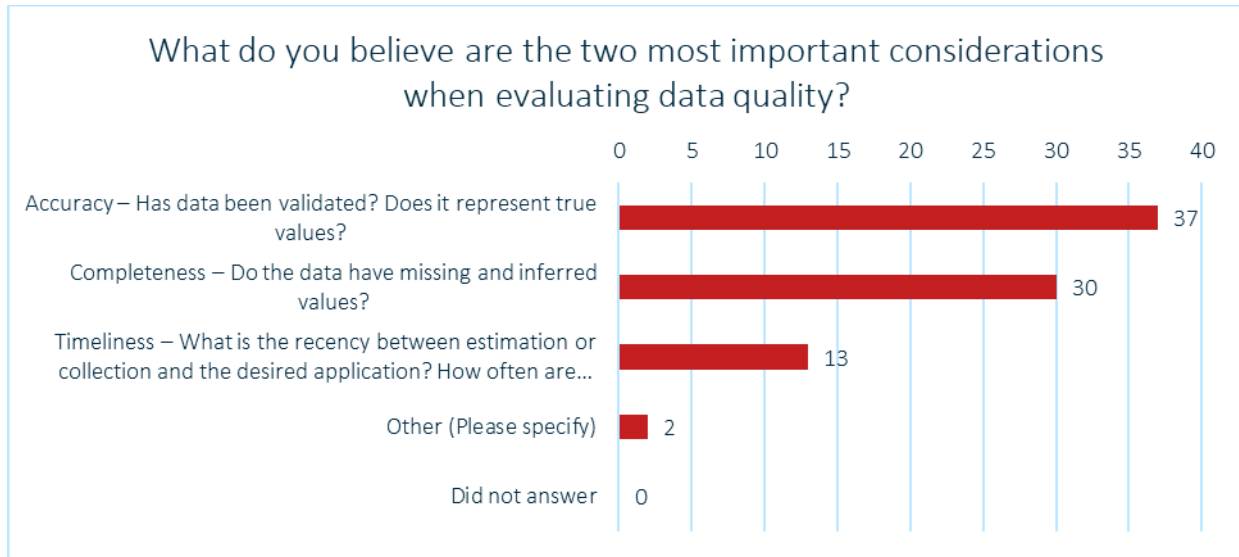


Figure 3.5 Respondents' most important considerations when evaluating data quality

Q13. What, if any, specific challenges or obstacles have you or your department encountered when attempting to verify data quality (e.g., accuracy, completeness, etc.)? [Open-ended]

Respondents reported challenges in verifying data quality such as:

- Departments are short staffed, and the verification process is time-consuming.
- No manuals for evaluation, so it can feel like recreating the wheel.
- Lack of access to vendor data.
- Disagreement on what is sufficiently accurate or the true values.

Q14. What are the two most important data integration considerations for your work? [Select two]

The final question asks about data integration with the primary considerations being usability, or ensuring the staff have the correct software and skills to analyze the data (**Figure 3.6**). The second highest priority was licensing or sharing, which has to do with who can access, review, or receive the data. These can be prioritized when evaluating third party vendors or even contracts. Privacy was ranked lower, supporting the conclusion that priorities differ across departments.

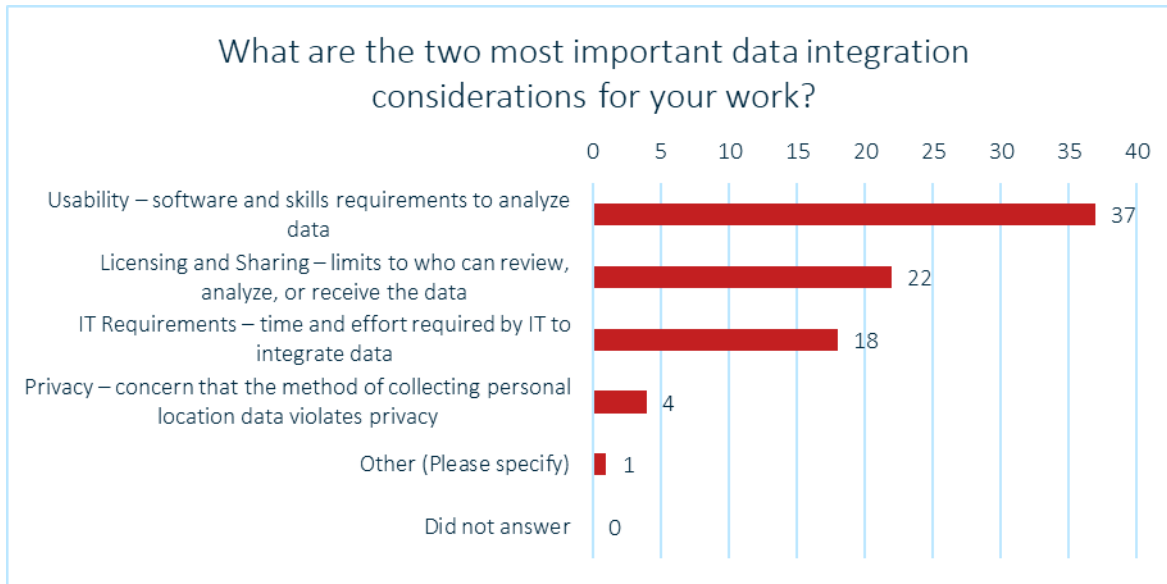


Figure 3.6 Respondents' most important data integration considerations

Third-party data experiences

The final section was designed to learn about ITD's experiences using third-party (i.e., vendor-provided commercial data available for purchase) datasets. The project team wanted to learn what to account for when evaluating data vendors. Summarized responses and an analysis of each question follow.

Q15. What third-party, commercial datasets have you utilized for your work at ITD? If none, please enter "None." [Open-ended]

The first open-ended question produced a list of vendors and datasets that have been used at ITD. These are included below.

- American Association of State Highway and Transportation Officials (AASHTO)
- American Transportation Research Institute
- Bidx
- Cambridge Systematics
- ConceptStation by Bentley
- Cyclomedia
- ESRI
- Georgetown Climate Center
- IHS Markit
- Inrix
- landropDATA
- Michelin
- Numetric
- QuestCDN
- Replica

- Telemetric
- TREDIS
- Terralytics

Q16. How do you or your department identify potential use cases for third-party commercial data?
[Select all that apply]

Respondents provided insight into how data vendors are identified (**Figure 3.7**). Most often staff responded that vendors are considered when there is an opportunity for increased efficiency or when staff indicate a clear need. Answers to this question allude to the importance of the tool because ITD staff state that they have identified efficiencies or a data need, but earlier reported difficulty articulating the data value. The tool could help evaluate the ROI of the opportunity and could avoid investment in well-presented but less useful products.

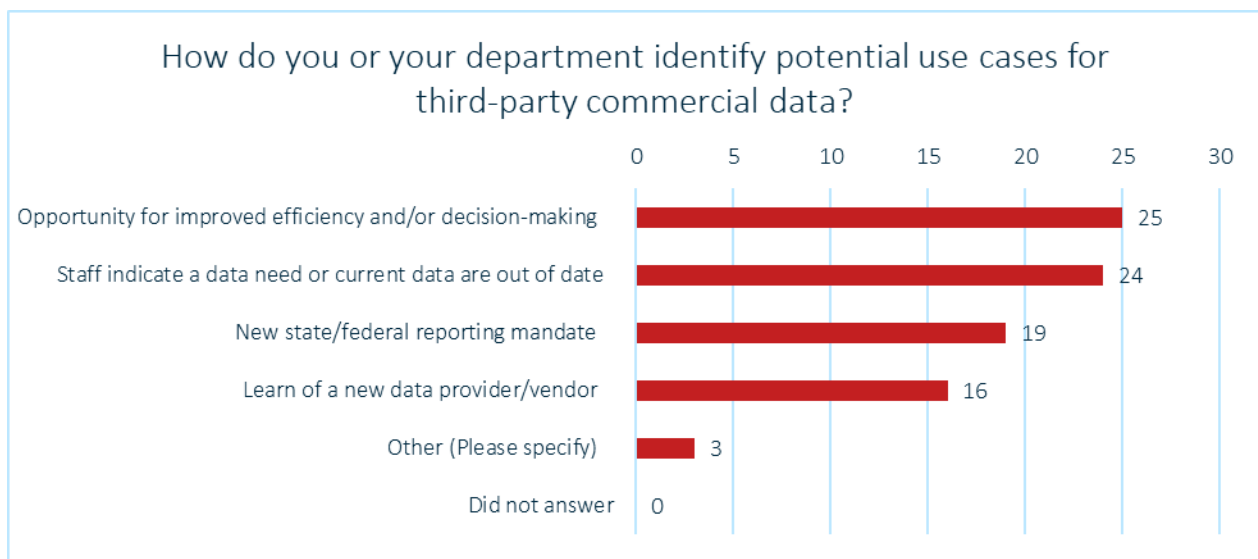


Figure 3.7 How respondents identify use cases for commercial data

Q17. What validation or evaluations do you or your department perform before purchasing third-party commercial data? Please include specific metrics if possible. [Open-ended]

Answers to this question include the following:

- ITD staff or consultant recommendation
- Political interests
- Review data samples and compare them to existing data
- Rough cost vs. service assessment
- Peer state research
- Security vs. cost assessment
- Use case assessment
- Pilot program

The tool could incorporate the use of data samples for evaluating vendors as well as the possibility for a pilot program, which can help ITD identify use cases.

Final comments

The internal survey ended with an open-ended question asking for any additional comments or considerations. Many spoke to the importance and value of this project with two comments specifically highlighting that many internal systems were developed “by us for us,” which has created pockets of governance. The tool could provide a step toward greater collaboration across the agency and help staff communicate data need and value to leadership.

External Survey

The external survey gathered information from peer state DOTs and fellow Idaho agencies to crowdsource processes and priorities around data investment decision making. Moreover, the project team was able to review the list of respondents to identify peers for interviews. This section highlights the external survey key takeaways and offers a review of the responses by section. The complete external survey form is provided in **Appendix A. Surveys**.

Key Takeaways

The following themes and practices emerged in the responses:

- Agencies focused on business use cases when evaluating data-needs.
- Agencies noted challenges in capturing the full value of data.
- Data accuracy is the highest priority for agencies, but it is difficult to assess.
- Agencies highlighted the importance of governance, data warehousing, and ownership.

Respondent information

Similar to the Internal Survey, the first section of the External Survey collected the respondents’ names, emails, department, position, and data-related responsibilities. This information, along with the rest of the respondent’s answers, was used to identify peer states to interview. The project team received 30 responses from agencies from across the country. Those agencies included:

- Arkansas DOT
- Caltrans
- Connecticut DOT
- Florida DOT
- Georgia DOT
- Idaho Department of Fish and Game



- Kentucky Transportation Cabinet
- Louisiana Department of Transportation and Development
- Maine DOT
- Minnesota DOT
- Montana DOT
- Nebraska DOT
- New Hampshire DOT
- New Jersey DOT
- New Mexico DOT
- New York State DOT
- North Carolina DOT
- Oklahoma DOT
- Oregon DOT
- Puerto Rico Highway and Transportation Authority
- South Carolina DOT
- Texas DOT
- Vermont Agency of Transportation
- Virginia DOT

Respondents' roles ranged from managers to engineers and planners:

- Chief Data Officer
- Enterprise Data Program Manager
- Deputy Chief Innovation Officer
- IT Manager
- Project Manager Chief
- System Performance Engineer
- Software and Data Bureau Chief
- Principal Engineer
- Strategic Data Scientist
- Geospatial Data Officer

The data responsibilities were similarly diverse. Respondents managed datasets including Highway Performance Monitoring System (HPMS), GIS, and safety, as well as oversaw processes and governance. A review of the external respondents indicates a broad sample of states, roles, and responsibilities, providing the project team with a thorough understanding of peer agency processes and considerations.

Current data asset procurement processes

This section of the external survey asks peer agencies how they address data needs and make data investment decisions. Summarized responses and an analysis of each question follow.

Q6. What leads you or your organization to consider data investment decisions (e.g., investing in internal data collection, purchasing third-party data, using a public resource, etc.)? [Select all that apply]

Responses were close with respect to events that trigger a data investment decision-making process. Among the top answers were staff indicating a need, new policy mandates, or an opportunity for

improved efficiency triggering a data investment consideration (**Figure 3.8**). Only a third of peer agencies indicated that they consider data investment when they learn about opportunities from vendors or providers. “Other” responses included peer agency and partnership knowledge sharing and new technologies or methods.

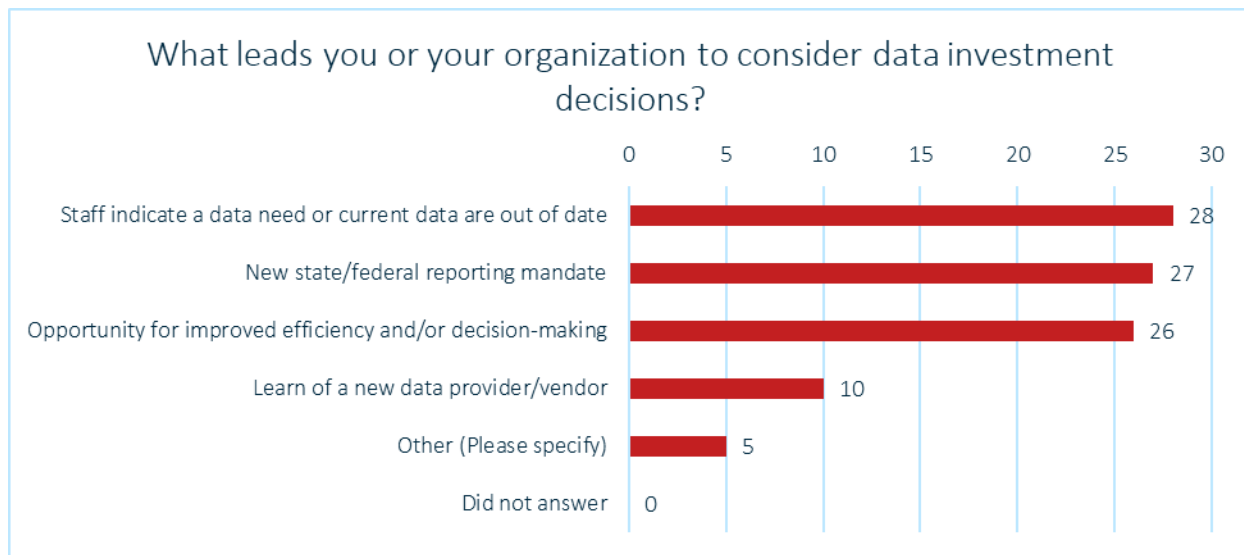


Figure 3.8 Survey of unmet data needs among external respondents

Q7. Which of the following approaches do you or your organization employ for addressing data needs? [Select all that apply]

Peer agencies commonly responded that their approaches to addressing data needs varied and primarily included investing in data collection and purchasing third-party datasets (**Figure 3.9**). Over two thirds of the agencies noted that they invest in data collection methods and/or recognized that different data needs required varying approaches. A slightly smaller number of respondents reported purchasing third-party data as an approach, while just over half use publicly available datasets. Notable “Other” responses included employing Artificial Intelligence or Machine Learning or relying on data sharing agreements.

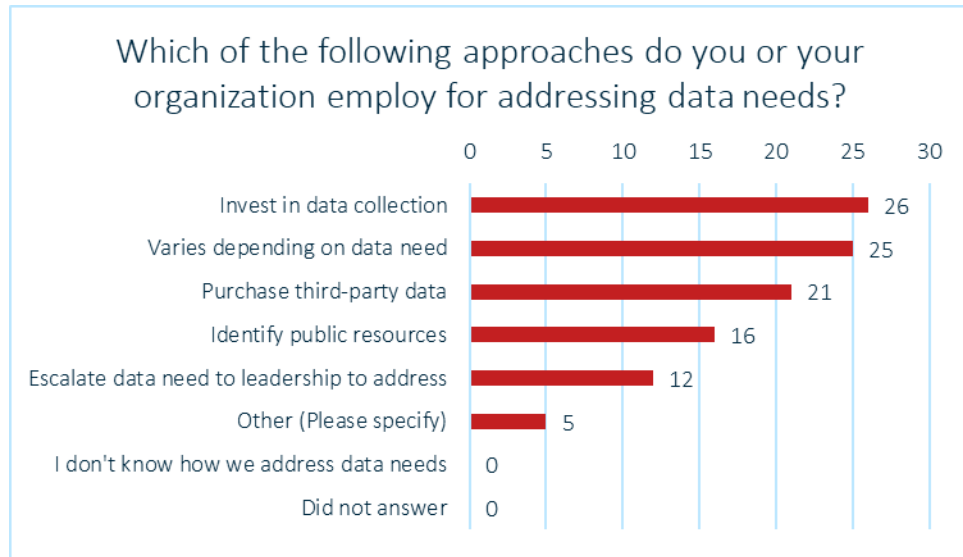


Figure 3.9 External respondents' approaches to addressing data needs

Q8. How do you or your organization primarily evaluate whether a dataset is important enough to invest in? [Select one]

Most peer agencies made data investment decisions based on business use cases (**Figure 3.10**). Professional judgement was also used, and a few agencies employ a formal BCA. One “Other” response specified that the agency reviews peer state research.

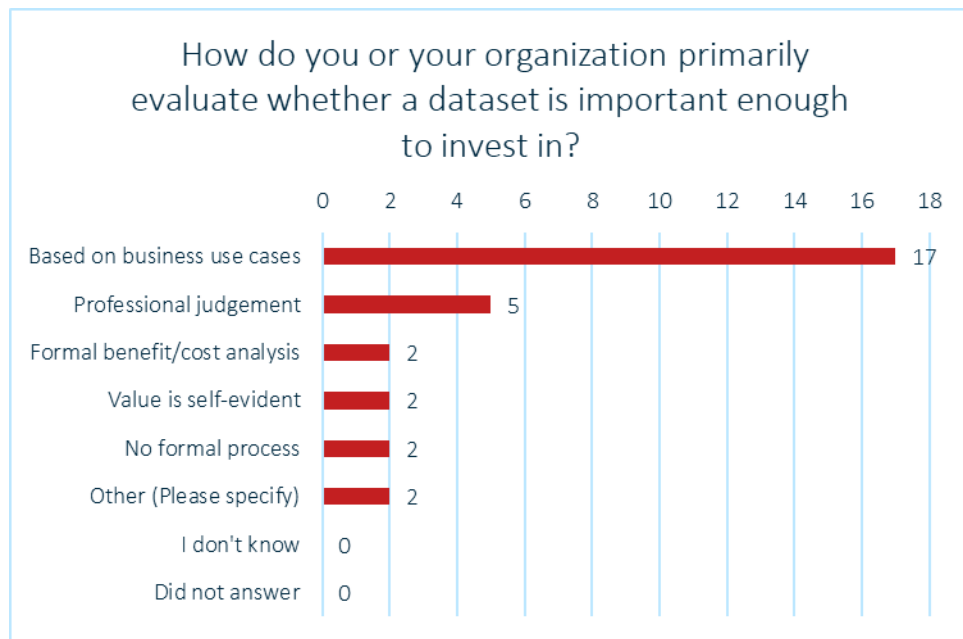


Figure 3.10 How external respondents evaluate a dataset's value

Q9. What specific challenges or limitations do you or your organization encounter when attempting to determine the value of a dataset? [Open-ended]

Common themes in responses to the final question included understanding and communicating the value of datasets, while ensuring that the entire value is captured. Making such decisions also brought up a few other challenges such as determining whether similar data exists, identifying stewards, and evaluating the data for accuracy. The specific challenges included:

- Determining, measuring, and communicating benefits to leadership while incorporating alignment with strategic directives
- Capturing the true, overall value of data; currently some have limited use cases
- Procurement and a BCA are lengthy processes
- The cost of data can be prohibitive
- Assessing the raw data vs processed data tradeoff
- Identifying stewards: managing and updating data
- Integrating new data into existing data systems
- Evaluating accuracy
- Determining whether a similar dataset already exists

Given this feedback, the tool could consider business use cases and address common data investment concerns. The project team should also ensure that using the tool does not require too much staff time.

Making data investment decisions

The questions in this section asked peer agencies about processes, important considerations, and challenges in making data investment decisions. Summarized responses and an analysis of each question follow.

Q10. What do you think are the three most important considerations when making data-related decisions (e.g., investing in internal data collection, purchasing third-party data, using a public resource, etc.)? [Select three]

When ranking the most important considerations for making data-related decisions, most respondents selected value-add, followed closely by integration (**Figure 3.11**). Staff resources, data cost, time savings, and privacy were not commonly selected as considerations. Value-add as the most common response aligns with earlier peer agency responses that business use cases were the primary driver behind data investment decisions. Integration as the second highest consideration could indicate an important component of the tool.

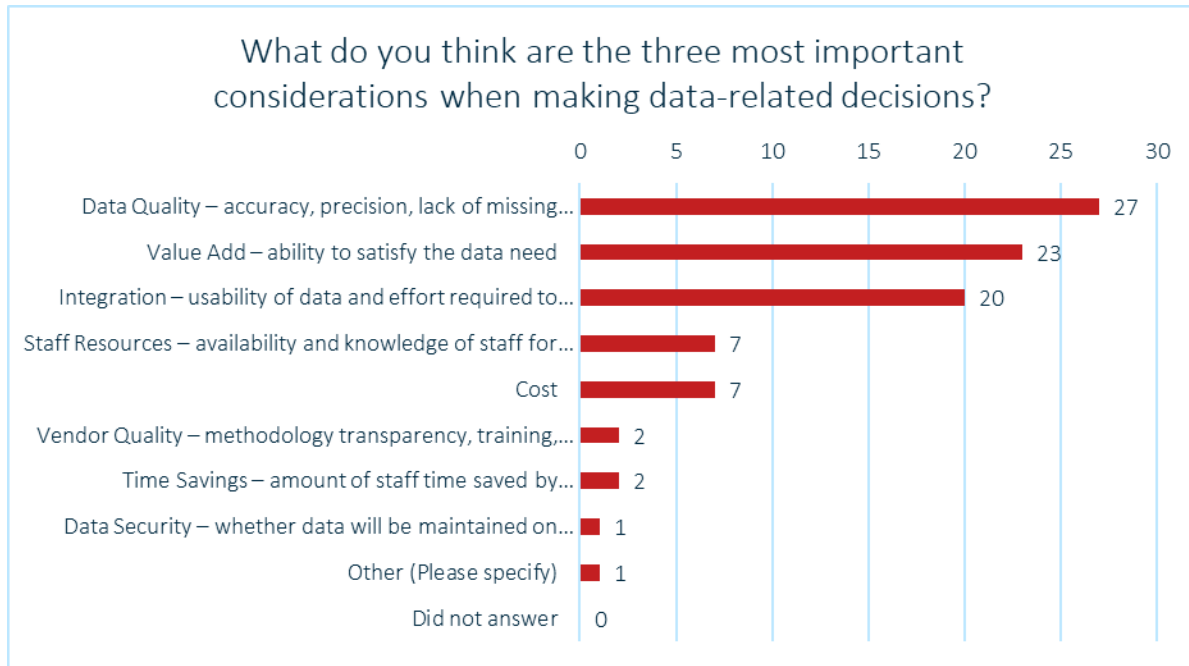


Figure 3.11 External respondents' most important considerations when making data-related decisions

Q11. What specific challenges or limitations have you or your organization encountered when attempting to determine whether and how to invest in a dataset and, if any, what strategies have your agency successfully used to overcome these challenges/limitations? [Open-ended]

When asked about challenges or limitations that arise when making data investment decisions, peer agencies responded with:

- Issues with vendors, including integration and method transparency
- Lack of governance around ownership and maintenance
- Cost
- Determining the quality: accuracy, completeness, and representativeness
- Sharing permissions
- Pilot potential
- Siloed departments
- Leadership advocating for a product with good marketing
- Data security and storage for external sources
- Staff time spent analyzing new data, which means abandoning analysis of existing data
- Needs being immediate rather than longer term, which leads to investments lacking broader value

Q12. What do you believe are the two most important considerations when evaluating data quality? [Select two]

Most agencies noted that the data's accuracy and completeness were the most important considerations when evaluating the quality of the data (**Figure 3.12**). This aligns with the internal survey results, indicating that accuracy and completeness could be weighted more heavily in a data quality evaluation.

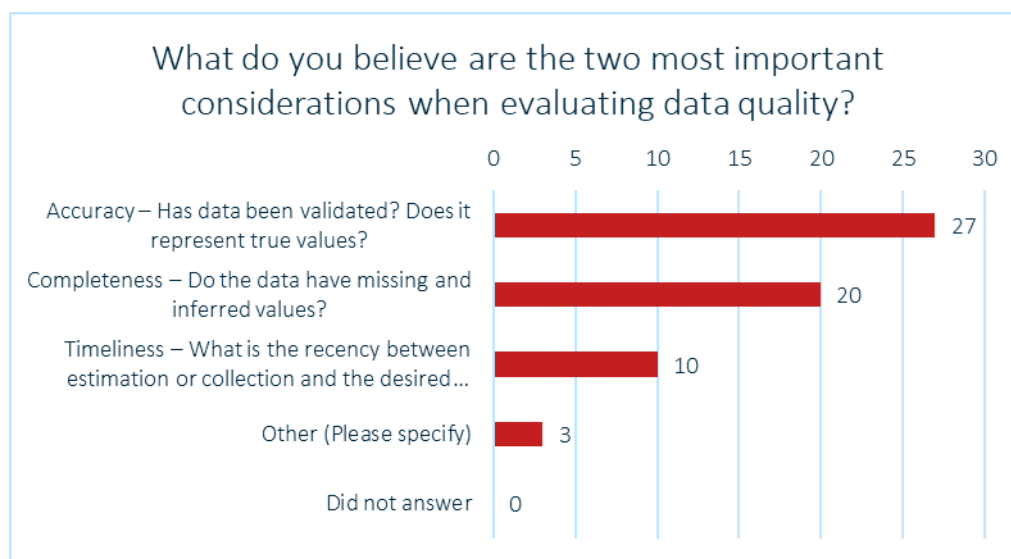


Figure 3.12 External respondents' most important considerations when evaluating data quality

Q13. What, if any, specific challenges or obstacles do you or your organization encounter when attempting to verify data quality (e.g., accuracy, completeness, etc.)? [Open-ended]

Peer agencies reported similar data quality verification challenges as the internal survey respondents. Challenges included:

- Lack of resources (time and staff) for proper quality assurance
- Lack of vendor transparency
- Insufficient documentation and agreement on accuracy metrics, (i.e. what are the thresholds for accuracy?)
- Data merging and resulting format inconsistencies

Q14. What are the two most important data integration considerations for your work? [Select two]

In terms of data integration respondents selected usability as most important, followed by IT effort requirements (**Figure 3.13**). Licensing or sharing was third with privacy coming in fourth. One "Other" response specified integration with other applications as important. Data with higher usability could ensure that the business use cases are met and could allow other staff to handle the data, identifying additional use cases.

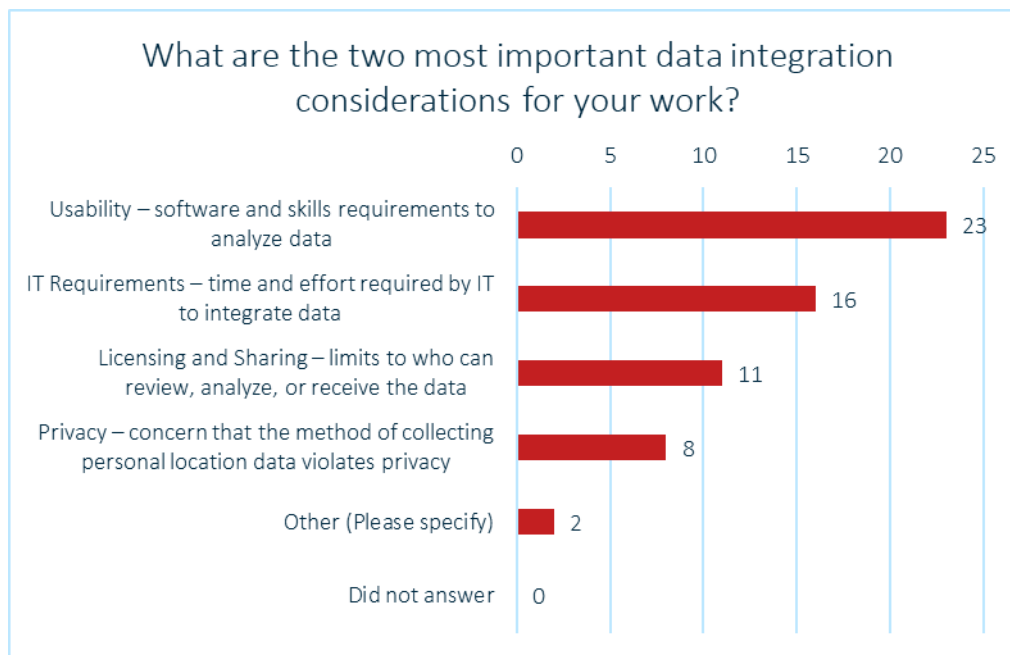


Figure 3.13 External respondents' most important data integration considerations

Third-party data experiences

Q15. What third-party, commercial datasets have you utilized for your work? If none, please enter “None.”

Third-party vendors and datasets used by peer agencies included:

- ATRI
- Drive Wyze
- ESRI
- FUGRO
- Google’s CARTO
- HERE Technologies
- Hexagon
- Info-Tech
- InfoUSA
- INRIX
- Iteris, e.g. ClearGuide
- Lightbox
- Mandli
- Replica
- Regional Integrated
- Transportation Information System
- TRANSCOM, e.g. Spatel
- Strava
- Streetlight
- TMC Transportation Company

- TomTom
- Wejo
- Transmetric America Inc.

Q16. How do you or your organization identify potential use cases for third-party commercial data?

Peer agencies reported a variety of paths that lead to data investment considerations:

- Needs identified by staff, divisions, or departments
- Business units develop use cases
- Learn about vendors at conferences
- Vendors approach with demos
- Gaps identified in methods which then support use cases
- Cross district and division focus groups/brainstorm sessions
- Clear business value with immediate return in value

Q17. What validation or evaluations do you or your organization perform before purchasing third-party commercial data? Please include specific metrics if possible.

When asked about third-party data validation, peer states employ a variety of methods:

- | | |
|--|---|
| • Cost, integration, and shareability | • Business unit decision |
| • Pilot study | • Bid process for identified business use case |
| • Cost/time savings | • Data resolution, acquisition timeframe, and use cases |
| • User feedback | • Quality assessment |
| • Request vendor accuracy metrics then evaluate internally | |

Peer agencies learn about data investment opportunities in similar fashions to ITD. However, some respondents have processes in place to identify needs and business across the agency.

Final comments

Other agencies reported performing similar evaluations around data investments and governance. One respondent recommended “start with the end in mind”, speaking to a focus on the purpose and not get lost in the process.

ITD Townhall and Technical Advisory Committee Meetings

In addition to the two surveys, the project team hosted two virtual meetings: a small “town hall” with ITD staff that have data-related responsibilities and two meetings with the research project Technical Advisory Committee (TAC), which was made up of professionals both internal to and outside of ITD. The goals of these meetings were to discover ways in which ITD and other TAC members’ agencies handle data decisions, discuss successes and challenges, and gather ideas for the most important considerations for the Tool. The takeaways spanned three themes:

ROI Calculation

- ROI calculation requires clear use cases.
- Pilots are an effective first step in identifying use cases.
- Thoroughly documenting needs can help guide decisions and inform specifics on data value (e.g. weight data depends on use; sensors vary with accuracy and cost).

Communication with Leadership

- Communicating easy wins is important – pilots and case studies can offer supportive documentation and materials.
- Tie investments to the agency’s strategic goals.

Data Investment Considerations

- Consider integration and data management.
- Evaluate the trade-off between owning and leasing data.
- Investigate potential for underutilized data that could fit data needs and gaps.

Peer State Interviews

Two state DOTs were selected for follow-up interviews to discuss the specifics of their responses: New Mexico Department of Transportation and Minnesota Department of Transportation.

New Mexico Department of Transportation

New Mexico Department of Transportation (NMDOT) was selected because they indicated using a BCA process when making data investment decisions. The project team met with Asset Management Bureau Chief, Phillip Montoya, for a brief discussion on August 14, 2024. Conversation takeaways include the following:

- Each NMDOT division determines the value of their own data needs and use cases and fills out a data quality plan to ensure data needs are adequately met.
- NMDOT holds cross-division meetings to help identify data integration requirements and ensure that relevant staff have input on decisions.
- It is up to each division to determine the “benefit” of a given data asset.
- The feasibility and relatability of the data are important to keep in mind.
 - Feasibility: Can the data be combined with existing data to get more use out of it?
 - Relatability: Who are the target audiences for the data?
- NMDOT provides data vendors with their own sample data and then requests that they integrate it into their products and submit integration documentation for NMDOT to review.

Minnesota Department of Transportation

The project team had a follow-up discussion with Minnesota DOT (MnDOT) because they reported using data pilots for evaluating data investment decisions. The project team met with Transportation Data & Analysis Program Manager, Benjamin Timerson, for a brief discussion on August 19, 2024. Conversation takeaways include the following:

- MnDOT frequently implements pilots to review data products before investing in them; pilots are driven by business use cases.
- Pre-established forms and tools help the offices evaluate data investment decisions.
 - Executive Briefing Tool (EBT): Helps offices calculate the return on investment.
 - IT Project Request Form: Helps ensure offices collaborate with interdepartmental partners.
- MnDOT utilizes university partners and resources to determine data quality and to apply research methods.
- Employees can use MnDOT’s “Ideascale”, a web platform used to collect research ideas, including data investment considerations.
- MnDOT requires vendors to do a live demonstration based on a business use case.
- MnDOT provided an example of a successful sixth month pilot with StreetLight Data. Highlights included:
 - Vendor provided a live demonstration of the product; other companies being considered fell short for this requirement.

- MnDOT calculated a return on investment of about 35 when compared to a more “traditional” approach that took staff costs into account.
- MnDOT used their EBT to present rationale for contract renewal with StreetLight Data. The document includes details on business impacts, including the benefits of one of the state’s Metropolitan Planning Organizations (Metropolitan Council) use of the platform. It also discusses risks, relevant metrics and data, and clearly states action needed from leadership. MnDOT shared the StreetLight EBT with ITD for reference.

4. Data Asset Evaluation Methodologies

Task Purpose

During this task, the project team synthesized the literature review and stakeholder outreach findings and drew from these to compile a complete list of data asset evaluation methodologies. This list ensured that the project team identified and considered a robust catalogue of approaches. Then, using the list as a reference, the team developed draft versions of the two tools and made recommendations that fit ITD's needs.

The project team recommended a two-form approach: a shorter Word-based form called the Data Purchase/Pilot Abstract Proposal (DPAP) and a longer, Excel-based tool called the Data Investment Needs and Evaluation (DINE). The DPAP evaluates data pilot and small, one-off data purchase requests while the DINE tool helps inform larger, longer-term investments. The next two sections include the complete list of methodologies and descriptions of the recommended approaches that resulted in the two tools.

Comprehensive Methodology List

Following a comprehensive literature review and supplementary research performed through surveys and interviews, the project team identified a wide array of methodologies for evaluating the return on investment for data related investment decisions. The following list outlines each method with a short description. The project team considered implementing one or more of these methods as it developed the draft data procurement or in-house collected data vetting tool.

- **Benefit Cost Analysis (BCA):** a BCA is a calculation that produces a profitability metric, or the benefit compared to the investment cost. The metric comes in the form of a ratio where a value of 1 indicates that the benefit matches the investment, a value below 1 signifies that the benefit is smaller than the investment, and a ratio greater than 1 indicates a benefit that is larger than the cost. For example, a ratio of .9 indicates the benefit is 90% of the cost, while 1.5 means that the benefit is 50% greater than the cost. Performing a BCA provides decision-makers with quantifiable information that enables them to more clearly articulate the value of an investment.
- **Business Use Case:** fitting investment decisions to a clear business use case ensures that they align with existing processes and value-add opportunities. Business use cases are the detailed processes and example projects for which the data would be used. The benefits for each business use case are then evaluated based on how the data investment is expected to impact it.
- **Council Review:** a council of decision makers and technical experts can be assembled to review and select the best data investment option using expert judgement. Whether the council is

standing or ad-hoc, it should include a cross-section of professionals, including but not limited to IT, GIS, legal, and data users and owners.

- **Criteria Evaluation:** criteria are the set of measures used to evaluate a decision. For a data-related decision, criteria can include questions or considerations related to the quality, effort needed to integrate the data into existing systems, vendor reputation, and more. Criteria evaluation requires a method for ranking or weighting the different criteria to determine how decisions compare. Evaluation methods can include:
 - **Binary Evaluation:** for each criterion a binary “yes” or “no” score can be applied, indicating whether a criterion is met and thus providing scores that can inform the vetting or decision-making process.
 - **Judgement Evaluation:** a judgement evaluation or description of whether a data investment option meets a criterion can be attached to each one. This information allows the decision-maker to consider the written comments and employ their own professional judgement to evaluate how well the criterion is met.
 - **Score Evaluation:** a score can be assigned to each criterion to indicate the degree to which the investment would meet it. The score can be a range of categories, for example, “Poor”, “Good”, “Great”, or it could a value out of 5 or 10.
 - **Weighting:** weighting different criteria allows for priorities to be reflected in the evaluation with more important criteria receiving higher weights. Weights are combined with the other evaluations and are multiplied by scores to show cross-criterion priority. The weights can also include dimensions or buckets, or the grouping of criteria by topic, such as quality, integration, and vendor reputation, each with multiple criteria and separate dimension weights.
- **Data Integration:** the ease with which new data can be integrated into existing datasets and/or software applications may impact the labor cost and overall success of actualizing identified benefits. It is essential to consult with data users and IT staff when considering integration. Evaluating integration can be assigned to vendors and included as a method of comparing alternatives.
- **Data Ownership and Maintenance:** planning for who and how the data will be updated and managed contributes to understanding the ease or difficulty of ensuring the long-term value of the data is realized.
- **Data Pilots:** piloting, or receiving and using a data sample, enables agencies to test the data, evaluate quality and integration, and identify additional use cases. Data pilots can help decision-makers with larger scale (i.e., more costly) decisions by clarifying risks and benefits and offering proof of concept.

- **Data Quality:** evaluating the quality of data investment opportunities ensures that the business use cases are met by the investment. Quality may include measuring accuracy, breadth, precision, and other metrics.
- **Decisions Trees:** a decision tree is a process that breaks down decisions into sequential choices, called branches, that the decision-maker follows to a final determination. Decisions trees can lead to decisions, like purchase or not to purchase or whether a pilot or further analysis is needed.
- **Evaluation Form:** forms populated with different ROI decision making methodologies can help guide data related decision-making by ensuring requests include similar and relevant information across options and decisions. The forms can be of varying lengths and detail and may depend on the investment size (e.g., cost). The forms could include evaluations, a BCA, descriptive information about the request and requester, and other information deemed important by the decision makers.
- **Existing Dataset Comparison:** identifying datasets that would be supplanted by a data investment helps prevent investing in duplicate datasets and allows decision-makers to evaluate the benefits that the updated data would provide.
- **Fit Gap Analysis:** a fit gap analysis generally includes identifying a gap or need and then evaluating options based on whether or to what degree the need is addressed and at what price. This analysis can also include comparing a no-investment or status quo alternative.
- **Intangible Benefits:** some investment benefits are not easily valued, such as community trust or enabling a partner agency to use the data. Creating a list of these benefits can help articulate the value of an investment opportunity otherwise not captured in an analytical evaluation.
- **Regulatory Requirements:** legislative authorities (federal, state, agency, etc.) may have data or analysis requirements, which can then be connected to funding. Therefore, decision-makers should consider whether a data investment will ensure or help meet any regulatory requirements.
- **Risk Analysis:** identifying the risks associated with a decision or a specific vendor can help guide further evaluation and increase confidence that the benefit of an investment will be realized.
- **Specificity Evaluation:** data needs can vary across use case. Even if two processes need the same type of data, the degree of specificity may differ. This can include, accuracy, the permitted error rate; granularity, the level of detail; and breadth, the geographic coverage.
- **Standing or Ad-Hoc Committees:** having committees to discuss potential use cases for data can help increase the value capture for investment opportunities. The committee could meet regularly or be assembled when a decision needs to be made.

- **Strategic Alignment:** aligning investment opportunities with the agency's strategic vision or goals can help underscore the importance of a data asset investment for decision-makers and agency leadership.
- **Total Cost:** calculating the cost over the time of an investment provides an easily interpreted metric for decision making. This includes upfront or initiation cost, subscription fees, staff time, storage of any purchased or collected data, and any other related costs incurred.
- **Vendor Evaluation:** data vendors will have different levels of support and training, which can increase the utilization of data investments. A vendor's reputation can be a good indicator of support and the level of integration effort required.

Recommended Approaches

From the methodology list and stakeholder engagement, the project team developed a two-form approach tailored to ITD's needs for evaluating data investments of varying sizes and pilot opportunities. The project team developed these tools to ensure adaptability while aligning with ITD's needs priorities. The first tool provides a concise review for smaller investments and pilot requests, while the second offers a comprehensive evaluation framework for larger investments.

These tools integrate a variety of evaluation methods, including data management, quality, and integration assessments, benefit-cost analysis (BCA), risk assessments, cost assessments, defining intangible benefits, and strategic alignment with ITD objectives. The short-form tool enables ITD to quickly assess smaller opportunities with minimal internal time investment, whereas the long-form tool supports a more thorough analysis for significant financial commitments. Both tools are summarized below.

Data Purchase/Pilot Abstract Proposal – Short Form Tool

When a data manager or user identifies a small data investment (i.e., lower cost) or pilot opportunity, the short form tool, or the Data Purchase/Pilot Abstract Proposal (DPAP), can be completed and included with the investment request. The DPAP provides decision-makers with a preliminary summary of business use cases, strategic alignments, cost, and management considerations. Then, a decision can be made whether to make the investment or pursue the pilot, request a longer evaluation (the DINE Tool), or defer investment. The DPAP is provided for reference in **Appendix B. Data Proposal/Pilot Abstract Proposal (DPAP)**.

When to use the DPAP

A one-time investment is a data purchase that fits a business use case with minimal, if any, additional cost beyond the initial payment. The investment size that warrants this shorter abstract or a longer evaluation is relative and may depend on department or division. For instance, \$60,000 could be a good

threshold, but professional judgement should be used to evaluate whether the request is considered small enough for this short form evaluation or should be subject to a more intensive, long form evaluation.

A data pilot provides ITD an opportunity to access a sample of a vendor's product with a smaller cost. ITD can test how the data could be implemented in existing processes and identify additional business use cases. Following the pilot, which is typically a shorter timeframe of about a year, then ITD can elect whether to enter a contract for a complete product.

The DPAP is a Word document that the requester fills out and submits to leadership. Reaching out to other divisions, offices, and partner agencies can help uncover additional use case opportunities; yet this short form DPAP is designed to be more limited in scope with more in-depth identification methods reserved for the longer DINE Tool. The surveys identified the time cost of filling out data evaluations as a concern, so this short form helps tie the time needed to evaluate data and the size of the investment.

Components of the DPAP

The DPAP has three sections designed to collect the most relevant information for decision making. The sections are outlined below with a version of the DPAP available in **Appendix B. Data Proposal/Pilot Abstract Proposal (DPAP)**:

- I. General Details of the Data Request
- II. Fiscal Requirements
- III. Data Ownership & Management

General details of the data request

The first section of the DPAP collects information about the data request, any related business use cases, and the strategic alignment. Following this section, the decision-maker should have a high-level understanding of the request and its potential benefits.

General detail

The general detail questions are designed to give the decision-maker an overview of the requesting department and the opportunity. It identifies the data, the requester, and a proposal title. This section also includes a question indicating whether the request is a pilot or a small data purchase. Finally, the DPAP asks about the vendor and for a description of the data.

Business use cases

Since clear business use cases help understand value-add potential, it is important to identify as many as possible as early as possible to help capture the full benefit of the investment. The importance of use cases arose multiple times in this project's research, including peer state interviews with NMDOT and

MnDOT, as well as during the TAC meetings. Listing the business use cases provides the decision makers with clear examples of how the data will impact and improve processes. Importantly, a pilot can be an opportunity to identify and evaluate additional business use cases.

A question about regulations is also included as federal and state regulations can influence funding opportunities. This is often a crucial consideration for decision-makers. Therefore, meeting regulatory compliance can drive use cases as new or updated data is required. For instance, the USDOT mandates State DOTs to collect Model Inventory Roadway Element (MIRE) data by 2026, so filling a MIRE data gap could be a critical business use case of a data investment.

Strategic alignment

Aligning data investment decisions with the agency's strategic direction helps communicate the broader value of the opportunity to leadership. This question was flagged as an important challenge to address in the External Survey as well as the Townhall meeting with the TAC. The alignment may reflect the agency's strategic vision, goals, or initiatives set by the governor, department, director, or department. By ensuring investments support larger agency objectives, decision-makers can more easily justify the investment.

Fiscal requirements

The DPAP only considers the costs over five years, limiting the cost to upfront initiation fees and subscriptions. Additional data investment costs can include storage and staff time for training and integration. However, due to the smaller investment associated with the DPAP, these costs are not included and are reserved for the long form.

The DPAP also gives the requester the opportunity to specify when they would like the investment to begin and to suggest the funding source that could be used for the purchase. Collectively, the costs and funding questions can give leadership the fiscal requirements to make small investment and pilot decisions.

Data ownership and management

This section helps decision-makers access how long the benefits of the data could last and who would be accountable for preserving its value. Both the literature review and stakeholder engagement revealed the importance of longer-term care of a data investment as an important consideration. The DPAP includes questions about the data owner and the nature of the investment, such as ongoing payments for access (leasing) or a one-time purchase that results in ownership. Leasing may limit sharing, while purchasing can ensure full control but comes with the responsibility of long-term upkeep. By treating data as a long-term investment, the value can be better sustained through tracking ownership and management responsibilities.

Data Investment Needs and Evaluation – Long Form Tool

The project team designed the Data Investment Needs and Evaluation (DINE) Tool to help ITD make informed decisions about larger data asset investments (e.g., more costly or resource intensive). The DINE is a spreadsheet-based, comprehensive evaluation form that requires collaboration across teams and departments to complete. Similar to the relative definition of a small investment, a “large” data investment can vary depending on the department. Because it is so comprehensive, completing the DINE requires a meaningful time investment. Therefore, the DINE Tool should be reserved for data investments large enough to justify the time required to complete it.

The DINE defines a “data investment” as “purchasing or leasing a dataset from a third-party vendor or allotting internal resources to collecting and maintaining data in-house.” It may often be completed following a data pilot, which will help with identifying and quantifying the benefits.

The DINE Tool is a Microsoft Excel workbook made of 14 tabs or sheets that can generally be divided into four categories. Each tab will be discussed in the sections that follow. A static pdf version of the form is available in **Appendix C. Data Investment and Needs Evaluation Long Form**.

The four main categories of information collected in the DINE Tool are

- I. Instructions and summary
- II. Descriptive information and management
- III. Business use cases and benefit cost analysis
- IV. Quality, integration, and risk evaluations

Instructions and summary

The first three tabs of the DINE Tool provide instructions for the requester (the person completing the form) on the layout and the format of the form. There are also definitions for frequently used terminology as well as a summary sheet which may be saved as or printed to PDF. The summary sheet is among the most important because it collects key information from each of the other sections. Since the instructions and summary tabs are instructive or for review, they are locked to prevent inadvertent editing.

Instructions

The “Instructions” tab provides information on both the purpose and structure of the form. It also specifies that each tab will mention whether it can be completed by the requester or requesting team or if collaboration with other departments, divisions, and ETS staff is recommended. The first tab also informs the requester that the tabs will be locked, excluding editable cells that are highlighted in blue or orange depending on whether they are required or optional. This tab should be read by first time

requesters and reviewed each time to ensure that the DINE Tool is used in its intended instances and completed correctly.

Terminology

The second tab, “Terminology,” provides the definitions for common terms used throughout the evaluation form. Examples of terms defined on this tab include data assessment-related words, such as transparency, recency, and extent, as well as commonly used terms, like data vendor. This tab is designed for the requester to return to as needed. The tab is locked from editing as it is purely informational.

Summary

The “I. Summary” tab appears empty at first, but it is auto populated as the DINE Tool is filled in. This tab is designed to provide decision makers with a high-level overview of the information deemed to be the most important. While the cells are not editable, the size of the rows may need to be adjusted depending on the text length of the referenced cells.

The Summary sheet is designed to be printed to or saved as a PDF. When printed, it creates a two-page document that can be submitted to reviewers, leadership, and decision-makers. If desired, reviewers may conduct a deeper review of the form should they need to further understand the information presented on the Summary sheet. The source tab for each row is referenced so the reviewer can navigate to sections if more information is desired. Some cells also provide ratings and colors provide context to certain numerical scores.

- For the BCA ratio, a BCA ratio below 1, meaning the costs outweigh the benefits, is highlighted in red to indicate that the reviewer should stop and reassess. A BCA ratio of 1 to 2, indicating a benefit equal to or double the investment cost, is highlighted in yellow to encourage the reviewer to take caution and review further. Finally, a BCA ratio greater than 2, meaning the benefits are over twice the cost, is highlighted in green to indicate that the reviewer may proceed with confidence.
- The three evaluation scores, “Data Quality”, “Data Integration”, and “Data Investment Risk,” are each out of 10 with a higher score indicating higher concern or risk. Similar to the BCA ratio, color highlights and context are provided to the number values. A score of 1 through 3 is highlighted in green and indicates “go with confidence,” a score of 3 through 7 is highlighted in yellow and indicates “caution and review,” and a score of 7 through 10 is highlighted in red and indicates “stop and reassess.”

Descriptive information and management

The next section of tabs is where the requester begins to enter information. They ask the requester for information about the data investment opportunity and how the data will be maintained. The content on these tabs aim to provide the necessary context about the investment opportunity.

General Details

The requester populates the “II. General Details” tab with background information related to the data investment proposal. This tab can be completed by the lead requester or the requesting team. From this section, a reviewer should understand the nature of the investment, such as when it would begin, what the investment entails, and who it would serve.

The DINE Tool includes and expands upon the general detail questions in the DPAP by asking about the data investment customers. ITD staff highlighted the importance of tying investments back to the customers, such as the public, Idaho businesses, and local government partners (e.g., MPOs).

Ownership and management

Treating data as an asset is a priority for ITD, so the “III. Ownership & Management” tab ensures that the requester considers who will own, manage, and maintain the data. This tab expands upon the DPAP by asking questions related to storage, maintenance, and data sharing. This tab provides the reviewer an understanding of the parties who will be responsible for the data and how its longevity will be ensured.

Business use cases and benefit cost analysis

This series of sheets includes the identification of business use cases and the evaluation of the return on investment through a BCA. For benefits that would be too time consuming or difficult to quantify, requesters can include them as intangible benefits.

Business use cases

Identifying business use cases is the first step in calculating the benefits of a data investment because they represent the ways data can create new value or improve existing workflows and businesses processes. The “IV. Business Use Cases” tab asks the requester to identify as many use cases as possible with collaboration from other departments or districts. Based on interviews with peer states, the following methods are recommended for requesters to identify business use cases:

- Assembling ad-hoc committees.
- Reach out to other potential ITD data users.
- Distribute a separate form to departments and districts requesting input.
- Consider the use cases of existing ITD data that will be replaced or supplemented by the investment.

The requester is encouraged to provide detailed business use case definitions, including specific data requirements. These specifications stem from discussions during the Technical Committee meetings, where members noted that while data may appear broadly useful across various use cases, specific

needs such as granularity, geographic extent, and recency often vary. By focusing on these specifications, the requester ensures alignment between business needs and data investment attributes.

Benefit cost analysis summary

The “V. BCA Summary” tab is mostly comprised of auto populated fields that reference the BCA cost and benefit tabs. It sums up these input tabs and calculates a BCA ratio, providing a ROI for the investment opportunity. The requester can only edit the cell that defines the number of years for the investment, which then updates the summary tab. The BCA ratio indicates whether the benefits outvalue the costs or vice versa.

Benefit cost analysis: costs

The “V.a BCA Costs” tab includes both one-time and recurring costs that then link back to the BCA summary tab. The one-time costs include initiation fees, integration and training costs, and any other one-time costs, such as equipment purchases. These costs are expected to occur only at or around the time of the investment.

Recurring costs are also included in the BCA calculation. These include charges such as subscription costs, storage costs, internal maintenance costs, and other yearly costs. The storage and maintenance costs were highlighted as important during the stakeholder engagement. The “Other Yearly Costs” are meant to be variable to the specific investment, for example, some vendors may include user fees for each staff member that can access their data.

The requester should use their best judgement on whether a cost is one-time or yearly and ensure costs are not double counted nor undercounted. For example, the instructions specify that if the initiation fee includes the first-year subscription fee, then the requester should remove this amount from the initiation fee as it will be included in the subscription fee.

Benefit cost analysis: process benefits

The DINE Tool includes two separate BCA benefit tabs to account for process improvements and added benefits. The “V.b BCA Process Benefits” tab quantifies the benefits from the amount of time the analyses would have required but is instead saved due to the new data investment. The benefits depend on the business use cases, or how the data is used. For process improvements, benefit quantification is the hours saved multiplied by an hourly rate, a method drawn from a MnDOT data investment evaluation.

This tab includes recommendations for the requester to collaborate with the staff that contributed to identifying the business use cases to also help quantify the time-saving benefits. The tab includes a table that the requester populates with the business use case, a use case title and a description, the hours saved annually, and the hourly rate, from which a formula calculates the total benefit for that use case. An example process benefit is included in the first row for the requester’s consideration.

Benefit cost analysis: added benefits

The second benefit quantification is included in the "V.c BCA Added Benefits" tab. These are the benefits that arise from the data enabling actions that have associated benefits, such as a new safety analysis that can reduce crashes. Therefore, the reduction of crashes would be the added benefit. Other added benefits could include replacing an existing, more expensive data license and meeting a regulatory requirement that is connected to funding. Quantifying additional benefits should also be done by collaborating with the staff members that helped populate the business use cases.

This tab provides a table for the requester to populate and includes an example for reference. The total benefits then link back to the BCA summary tab. After defining the use case, the requester describes the benefit, with instructions to be specific and reference sources. The description should include details or methodology for how the units and unit values were selected. Then, the requester indicates whether this added benefit occurs one-time over the investment period or yearly, ensuring that the benefit is properly scaled. Finally, the quantification is divided into two columns, the "Quantified Unit Change" and the "Per Unit Outcome Value." The unit change is the change caused by integrating the data, such as reducing a crash. The column called "Per Unit Outcome Value" is the dollar value attributed to a single unit change, so referring to the provided example, a distracted driver crash can be valued as a \$231,238 cost.

Intangible benefits

Some business use cases and benefits are more difficult or require too much time to quantify, but they still should be considered in the decision-making process. For instance, if investing in a dataset means that it would also be available to partner agencies, then the benefit may be stronger relationships. The "VI. Intangible Benefits" tab provides a table for the requester to include benefits not quantified in the BCA. These additional, intangible benefits should be derived and developed from identified business use cases.

The table on this tab has example information in three columns to provide the requester with some guidance. The requester titles the intangible benefit, describes it, and identifies the customer that would benefit. Since these benefits do not have value, the Summary tab provides a count of intangible benefits.

Quality, integration, and risk evaluations

The research described in the prior sections highlights a plethora of evaluations beyond the data management and ROI. For instance, the internal and external surveys identified data quality and integration effort as two of the most important considerations. Moreover, the TAC meetings raised a series of risks that could impact a data investment decision. This final section of the DINE Tool includes three tabs of evaluation and aggregation that generate data quality, integration, and risk scores.

Data quality evaluation

The external and internal surveys identified data quality as the most important data evaluation element. A data quality evaluation can be considered a measure of trust. Better quality indicates a closeness to the truth (accuracy), a lack of missing or imputed values (completeness), or data collection or calculation close to the desired time period (recency). A clear and detailed methodology (transparency) can also instill trust. Collectively, when these evaluations are scored, a data quality metric can provide a measure of confidence that the data asset will have the identified efficacy when applied to the business use cases.

Quality can be evaluated with quantitative methods, such as validation by comparing data to known values or evaluating the proportion of missing or imputed values. However, the survey results pointed to the extensive time requirements in evaluating quality quantitatively and the difficulty in determining the truth for known values and identifying benchmarks. Therefore, the “VII. Data Evaluation” tab leverages criteria evaluation, a method mentioned in many of the literature review sources including ITD.

In this tab, the requester considers various statements about the quality of the data at the time of the purchase. The statements cover three categories: accuracy, completeness, and recency. The requester then uses their expertise and best judgement to rate them on a scale of 1 to 3 with a 1 indicating “agree,” a 2 indicating “neutral,” and a 3 indicating “disagree.” The individual scores are then aggregated into a final data quality score out of 10 that links to the Summary tab following discussions with ITD. A score of 1 through 3 (highlighted green) indicates “go with confidence”, of 4 through 6 (highlighted yellow) indicates “caution and review,” and of 7 through 10 (highlighted red) indicates “stop and reassess.”

Data integration evaluation

The survey respondents ranked data integration as the second most important data evaluation metric. Integration accounts for the amount of effort required to incorporate the data into the existing data system, how easily the data can be employed, and any restrictions on sharing or publishing the data. Data integration provides an understanding for the level of effort required to actualize the benefits of the data investment. While these costs are in part included in the BCA, the “VIII. Data Integration” sheet provides a score for evaluating the estimated effort. This tab is best completed collectively with fellow ITD staff and conversations with the data vendor.

The data integration questions cover four categories: integration effort, usability, licensing and sharing, training and support. The integration and usability questions are best completed with input from data users and ETS staff. The licensing and sharing should be evaluated following a discussion with the data vendor to identify potential data sharing restrictions. Finally, the training and support inputs act as a sort of vendor evaluation because the questions consider vendor offerings that can limit or enhance the integration effort, such as demonstrating how the collected data can be integrated into a system and providing staff training.

This tab employs the same evaluation method as the Data Quality Evaluation tab, with a 1 indicating agree, a 2 indicating neutral, and a 3 indicating disagree. The aggregated Integration Score, which is out of 10, links to the Summary tab with the same score classifications as data quality: 1 through 3 indicates "go with confidence", 4 through 6 indicates "caution and review," and of 7 through 10 indicates "stop and reassess."

Risk evaluation

Investing in and utilizing a new data source comes with risks that may arise overtime. The surveys, interviews, and meetings raised additional potential risks to investing in data, such as privacy and security. While risk evaluation was not necessarily the highest priority among the stakeholders, it still warranted evaluation. The final tab, "IX. Risk Evaluation," includes considerations for seven data investment risks and provides space for requesters to add any additional risks that come to mind.

The project team selected risks that were previously identified and not already included in the data quality and integration evaluations. The risk evaluation provides an aggregate risk score out of 10 that is linked to the Summary tab by aggregating the seven individual risk scores, and any added ones, which employ the same 1 through 3 rating of the other evaluation tabs. The risk scores use the same levels as the data quality and integrations scores: 1 through 3 indicates "go with confidence", 4 through 6 indicates "caution and review," and of 7 through 10 indicates "stop and reassess."

5. Test Case Evaluations

Task Purpose

For the final task in the research project, ITD staff completed the DINE Tool to evaluate three ITD data investment opportunities:

- Michelin Driving Data to Intelligence (DDi)
- INRIX Travel Time Data
- Continuous Count Data (produced in house)

These datasets include both third-party data and in-house data collection investments. These test cases enabled the project team to assess the DINE Tool performance across various data investment scenarios and make refinements. The project team presented the findings to the TAC and after some discussion, the committee concluded that ITD had successfully used the DINE Tool to evaluate data investment opportunities.

Vetting Tool Evaluation Summary

ITD had three separate staff members each fill out the DINE Tool based on a dataset their department is considering. ITD provided the completed forms along with comments and findings to the project team. The feedback and corresponding project team response are summarized in **Table 5.1**.

Table 5.1 ITD DINE feedback and project team response

Test Case	DINE Location	Feedback	Vetting Tool Adjustment
All Test Cases	5.a BCA Costs	The formulas for [Storage Cost] and [Internal Maintenance Costs] did not compute automatically.	Added the formulas and locked the cells to prevent accidental deletion.
All Test Cases	4. Business Use Cases through 6. Intangible Benefits	Users expressed uncertainty on column definitions.	Edited the definitions and provided examples to further clarify column definitions.
INRIX	5.a. BCA Costs	Users expressed uncertainty on the definition of “Yearly Recurring Cost” versus “One-Time Cost”	Added a [Length of Investment in Years] field to page 5. Benefit Cost Analysis Ratio.

Test Case	DINE Location	Feedback	Vetting Tool Adjustment
Michelin DDi	2. General Details	User input “Ongoing” as the [Fiscal Year]	Set the formatting for the [Fiscal Year] to “YYYY” and will accept years greater than 2020.
Michelin DDi	1. Summary	The [DINE Form Submission Data] field was not properly formatted.	Set the formatting for the field to a proper date field.
Michelin DDi	5. BCA Summary	The [Investment Fiscal Year] field was not properly formatted.	Set the formatting for the field to “YYYY” and will accept years greater than 2020.

As part of testing the DINE tool, the project team also completed various validations, which are included in **Table 5.2**.

Table 5.2 Additional DINE formatting and formula checks

Check Name	Description of Check
Math Validation	Ensure all items are summed where appropriate
Math Validation	Ensure all items are counted where appropriate
Math Validation	Ensure all items are multiplied where appropriate
Math Validation	Ensure all items are subtracted where appropriate
Number Validation	Ensure number format for dollar amounts
Number Validation	No negative numbers where appropriate
Scoring Validation	Ensure scores in decimals are accounted for in scoring ranges
Text Validation	Ensure text wraps when appropriate

Check Name	Description of Check
Date Validation	Ensure date format when appropriate

Test Case Details

For each of the test cases, the project team reviewed and summarized ITD’s DINE Tool results. The provided files, while not always completed, offered insight into how the results can be interpreted. Ultimately, the tool was deemed successful at evaluating these preliminary test cases.

Michelin Driving Data to Intelligence data

Dataset description

Michelin DDi maintains data from more than 40 million connected drivers in North America. This data provides insights into driving behavior that transportation departments may find beneficial to their planning efforts (MichelinDDi n.d.). ITD, in developing federally mandated Highway Safety Improvement Program (HSIP) projects, turned to Michelin DDi to purchase near miss, harsh acceleration, harsh braking, and other behavioral data to assess roadway risks for mitigation.

Description of request

Based on the information captured in the Summary tab of the DINE Tool, the project team compiled the data’s descriptive information, shown in **Figure 5.1**, into the following statement:

In November of 2024, Margaret Pridmore and Bill Kotowski submitted a request on behalf of the HSIP Administrator and Office of Highway Safety (OHS) for Michelin DDi Behavioral Statistics, for use in a pilot in Kootenai County. The data includes near miss, harsh acceleration, harsh braking, and other behavioral data to assess roadway risks for mitigation. The data can help ITD meet legislative obligations, and it can assist in addressing significant safety needs in ITD Strategic Highway Safety Plan (SHSP) identified focus areas. If approved, ITD would purchase the dataset in fiscal year 2024, and data management would fall under the responsibility of the Roadway Data and OHS.

1	Title of Request (II. General Details)	Michelin DDI Behavioral Statistics - Kootenai County Only - Pilot
2	DINE Form Submission Date (II. General Details)	
3	Requester Name(s) (II. General Details)	Margaret Pridmore and Bill Kotowski
4	Proposed Investment Fiscal Year (II. General Details)	2024
5	Vendor/Provider (II. General Details)	Michelin DDI
6	Data Investment Description (II. General Details)	Purchase of near miss, harsh acceleration, harsh braking, and other behavioral data to assess roadway risks for mitigation.
7	Strategic Alignment (II. General Details)	ITD has a strong focus on Safety, and this would benefit addressing significant safety needs, particularly addressing the ITD SHSP identified focus areas.
8	Legislative Requirement (II. General Details)	It could potentially identify HSIP projects that require a data-driven approach to meet federal obligation requirements.
9	Who are the data customers? (II. General Details)	OHS, KMPO, Dist 1 Traffic Engineering, LEAs, Planners
10	Data Owner (III. Ownership & Management)	Roadway Data and OHS
11	Data Purchase or Lease (III. Ownership & Management)	Owned

Figure 5.1: Description and ownership details from Summary Page (MichelinDDi)

Business Use Cases and Benefit-Cost Analysis Ratio

Based on the information compiled in the Summary tab, the project team wrote up the ROI results, shown in **Figure 5.2**, in the following statement:

The data requester provided nine (9) business use cases. Additionally, the benefit-cost analysis ratio for the data was 14.9, indicating a positive return-on-investment of \$14.90 per \$1 spent. Lastly, the requester provided two (2) additional intangible benefits that could not be readily quantified.

Business use cases are the opportunities where the data could improve processes and provide benefits.		
12	Count of Business Use Cases (IV. Business Use Cases)	9
The return on investment for data decisions are integral to determining whether to invest. The cost must be reasonable and be balanced with the expected return ratio. Yet, some benefits are not easy to quantify, but should still be considered.		
13	Cost (V.a BCA Cost)	\$ 250,000.00
14	BCA Ratio (V. BCA Summary)	14.9
15	Count of Intangible Benefits (VI. Intangible Benefits)	2

Figure 5.2 Use cases and BCA ratio details from Summary Page (MichelinDDi)

Data quality, integration, and risks

Regarding the data's quality, associated integration effort, and potential risks, as shown in **Figure 5.3**, the project team summarized the information into the following statement:

The Data Quality Score was 2.5 out of 10, indicating a low concern with the data's quality. The Data Integration Score was 2.5 out of 10, indicating a low concern with ITD's ability to integrate the data into existing infrastructure. The Data Investment Risk Score was 3.3 out of 10, indicating a low concern with any of the additional risks in purchasing this large dataset.

Considering the data quality, integration effort, and potential risks of the data investment can provide confidence ratings on whether the benefits can be actualized. The following scores provide high-level summaries on each of these topics.		
16	Data Quality Score (VII. Quality Evaluation, a score out of 10 with a higher score indicating quality concerns)	2.5
17	Data Integration Score (VIII. Integration Evaluation, a score out of 10 with a higher score indicating integration concerns)	2.5
18	Data Investment Risk Score (IX. Risk Evaluation, score out of 10 with a higher score indicating additional risk)	3.3

Figure 5.3 Data quality, integration, and risk scores (MichelinDDi)

Evaluation Results

Considering the results of the DINE tool, investing in the Michelin DDI data would provide a high return-on-investment of 14.9, nine additional business use cases, and high data quality, integration, and risk scores, the decision maker can reasonably conclude that the MichelinDDi data could be a good investment.

INRIX Travel Time data

Dataset Description

INRIX collects historic traffic time and traffic speed data from connected vehicles, making them available to transportation planners and consultants who are “focused on improving traffic flow, capacity, and signal timing” (INRIX n.d.). ITD, in its efforts to monitor traffic speed at a granular level, is interested in purchasing traffic speed statistics.

Description of request

Based on the information captured in the Summary tab, the project team compiled the data's descriptive information, shown in **Figure 5.4**, into the following statement:

On November 12, 2024, David Coladner submitted a request INRIX Travel Time and Speed Data. The data, provided by INRIX, provides a view of passively obtained travel speed statistics on very granular sections down to all collector roads in Idaho. The data is not required to meet legislative obligations; however, it can help ITD in its goal to monitor traffic speed performance at a planning level. If approved, ITD would purchase to own the data, and data management would fall under the responsibility of the Data Analytics and Asset Management teams.

The following topics provide descriptive information about the proposed investment and the requester.		
1	Title of Request (II. General Details)	Inrix Travel Time/Speed Data Request
2	DINE Form Submission Date (II. General Details)	11/12/2024
3	Requester Name(s) (II. General Details)	David Coladner
4	Proposed Investment Fiscal Year (II. General Details)	ongoing
5	Vendor/Provider (II. General Details)	Inrix Corporation
6	Data Investment Description (II. General Details)	This data product provides a view of passively obtained travel speed statistics on very granular sections down to all collector roads in Idaho. It is shareable with whomever our agency designates as a partner subject to Data Use Agreement compliance.
7	Strategic Alignment (II. General Details)	Allows the agency to monitor traffic speed performance at a planning level
8	Legislative Requirements (II. General Details)	
9	Who are the data customers? (II. General Details)	Planning staff, Construction Project planning, MPOs, Highway Districts, etc.
10	Data Owner (III. Ownership & Management)	Data Analytics/Asset Management
11	Data Purchase or Lease (III. Ownership & Management)	Owned

Figure 5.4: Description and ownership details from Summary Page (INRIX)

Business Use Cases and Benefit-Cost Analysis Ratio

Based on the information compiled in the Summary tab, the project team captured ROI information, shown in **Figure 5.5**, in the following statement:

The data requester provided four (4) business use cases. Additionally, the benefit-cost analysis ratio for the data was 0.4, indicating a negative return-on-investment. Lastly, the requester provided one (1) intangible benefit, for which a readily accessible benefit was not quantifiable.

Business use cases are the opportunities where the data could improve processes and provide benefits.		
12	Count of Business Use Cases (IV. Business Use Cases)	4
The return on investment for data decisions are integral to determining whether to invest. The cost must be reasonable and be balanced with the expected return ratio. Yet, some benefits are not easy to quantify, but should still be considered.		
13	Cost (V.a BCA Cost)	\$ 600,000.00
14	BCA Ratio (V. BCA Summary)	0.4
15	Count of Intangible Benefits (VI. Intangible Benefits)	1

Figure 5.5: Use cases and BCA ratio details from Summary Page (INRIX)

Data quality, integration, and risks

Regarding the data's quality, associated integration effort, and potential risks, as shown in **Figure 5.6**, the project team summarized this information into the following statement:

The Data Quality Score was three (3) out of 10, indicating a low concern with the data's quality. The Data Integration Score was 2.5 out of 10, indicating a low concern with ITD's ability to integrate the data into existing infrastructure. The Data Investment Risk Score was 2.5 out of 10, indicating a low concern with any of the additional risks associated with investing in large datasets.

Considering the data quality, integration effort, and potential risks of the data investment can provide confidence ratings on whether the benefits can be actualized. The following scores provide high-level summaries on each of these topics.		
16	Data Quality Score (VII. Quality Evaluation, a score out of 10 with a higher score indicating quality concerns)	3.0
17	Data Integration Score (VIII. Integration Evaluation, a score out of 10 with a higher score indicating integration concerns)	2.5
18	Data Investment Risk Score (IX. Risk Evaluation, score out of 10 with a higher score indicating additional risk)	2.3

Figure 5.6: Data quality, integration, and risk scores (INRIX)

Evaluation results

ITD performed a preliminary evaluation of INRIX, so the results are not considered complete, but rather offer an opportunity to gain understanding from the Dine Tool. Therefore, considering the partially completed form, INRIX has a low BCA ratio of 0.4 and a decision maker could reasonably conclude that purchasing the INRIX was not a good investment. However, the DINE Tool provides fields for the requestor to include information which may still justify the investment. The decision maker has the opportunity to review the four business use cases, one intangible benefit, and the high data quality, integration, and risk scores.

Continuous Count data

Dataset Description

Continuous counters are devices that can count and classify vehicle traffic. Data is collected in-house, and maintenance of the devices and data is typically the responsibility of the agency that owns them. ITD, in its efforts to meet federal requirements outlined in MAP-21 and the FAST Act, has decided to purchase continuous counters.

Description of Request

Based on the information captured in the Summary tab, the project team summarized the data's descriptive information, shown in **Figure 5.7**, into the following statement:

On November 5, 2024, Margaret Pridmore submitted a request on behalf of Traffic Operations for Continuous Counters. The devices, provided by multiple vendors and maintained by ITD, provide traffic data which is utilized for roadway design, safety analysis, mobility analysis, and economic impact. The data is required to meet legislative obligations, and it can help ITD by providing a resource for local agencies when applying for grants and other funding. If approved, ITD would purchase the devices in fiscal year 2024, and the data would fall under the responsibility of Roadway Data and Traffic Operations.

1	Title of Request (II. General Details)	ITD Continuous Counters
2	DINE Form Submission Date (II. General Details)	45601
3	Requester Name(s) (II. General Details)	Margaret Pridmore
4	Proposed Investment Fiscal Year (II. General Details)	2024
5	Vendor/Provider (II. General Details)	Multiple Vendors and ITD
6	Data Investment Description (II. General Details)	Maintenance of equipment to collect continuous count data
7	Strategic Alignment (II. General Details)	Traffic data is utilized for roadway design, safety analysis, mobility analysis, and economic impact. It also shows growth across the state and provides a resource for local agencies when applying for grants and other funding.
8	Legislative Requirement (II. General Details)	Yes. It will meet requirements in MAP-21 and FAST Act.
9	Who are the data customers? (II. General Details)	MPOs, Travel Demand Modelers, Planners, Engineers, Businesses, Public, FHWA, neighboring states, and more
10	Data Owner (III. Ownership & Management)	Roadway Data, Traffic Ops
11	Data Purchase or Lease (III. Ownership & Management)	Owned

Figure 5.7: Description and ownership details from Summary Page (Continuous Counters)

Business Use Cases and Benefit-Cost Analysis Ratio

Based on the information compiled in the Summary tab, the project team captured the ROI information, shown in **Figure 5.8**, in the following statement:

The data requester provided five (5) business use cases. Additionally, the benefit-cost analysis ratio for the data was 57, indicating a positive return-on-investment. Lastly, the requester provided three (3) additional intangible benefits, for which a readily accessible benefit was not quantifiable.

Business use cases are the opportunities where the data could improve processes and provide benefits.		
12	Count of Business Use Cases (IV. Business Use Cases)	5
The return on investment for data decisions are integral to determining whether to invest. The cost must be reasonable and be balanced with the expected return ratio. Yet, some benefits are not easy to quantify, but should still be considered.		
13	Cost (V.a BCA Cost)	\$ 3,500,000.00
14	BCA Ratio (V. BCA Summary)	57.0
15	Count of Intangible Benefits (VI. Intangible Benefits)	3

Figure 5.8: Use cases and BCA ratio details from Summary Page (Continuous Counters)

Data quality, integration, and risks

Regarding the data's quality, associated integration effort, and potential risks, as shown in **Figure 5.9**, we can compile this information into the following statement:

The Data Quality Score was one (1) out of 10, indicating a low concern with the data's quality. The Data Integration Score was 1.8 out of 10, indicating a low concern with ITD's ability to integrate the data into existing infrastructure. The Data Investment Risk Score was one (1) out of 10, indicating a low concern with any of the additional risks associated with investing in large datasets.

Considering the data quality, integration effort, and potential risks of the data investment can provide confidence ratings on whether the benefits can be actualized. The following scores provide high-level summaries on each of these topics.		
16	Data Quality Score (VII. Quality Evaluation, a score out of 10 with a higher score indicating quality concerns)	1.0
17	Data Integration Score (VIII. Integration Evaluation, a score out of 10 with a higher score indicating integration concerns)	1.8
18	Data Investment Risk Score (IX. Risk Evaluation, score out of 10 with a higher score indicating additional risk)	1.0

Figure 5.9 Data quality, integration, and risk scores (Continuous Counters)

Evaluation Results

Initially, the continuous count data has a large BCA ratio of 57. While this would justify a purchase, in this instance, it revealed a needed refinement for the tool. Moreover, the count data is needed for ITD

to receive federal funding, which is a large added benefit. In addition to the large BCA ratio, the evaluated data had five (5) additional business use cases and high data quality, integration, and risk scores.

6. Conclusions and Next Steps

Data offerings may continue to grow as technologies advance and adoption increases, such as Artificial Intelligence. This could mean new vendors, existing vendors offering new products, or established providers going out of business. This changing landscape will mean new decisions and opportunities for ITD to consider. The DPAP and DINE Tool can help ITD make objective, consistent decisions that consider use cases, value, quality, integration, and risks.

This project sets the path for ITD to develop a more comprehensive data management plan. Many survey respondents agreed that department-wide data prioritization and governance would address many challenges. As ITD continues to employ the DPAP and DINE Tool, ITD may benefit from assessing its approach to data management and move beyond standardizing new data acquisitions. By adapting the method for valuing data to current datasets, ITD can better understand data as an asset. Moreover, the research and cross-department collaboration encouraged by the DINE Tool can further help ITD staff improve their own understanding, along with that of leadership, of the data used across the agency, creating opportunities for additional efficiencies and reinforcing the view of data as a valuable asset.

7. Cited Works

- AICPA Assurance Services Executive Committee. 2020. *Criteria for describing a set of data and evaluating its integrity*. ICPA & CMA.
- APTA Standards Development Program. 2020. *Procuring Software to Support Transit Asset Management*. American Public Transportation Association.
- California Department of Transportation. 2019. *Data Quality Management Plan*. California Department of Transportation.
- Data.gov Working Group. 2010. *Data.gov Interim Identification & Prioritization Process and Guidelines v1.0*. United States Department of Transportation.
- Federal Highway Administration. 2012. *Benefit-Cost Analysis of Investing in Data Systems and Processes for Data Driven Safety Programs: Decision-Making Guidebook*. . United States Department of Transportation.
- Federal Highway Administration. 2008. *Data Quality White Paper*. United States Department of Transportation.
- Gartner. n.d. *Gartner Magic Quadrant*. Gartner.
- Gladys Kong. 2019. *How To Choose The Right Data Vendor*. Forbes Technology Council.
- Idaho Transportation Department. n.d. *ETS Technology Request Intake Process & Intake Council Overview*. Idaho Transportation Department.
- Idaho Transportation Department. 2022. *ITIP Project Key #22185, Statewide Asset Attribute Inventory*. Idaho Transportation Department.
- INRIX. n.d. *INRIX Analytics: Speed*. <https://inrix.com/products/speed/>.
- Institution of Engineering and Technology. n.d. *Big Data in Transport*. Institution of Engineering and Technology.
- Iowa Department of Transportation. 2020. *Data Management Strategic Plan*. Iowa Department of Transportation.
- Metropolitan Washington Council of Governments. 2021. *Big Data for Regional Travel and Mobility Analyses*. . National Capital Region Transportation Planning Board.
- MichelinDDi. n.d. *Michelin Mobility Intelligence*. <https://mobilityintelligence.michelin.com/us/>.
- National Academies of Sciences, Engineering, and Medicine. 2011. *Improving Employment Data for Transportation Planning*. Washington, DC: The National Academies Press.

National Research Council. 2011. *Allocating Federal Funds for State Programs for English Language Learners*. Washington, DC: The National Academies Press.

Appendix A. Surveys

ITD Internal Data Investment ROI Survey



* Required Information

ITD External Data Investment ROI Survey

The Idaho Transportation Department (ITD) is researching methods for evaluating the return on investment (ROI) of data investments to aid in decision making processes. Data investments could include purchasing data from third-party vendors or allotting resources for internal data collection and management. This survey aims to gather information from peer organizations on their current processes around data investment decision making. It should take less than 10 minutes to complete. Your time and input are greatly appreciated. If desired, an anonymized summary of the survey responses can be provided for your consideration.

Tip: Data investments" refers to the range of different dataset types (pavement condition, land use, traffic, safety, etc.) and collection methods (LiDAR, aerial imaging, sensors, etc.). The dataset types and collection methods continue to grow and we want to account for this evolving space. As is relevant to your responses, please reference any data types or collection methods in the open-ended response questions.

* 1. First & Last Name

* 2. Email

*** 3. Organization**

*** 4. Position**

*** 5. Data-related Responsibilities**

Page 1 of 6

Current Data Asset Procurement Processes

The following questions will focus on how your organization currently makes data investment decisions. This could include purchasing third-party data or collecting data in-house.

Tip: Data investments" refers to the range of different dataset types (pavement condition, land use, traffic, safety, etc.) and collection methods (LiDAR, aerial imaging, sensors, etc.). The dataset types and collection methods continue to grow and we want to account for this evolving space. As is relevant to your responses, please reference any data types or collection methods in the open-ended response questions.

*** 6. What leads you or your organization to consider data investment decisions (e.g., investing in internal data collection, purchasing third-party data, using a public resource, etc.)?**

[Select all that apply]

- ☐ Staff indicate a data need or current data are out of date
- ☐ New state/federal reporting mandate
- ☐ Opportunity for improved efficiency and/or decision-making
- ☐ Learn of a new data provider/vendor
- ☐ Other (Please specify) _____

*** 7. Which of the following approaches do you or your organization employ for addressing data needs?**

[Select all that apply]

- ☐ Purchase third-party data
- ☐ Invest in data collection
- ☐ Identify public resources
- ☐ Varies depending on data need
- ☐ Escalate data need to leadership to address
- ☐ I don't know how we address data needs
- ☐ Other (Please specify) _____

*** 8. How do you or your organization primarily evaluate whether a dataset is important enough to invest in?**

[Select one] (Select one option)

- ☐ Formal benefit/cost analysis
- ☐ Based on business use cases
- ☐ Value is self-evident
- ☐ Professional judgement
- ☐ No formal process
- ☐ I don't know
- ☐ Other (Please specify) _____

9. What specific challenges or limitations do you or your organization encounter when attempting to determine the value of a dataset?

Tip: Data investments" refers to the range of different dataset types (pavement condition, land use, traffic, safety, etc.) and collection methods (LiDAR, aerial imaging, sensors, etc.). The dataset types and collection methods continue to grow and we want to account for this evolving space. As is relevant to your responses, please reference any data types or collection methods in the open-ended response questions.

Page 2 of 6

Making Data Investment Decisions

These questions are targeting processes, important considerations, and challenges in making data investment decisions.

Tip: Data investments" refers to the range of different dataset types (pavement condition, land use, traffic, safety, etc.) and collection methods (LiDAR, aerial imaging, sensors, etc.). The dataset types and collection methods continue to grow and we want to account for this evolving space. As is relevant to your responses, please reference any data types or collection methods in the open-ended response questions.

- * 10. What do you think are the three most important considerations when making data-related decisions (e.g., investing in internal data collection, purchasing third-party data, using a public resource, etc.)?**
[Please select exactly 3 options.]

- ☐ Value Add – ability to satisfy the data need
- ☐ Data Quality – accuracy, precision, lack of missing values, etc.
- ☐ Vendor Quality – methodology transparency, training, client referrals, etc.
- ☐ Integration – usability of data and effort required to integrate and maintain data
- ☐ Staff Resources – availability and knowledge of staff for data collection and management
- ☐ Time Savings – amount of staff time saved by purchasing data versus in-house collection
- ☐ Data Security – whether data will be maintained on agency servers or accessed via the web/cloud
- ☐ Cost
- ☐ Other (Please specify) _____

- 11. What specific challenges or limitations have you or your organization encountered when attempting to determine whether and how to invest in a dataset and, if any, what strategies have your agency successfully used to overcome these challenges/limitations?**

* 12. **What do you believe are the two most important considerations when evaluating data quality?** [Please select exactly 2 options.]

- ☐ Accuracy – Has data been validated? Does it represent true values?
- ☐ Completeness – Do the data have missing and inferred values?
- ☐ Timeliness – What is the recency between estimation or collection and the desired application? How often are data updated?
- ☐ Other (Please specify) _____

13. **What, if any, specific challenges or obstacles do you or your organization encounter when attempting to verify data quality (e.g., accuracy, completeness, etc.)?**

Tip: Data investments" refers to the range of different dataset types (pavement condition, land use, traffic, safety, etc.) and collection methods (LiDAR, aerial imaging, sensors, etc.). The dataset types and collection methods continue to grow, and we want to account for this evolving space. As is relevant to your responses, please reference any data types or collection methods in the open-ended response questions.

* 14. **What are the two most important data integration considerations for your work?** [Please select exactly 2 options.]

- ☐ IT Requirements – time and effort required by IT to integrate data
- ☐ Usability – software and skills requirements to analyze data
- ☐ Licensing and Sharing – limits to who can review, analyze, or receive the data
- ☐ Privacy – concern that the method of collecting personal location data violates privacy
- ☐ Other (Please specify) _____

Page 3 of 6

Third-party Data Experiences

These questions are designed to help us learn about your and your organization's experience(s) using third-party datasets. Third-party data providers are vendors that either sell or provide access to data or collect data on behalf of an agency.

Tip: Data investments" refers to the range of different dataset types (pavement condition, land use, traffic, safety, etc.) and collection methods (LiDAR, aerial imaging, sensors, etc.). The dataset types and collection methods continue to grow, and we want to account for this evolving space. As is relevant to your responses, please reference any data types or collection methods in the open-ended response questions.

- * 15. **What third-party, commercial datasets have you utilized for your work? If none, please enter "None."**

Tip: Data investments" refers to the range of different dataset types (pavement condition, land use, traffic, safety, etc.) and collection methods (LiDAR, aerial imaging, sensors, etc.). The dataset types and collection methods continue to grow, and we want to account for this evolving space. As is relevant to your responses, please reference any data types or collection methods in the open-ended response questions.

16. How do you or your organization identify potential use cases for third-party commercial data?

17. What validation or evaluations do you or your organization perform before purchasing third-party commercial data? Please include specific metrics if possible.

Page 4 of 6

18. Would you be willing to participate in a follow-up interview or meeting on this topic? (Select one option)

- ☐ Yes
☐ No

19. Would you like to receive an anonymized summary of the survey responses? (Select one option)

- ☐ Yes
☐ No

20. Do you have any additional comments or considerations you'd like us to know?

Tip: Data investments" refers to the range of different dataset types (pavement condition, land use, traffic, safety, etc.) and collection methods (LiDAR, aerial imaging, sensors, etc.). The dataset types and collection methods continue to grow, and we want to account for this evolving space. As is relevant to your responses, please reference any data types or collection methods in the open-ended response questions.

Page 5 of 6

Thank you for your time and responses!
If you have any questions, please direct them to the
Project Manager: Rebecca Van Dyke of High Street
Consulting Group
vandyke@highstreetconsulting.com

Page 6 of 6

ITD Internal Data Investment ROI Survey



* Required Information

ITD Internal Data Investment ROI Survey

The Idaho Transportation Department (ITD) is researching methods for evaluating the return on investment (ROI) of data investments to aid in decision making processes. Data investments could include purchasing data from third-party vendors or allotting resources for internal data collection and management. This survey aims to gather information from ITD on current processes around data investment decision making. It should take less than 10 minutes to complete. Your time and input are greatly appreciated.

* 1. First & Last Name

* 2. Email

*** 3. Department**

*** 4. Position**

*** 5. Data-related Responsibilities**

Page 1 of 6

Current Data Asset Procurement Processes

The following questions will focus on how you or your department currently make data investment decisions. This could include purchasing third-party data or collecting data in-house.

*** 6. Do you or your department currently have any of the following unmet data needs?**

[Select all that apply]

- ☐ Data gap needs to be filled (e.g. insufficient variable coverage)
- ☐ Need for more recent data
- ☐ Need for more granular data (e.g. smaller geographies)
- ☐ Need for better quality data
- ☐ I don't know of any data needs
- ☐ Other (Please specify) _____

*** 7. Which of the following approaches do you or your department employ for addressing data needs?**

[Select all that apply]

- ☐ Purchase third-party data
- ☐ Invest in data collection
- ☐ Identify public resources
- ☐ Varies depending on data need
- ☐ Escalate data need to leadership to address
- ☐ I don't know how we address data needs
- ☐ Other (Please specify) _____

*** 8. How do you or your department primarily evaluate whether a dataset is important enough to invest in?**

[Select one] (Select one option)

- ☐ Formal benefit/cost analysis
- ☐ Based on business use cases
- ☐ Value is self-evident
- ☐ Professional judgement
- ☐ No formal process
- ☐ I don't know
- ☐ Other (Please specify) _____

9. What specific challenges or limitations do you or your department encounter when attempting to determine the value of a dataset?

Page 2 of 6

Making Data Investment Decisions

These questions are targeting processes, important considerations, and challenges in making data investment decisions.

- * 10. What do you think are the three most important considerations when making data-related decisions (e.g., investing in internal data collection, purchasing third-party data, using a public resource, etc.)? [Please select exactly 3 options.]

- ☐ Value Add – ability to satisfy the data need
- ☐ Data Quality – accuracy, precision, lack of missing values, etc.
- ☐ Vendor Quality – methodology transparency, training, client referrals, etc.
- ☐ Integration – usability of data and effort required to integrate and maintain data
- ☐ Staff Resources – availability and knowledge of staff for data collection and management
- ☐ Time Savings – amount of staff time saved by purchasing data versus in-house collection
- ☐ Data Security – whether data will be maintained on agency servers or accessed via the web/cloud
- ☐ Cost
- ☐ Other (Please specify) _____

11. What specific challenges or limitations have you or your department encountered when attempting to determine whether and how to invest in a dataset and, if any, what strategies have you or your department successfully used to overcome these challenges and limitations?

12. What do you believe are the two most important considerations when evaluating data quality? [Please select exactly 2 options.]

- ☐ Accuracy – Has data been validated? Does it represent true values?
- ☐ Completeness – Do the data have missing and inferred values?
- ☐ Timeliness – What is the recency between estimation or collection and the desired application? How often are data updated?
- ☐ Other (Please specify) _____

13. What, if any, specific challenges or obstacles have you or your department encountered when attempting to verify data quality (e.g., accuracy, completeness, etc.)?

*** 14. What are the two most important data integration considerations for your work? [Please select exactly 2 options.]**

- ☐ IT Requirements – time and effort required by IT to integrate data
- ☐ Usability – software and skills requirements to analyze data
- ☐ Licensing and Sharing – limits to who can review, analyze, or receive the data
- ☐ Privacy – concern that the method of collecting personal location data violates privacy
- ☐ Other (Please specify) _____

Page 3 of 6

Third-party Data Experiences

These questions are designed to help us learn about your experience(s) using third-party datasets. Third-party data providers are vendors that either sell or provide access to data or collect data on behalf of an agency.

- * 15. **What third-party, commercial datasets have you utilized for your work at ITD? If none, please enter "None."**

- * 16. **How do you or your department identify potential use cases for third-party commercial data?**
[Select all that apply]

- ☐ Staff indicate a data need or current data are out of date
- ☐ New state/federal reporting mandate
- ☐ Opportunity for improved efficiency and/or decision-making
- ☐ Learn of a new data provider/vendor
- ☐ Other (Please specify) _____

17. **What validation or evaluations do you or your department perform before purchasing third-party commercial data? Please include specific metrics if possible.**

18. Would you be willing to participate in a follow-up interview or meeting on this topic? (Select one option)

☐ Yes

☐ No

19. Do you have any additional comments or considerations you'd like us to know?

Page 5 of 6

Thank you for your time and responses!
If you have any questions, please direct them to Project
Manager Rebecca Van Dyke at
vandyke@highstreetconsulting.com

Page 6 of 6

Appendix B. Data Proposal/Pilot Abstract Proposal (DPAP)

DATA PURCHASE/PILOT ABSTRACT PROPOSAL (DPAP)

INSTRUCTIONS: This Data Purchase/Pilot Abstract Proposal (DPAP) should be completed for data vendor pilot or small data procurement requests. Please answer the questions as accurately as possible. This form will be used to understand your request and help determine whether the pilot or small data purchase request should be approved or if more information is needed.

I. General Details of the Data Request

1	Submittal Date:	
2	Requester Name(s):	
3	Requester Email:	
4	ITD Business Area or Department:	
5	Proposal Title:	
6	Is your request for a pilot or small data purchase?	
7	Who is the data vendor/provider?	
8	Describe the data you would like to pilot /purchase and whether ITD has similar or complimentary data already.	
9	Describe the data's business use cases and the related benefits. Include any associated timeframe requirements. [Is there Legislative Authority (Federal, State, MOU, or Agency) that requires the work?]	
10	How does the purchase/pilot contribute to or align with ITD's strategic vision or initiatives?	

II. Fiscal Requirements

		Year 1	Year 2	Year 3	Year 4	Year 5
1	Initiation & Subscription Cost					
2	Desired Starting and/or Implementation Fiscal Year (FYxx)					
3	What funding will be used to complete the effort? [Current FY, Next FY, Uncertain]					

III. Data Ownership & Management

1	Will the data be owned or leased (ITD will lose access once payments stop)?	
2	What department and who, if known, will own and maintain the data?	

Appendix C. Data Investment and Needs Evaluation Long Form

	A	B
1		The Data Investment Needs & Evaluation (DINE) Tool
2		Instructions
3		About the DINE Tool
4		The DINE Tool is designed to help the Idaho Transportation Department (ITD) make informed decisions about data asset investments. A “data investment” includes purchasing/leasing a dataset from a third-party vendor or allotting internal resources to collecting and maintaining data in-house. This form serves as a guide for larger investments. For smaller investment opportunities or to evaluate the benefits of a data pilot, the short form, the Data Pilot/Purchase Abstract Proposal (DPAP), should be used. The data requestor should provide as much detail as possible to help evaluators adequately consider requests. This form is best completed following a pilot, which will help with identifying and quantifying the benefits.
5		Guidance on Completing the DINE Tool
6		The DINE Tool should be primarily filled out and owned by a single requester or small team. Some of the tabs can be completed by the requester(s), but others require input from potential data users and ETS staff. The instructions on each tab include recommendations for the level of collaboration. The DINE Tool should be considered a living document that is completed over time and continuously revisited and reviewed. Completed forms can also help inform additional investment requests, act as a reminder for investment decisions, and a consideration for reinvestment. Therefore, as possible, they should be shared as a resource for other ITD staff.

Do not attempt to edit cells highlighted in grey as they reference other sections or include instructions and examples.

The cells highlighted in blue are for editing. Populate all blue cells, referencing the instructions and examples for guidance.

The cells highlighted in orange indicate list items to be edited, but are not required. Populate as many orange cells as relevant and you can add more rows if necessary.

7	Outline of the DINE Tool
8	The DINE Tool contains the following sections:
9	I. Summary - an "at a glance" view of key information from sections II-IX.
10	II. General Details - basic information about the data request.
11	III. Ownership & Management - information on asset longevity and data management.
12	IV. Business Use Cases - specific examples of how the data may be used.
13	V. Benefit-Cost Analysis (BCA) Summary - the potential return on investment.
14	V.a. BCA Cost - the initial and yearly costs.
15	V.b. BCA Process Benefits - the time-value savings from process improvements.
16	V.c. BCA Added Benefit - the direct and indirect benefits.
17	VI. Intangible Benefits - potential benefits that are not easily quantified.
18	VII. Quality Evaluation - an evaluation of the quality of the data.
19	VIII. Integration Evaluation - an evaluation of the effort required to integrate the data.
20	IX. Risk Evaluation - an evaluation of potential risks for the data investment.
21	

	A	B	C
1		Terminology	
2		The following terms are used in the form.	
3			
4		Term	Definition
5		Accuracy	How closely the data reflects the true values or conditions it represents.
6		Benefit-Cost Analysis (BCA)	A financial evaluation method that compares the total expected benefits of a project to its total expected costs, often presented as a ratio where greater than 1 indicates that benefits outweigh costs, and a ratio less than 1 means costs exceed benefits.
7		Business Use Case	The ITD processes that the data investment will impact or be implemented in.
8		Completeness	Whether all required data points are present, with no missing or incomplete values, also the number of missing or incomplete values that are imputed.
9		Data Investment	Purchasing/leasing a dataset from a third-party vendor or allotting internal resources to collecting and maintaining data in-house.
10		Data Investment Needs & Evaluation (DINE)	This form, the longer data investment form.

11	Data Pilot/Purchase Abstract Proposal (DPAP)	ITD's short data investment request form meant for small data purchases or data pilot requests.
12	Extent	The scope, range, or breadth of the data, typically in terms of geographic coverage or time span.
13	Precision	The level of detail or exactness in the data, indicating how finely measurements or values are specified.
14	Recency	How current the data is, indicating how recently it was collected or updated.
15	Requester	The person filling out the DINE Tool and requesting the data investment.
16	Transparency	Openness and clarity around methodology and underlying data sources.
17	Usability	How easily the data can be accessed, interpreted, and applied for its intended purpose.
18	Vendor/Provider	The organization from which data would be purchased or leased.

	A	B	C	D	E
1	I. Summary This sheet serves as a snapshot of the DINE Form to give decision-makers a high-level summary of relevant details. The information will automatically populate from the other sections. The tab is locked to prevent editing, but you can adjust the height of the rows to view longer text by manually adjusting the row height via the Row Headers on the left or highlighting the desired rows and searching autofit in the Search bar above or from the Home ribbon tab, Cells group, Format drop down, and selecting AutoFit Row Height under Cell Size.				
2					
3					
4	The following topics provide descriptive information about the proposed investment and the requester.				
5	1	Title of Request (II. General Details)			
6	2	DINE Form Submission Date (II. General Details)			
7	3	Requester Name(s) (II. General Details)			
8	4	Proposed Investment Fiscal Year (II. General Details)			
9	5	Vendor/Provider (II. General Details)			
10	6	Data Investment Description (II. General Details)			

11	7	Strategic Alignment (II. General Details)			
12	8	Legislative Requirements (II. General Details)			
13	9	Who are the data customers? (II. General Details)			
14	Data ownership & management are integral to maintaining the value of data assets. The following topics cover key considerations on data ownership.				
15	10	Data Owner (III. Ownership & Management)			
16	11	Data Purchase or Lease (III. Ownership & Management)			
17	Business use cases are the opportunities where the data could improve processes and provide benefits.				
18	12	Count of Business Use Cases (IV. Business Use Cases)	0		

19	The return on investment for data decisions is integral to determining whether to invest. The cost must be reasonable and be balanced with the expected return ratio. A BCA ratio of 2 or above (highlighted green) indicates "go with confidence", a ratio of 1-2 (highlighted yellow) indicates "caution and review," and a ratio of less than 1 (highlighted red) indicates "stop and reassess."		
20	13	Cost (V.a BCA Cost)	\$ -
21	14	BCA Ratio (V. BCA Summary)	
22	15	Count of Intangible Benefits (VI. Intangible Benefits)	0
23	Considering the data quality, integration effort, and potential risks of the data investment can provide confidence ratings on whether the benefits can be actualized. The following scores provide high-level summaries on each of these topics. Scores of 1 through 3 (highlighted green) indicate "go with confidence", scores of 4 through 6 (highlighted yellow) indicate "caution and review," and scores of 7 to 10 (highlighted red) indicate "stop and reassess."		
24	16	Data Quality Score (VII. Quality Evaluation, a score out of 10 with a higher score indicating quality concerns)	
25	17	Data Integration Score (VIII. Integration Evaluation, a score out of 10 with a higher score indicating integration concerns)	
26	18	Data Investment Risk Score (IX. Risk Evaluation, score out of 10 with a higher score indicating additional risk)	
27			

	A	B	C	D
1	II. General Details			
2	Please provide the background information related to the data investment proposal. This page can be completed by the lead requester or the requesting team. From this section, a reviewer should understand the nature of the investment, such as when it would begin, what the investment entails, and who it would serve.			
3				
4		#	Question	Answer
5		1	Title of Request	
6		2	DINE Form Submission Date	
7		3	Requester Name(s)	
8		4	Requester Email	
9		5	Requester ITD Business Area or Department	
10		6	Proposed Investment Fiscal Year (YYYY)	
11		7	Who is the vendor/provider or what is the data collection process?	
12		8	Please describe the data investment request.	
13		9	How does the investment contribute to or align with ITD's strategic vision or initiatives?	
14		10	Will the investment help meet a Legislative Authority's (Federal, State, MOU, or Agency) requirement? If yes, please describe.	
15		11	Who are the customers?	

1

2

3

4

5

6

7

8

9

10

A

B

C

D

E

III. Data Ownership & Management

Data ownership and management are important for maintaining the value of data assets. This page ensures that the requester considers who will own, manage, and maintain the data. This page can be completed by the lead requester or the requesting team, unless they do not include the data owner. A conversation with ETS staff may be helpful.

#	Question	Answer
1	Will the data be owned or leased (i.e., will ITD lose access if it stops paying)?	
2	Will the data be accessed through a vendor/provider platform or stored on ITD servers?	
3	What department and who will own and maintain the data or manage access to a provider/vendor platform?	
4	Will the data require any maintenance (recollected, adjusted, updated, etc.)?	
5	Who can access the data? Is there a cost attributed to each user?	
6	Are there any restrictions for sharing or publishing the data (e.g. can only aggregated data be shared, only paying users can use the data, partners cannot access raw data, etc.)?	

	A	B	C	D	E	F
1	IV. Business Use Cases					
2	Business use cases should be identified by collaborating with other departments or districts to identify potential data users. Business use cases explain how the data may be used. They give decision-makers an understanding of how the data could create new value or improve existing workflows and businesses processes. Identifying clear business use cases helps support the value-add potential, so it is important to identify as many as possible as early as possible. Methods for identifying business use cases include:					
3	- assembling ad-hoc committees. - reaching out to other potential ITD data users. - distributing a separate form to departments and districts requesting input. - considering the use cases of existing ITD data that will be replaced or supplemented by the investment.					
4	For each identified business use case, complete a row in the table. Additional rows can be added if needed. Each business use case should include a benefit and therefore be listed in either the "V.b BCA Process Benefits", "V.c BCA Added Benefits", and/or "VI. Intangible Benefits" tabs.					
5		Business Use Case Title # (e.g. project title)	Business Use Case Description	Does similar data exist? If yes, why do you recommend supplementing or replacing it?	Data Requirement Specifics	
6	0	Title of the project or process that will be impacted by the investment.	Provide a description of the business use case, including how the data will be used and the potential impact of the investment.	Discuss similar datasets or datasets that will be improved by investing in the requested data.	As applicable, specify the data requirements necessary to meet business use case needs. This can include precision, extent (geographical coverage), recency, and accuracy requirements.	
7	e.g.	Safety & Capacity Program Analysis	Used in understanding travel time performance as is. Generally used as a check to help quantify the present-day traffic performance issue. Sometimes used in conjunction with other traffic studies and is a backup source of performance information when there is no other source.	This data is very similar to the NPMRDS dataset that we get from FHWA to support TPM. But this data is on more granular segments and is available on more than just the National Highway System.	Would need to cover SHS and granular sections are desirable to allow for more flexibility to project scope.	
8	1					
9	2					
10	3					
11	4					
12	5					
13	6					
14	7					
15	8					
16	9					
17	10					
18	11					
19	12					
20	13					
21	14					
22	15					
23	16					
24	17					
25	18					
26	19					
27	20					

	A	B	C	D	E	F	G
1	V. Benefit Cost Analysis Ratio						
2	The return on investment for data decisions, here in the form of a BCA ratio, is integral to determining whether to invest. The cost must be reasonable and be balanced with the expected return ratio. This page summarizes the cost, benefits, and Benefic Cost Analysis (BCA) ratio. The information featured in the Costs, Time Benefit, and Added Benefits rows will automatically populate based on your responses in tabs "IV.a BCA Costs," "IV.b BCA Process Benefits," and "IV.c. BCA Added Benefits." A BCA ratio of 2 or above (highlighted green) indicates "go with confidence", a ratio of 1-2 (highlighted yellow) indicates "caution and review," and a score of less than 1 (highlighted red) indicates "stop and reassess."						
3	The Investment Fiscal Year is populated by the "II. General Details" tab; if the Fiscal Year in the "II. General Details" is blank, then the current year is used. The requester can enter a value of 1 through 5 for the Length of Investment in Years in the blue box.						
4	Investment Fiscal Year (linked to 'II. General Details' Investment Fiscal Year)		2024				
5	Discount Rate		7%				
6	Length of Investment in Years, Max of 5		5				
7							
8		2024	2025	2026	2027	2028	
9	Costs	\$ -	\$ -	\$ -	\$ -	\$ -	
10	Process Benefits	\$ -	\$ -	\$ -	\$ -	\$ -	
11	Added Benefits	\$ -	\$ -	\$ -	\$ -	\$ -	
12							
13	Total Cost and the Benefit Needed to Reach BCA Ratio of 1		\$ -				

	A	B	C	D
1	V.(a) Benefit Cost Analysis: Costs This page describes costs associated with the data investment. Both one-time and recurring costs are included and link back to the BCA summary tab. The one-time costs include initiation fees, integration and training costs, and any other one-time costs, such as equipment purchases. These costs are expected to occur only at or around the time of the investment. Recurring costs include charges such as subscription costs, storage costs, internal maintenance costs, and other yearly costs. Conversations with ETS staff and the vendor or process experts can help identify costs. Please fill in the blue cells and be as accurate as possible. The One-Time Investment Costs will only populate investment year. The Yearly Recurring Costs will be applied to each year of the investment, so make sure the Yearly Investment Costs are not double counted. For instance, if the initiation fee includes the first year subscription fee, remove this amount from the initiation fee and only include the subscription fee. Complete the blue cells and the grey cells will automatically update.			
2				
3				
4	One-Time Investment Costs			
5	Initiation Fee	\$	-	
6	Integration Cost	\$	-	
7	Estimated Labor Hours Required for Integration			
8	Hourly Integration Labor Rate	\$	-	
9	Training Cost	\$	-	
10	Estimated Hours Required for Training			
11	Hourly Training Cost	\$	-	

12	Other One-Time Costs (e.g. equipment purchases, etc.)	\$	-
13	Total One-Time Investment Costs	\$	-
14			
15	Yearly Recurring Costs		
16	Subscription Cost	\$	-
17	Storage Cost	\$	-
18	Data Size in Gigabytes (GB)		
19	Yearly GB Cost	\$	-
20	Internal Maintenance Cost	\$	-
21	Estimated Yearly Labor Maintenance Hours		
22	Hourly Maintenance Labor Rate	\$	-
23	Other Yearly Costs (e.g. equipment rentals, user fees, etc.)	\$	-
24	Total Recurring Investment Costs	\$	-
25			
26	Total Costs Across Investment Period	\$	-

	A	B	C	D	E	F	G
1	V.(b) Benefit Cost Analysis: Process Benefits						
2	This page quantifies the benefits from the amount of time the analyses would have required but is instead saved due to the new data investment. The benefits depend on the business use cases, or how the data is used. Quantifying the time saving benefits from employing the data investment in business use case processes should be done by collaborating with the parties that have identified the use cases. The table below provides the outline for calculating benefit considering business use cases, projects, and project costs. If you can identify a potential benefit but are unable to quantify it at this time, please include it in tab VI. Intangible Benefits.						
3							
4	#	Business Use Case	Description	Hours Saved Annually	Hourly Rate	Total Benefit	
5	0	Title of the business use case that the data would impact.	Describe how the data would improve the business use case process. Please be descriptive.	Approximate, number of hours saved annually on the business use case process due to purchasing the data. Could be the number of times a project or analysis is run multiplied by the time spent to perform.	The approximate average hourly rate for the time spent on the process.	Hours Saved Annually multiplied by the Hourly Rate. The cells have formulas that will calculate the total benefit.	
6	e.g.	Safety & Capacity Program Analysis	Varies from project to project. But saving us from doing a single traffic study would be valuable. For this example, coarsely speaking, let's say we save 5 traffic speed studies.	500	\$ 100	\$ 50,000	

7	1				\$ -	\$ -
8	2				\$ -	\$ -
9	3				\$ -	\$ -
10	4				\$ -	\$ -
11	5				\$ -	\$ -
12	6				\$ -	\$ -
13	7				\$ -	\$ -
14	8				\$ -	\$ -
15	9				\$ -	\$ -
16	10				\$ -	\$ -
17	11				\$ -	\$ -
18	12				\$ -	\$ -
19	13				\$ -	\$ -
20	14				\$ -	\$ -
21	15				\$ -	\$ -
22	16				\$ -	\$ -
23	17				\$ -	\$ -
24	18				\$ -	\$ -
25	19				\$ -	\$ -
26	20				\$ -	\$ -

	A	B	C	D	E	F	G	H
1	V.(c) Benefit Cost Analysis: Additional Quantifiable Benefits							
2	Benefits of a data investment may also result in other measurable outcomes. This page includes the benefits that arise from the data enabling actions that have associated benefits, such as a new safety analysis that can reduce crashes. Therefore, the reduction of crashes would be the added benefit. Other added benefits could include replacing an existing, more expensive data license and meeting a regulatory requirement that is connected to funding. Quantifying additional benefits from employing the data investment into business use case processes should be done by collaborating with the parties that have identified the use cases.							
3								
4		#	Business Use Case	Description	Yearly vs. One-Time	Quantified Unit Change	Per Unit Outcome Value	Total Yearly Benefit
5		0	Title of the business use case that the data would impact.	Describe the readily quantifiable benefit if the data were employed in the business use case, including the unit of improvement. Please be descriptive by adding information sources and explaining any unit changes and unit values.	Select "One-time" or "Yearly" to indicate whether the benefit occurs once or each year of the investment window.	Quantify the yearly or one-time change from integrating the data. Do not scale to investment window.	The dollar value of the unit change.	Yearly Total, calculated with a formula taking into account unit change and per unit value. One-time benefits are converted to yearly values.
6		e.g.	Distracted Driving	We estimate \$1.1 Billion is spent on distracted driving each year for an estimate 4,757 distracted driving crashes in 2023. This comes out to \$231,238 spent per distracted driving crash in 2023. Assuming we can identify the areas of distracted driving, we can target messaging and law enforcement agency efforts. The assumption would be that we can reduce at least one distracted driving crash per year.	Yearly	1	\$ 231,238	\$ 231,238
7		1					\$ -	\$ -
8		2					\$ -	\$ -

9	3					\$ -	\$ -
10	4					\$ -	\$ -
11	5					\$ -	\$ -
12	6					\$ -	\$ -
13	7					\$ -	\$ -
14	8					\$ -	\$ -
15	9					\$ -	\$ -
16	10					\$ -	\$ -
17	11					\$ -	\$ -
18	12					\$ -	\$ -
19	13					\$ -	\$ -
20	14					\$ -	\$ -
21	15					\$ -	\$ -
22	16					\$ -	\$ -
23	17					\$ -	\$ -
24	18					\$ -	\$ -
25	19					\$ -	\$ -
26	20					\$ -	\$ -

	A	B	C	D	E
1	VI. Intangible Benefits				
2	Some business use cases and benefits are more difficult or require too much time to quantify, but they still should be considered in decision-making. For instance, if investing in a given dataset means that it would also be available to partner agencies, then the benefit may be stronger relationships with these partners. These additional, intangible benefits should be identified by collaborating with the parties that have identified the use cases. Populate a row in the table below for each intangible benefit.				
3					
4		#	Intangible Benefit	Intangible Benefit Description	Customers
5		0	Include the title of the potential benefit.	Provide a description of how the investment could result in the identified benefit.	Who are the customers that will benefit? The public, partner agencies, specific communities, etc.
6		e.g.	Business Location	Businesses moving to Idaho will typically look at traffic counts to verify that there is enough traffic to keep the business profitable.	Businesses
7		1			
8		2			
9		3			
10		4			
11		5			
12		6			
13		7			
14		8			
15		9			

16	10			
17	11			
18	12			
19	13			
20	14			
21	15			
22	16			
23	17			
24	18			
25	19			
26	20			

	A	B	C	D	E	F
1	VII. Data Quality Evaluation The following questions can be done by the requestor(s) following the business use case identification and quantification, but some questions may be more easily answered through collaboration. A data quality evaluation can be considered a measure of trust. Better quality indicates that the results from data analysis can be trusted, whether that is due to the closeness to the truth (accuracy), the lack of missing or imputed values (completeness), or the recency of the data collection or creation (recency). Sometimes to consider a data's quality, a sample and a description of the methodology is required for review (transparency). Collectively, when these evaluations are scored, a data quality metric can provide a measure of confidence that the data asset will have the identified efficacy when applied to the business use cases. Consider the statements below about the quality of the data at the time of the purchase, and use the drop-down menu, indicate on a scale of 1 to 3 whether you agree. 1 = agree; 2 = neutral; 3 = disagree. Please respond to each statement. The scores are then aggregated into a Final Quality Score. A score of 1 through 3 (highlighted green) indicates "go with confidence", of 4 through 6 (highlighted yellow) indicates "caution and review," and of 7 to 10 (highlighted red) indicates "stop and reassess."					
2						
3						
4	Final Quality Score			out of 10, where 10 is high concern about data quality.		
5						
6	Score (1 agree to 3 (disagree)					
7	#	Statements		Comments & Explanation		
8	1	Accuracy				
	1.A	I have few to no concerns about the quality of the underlying data sources(s) and/or collection methods.				

9	1.B	The data methodology is transparent (clearly articulated) and I feel confidence in the accuracy of the data.		
10	1.C	There are whitepapers and/or benchmark reports that help demonstrate the accuracy of the methodology & results.		
11	2 Completeness			
12	2.A	The amount of data imputed or inferred values due to missing data is clearly articulated and I feel confident in the results.		
13	2.B	I feel confident that the amount of missing data that is imputed or inferred does not harm the quality of the data.		
14	2.C	The data covers the necessary extent, or geographical coverage, required for the business use cases.		
15	3 Recency			

16	3.A	Based on the most recent date of data collection or calculation, I feel confident that it meets the recency or timeliness requirements of the business use cases.		
17	3.B	Based on how often the data is updated, I feel confident that it will continue to meet the recency or timeliness requirements of the business use cases.		
18	3.C	I feel confident that as the data is updated, it will remain compatible with older data, allowing for analyses across time.		

	A	B	C	D	E
1	VIII. Data Integration Evaluation				
2	The integration evaluation questions should be done with input from data users and IT staff. Integration accounts for the amount of effort required to incorporate the data into the exiting data system, how easily the data can be employed, and any restrictions on sharing or publishing the data. Considering data integration is key to understanding the level of effort required to actualize the benefits of the data investment. While these costs are in part included in the BCA, this evaluation can provide a confidence score on the estimated effort.				
3	Consider the statements below and using the drop-down menu, indicate on a scale of 1 to 3 whether you agree. 1 = agree; 2 = neutral; 3 = disagree. Please respond to each statement. The scores are then aggregated into a Final Integration Score. A score of 1 through 3 (highlighted green) indicates "go with confidence", of 4 through 6 (highlighted yellow) indicates "caution and review," and of 7 to 10 (highlighted red) indicates "stop and reassess."				
4	Final Integration Score			out of 10, where 10 is high concern about data integration.	
5					
6	Score (1 agree to 3 (disagree)				
7	#	Statements		Comments & Explanation	
8	1	Integration Effort			
9	1.A	The data can be integrated into the business use cases with little difficulty.			
	1.B	Integrating data does not require repeated actions by different user, such as reformatting every time the data is accessed through a portal.			

10	2 Usability		
11	2.A	Staff require limited training to utilize the data.	
12	2.B	The data is easily accessed by different users.	
13	2.C	The data format is clearly labeled and documented.	
14	2.D	ITD has the necessary skillsets and software to utilize the data.	
15	3 Licensing & Sharing		
16	3.A	ITD can share the data in ways that meets any and all regulatory requirements.	
17	3.B	ITD can share the data in ways with partners or public that will enable relationship cultivation and strengthening.	
18	3.C	All ITD staff that wants to access and employ the data is able to.	

19	4 Training & Support		
20	4.A	The data vendor provides adequate training or training is available on how to use the collected data.	
21	4.B	The data vendor demonstrates how their data can be integrated into business use cases, even helping with any required data manipulation; or it is clear how the collected data can be integrated into the use cases.	
22	4.C	There is ongoing support by the provider or process experts to help with questions, challenges, and/or new users.	

	A	B	C	D	E	F												
1	IX. Risk Evaluation The risks identified below can be completed by the requester(s) following the business use case identification and quantification, but some questions may be more easily answered through collaboration. Investing in and utilizing a new data source comes with risks that may arise overtime and they should be evaluated. The table below includes risks that have been identified in advance, as well as additional space for risks to be added. Certain risks were incorporated into other areas of the DINE Form, such as data quality and integration risks. Try to avoid adding additional risks that are incorporated elsewhere to avoid double counting. Consider the statements below and using the drop-down menu, indicate on a scale of 1 to 3 whether you agree. 1 = agree; 2 = neutral; 3 = disagree. Please respond to each statement. The scores are then aggregated into a final Risk Score. A score of 1 through 3 (highlighted green) indicates "go with confidence", of 4 through 6 (highlighted yellow) indicates "caution and review," and of 7 to 10 (highlighted red) indicates "stop and reassess."																	
2																		
3																		
4	<table border="1"> <tr> <td>Risk Score</td> <td></td> <td>out of 10, where 10 is high concern about data integration.</td> </tr> </table>		Risk Score		out of 10, where 10 is high concern about data integration.													
Risk Score		out of 10, where 10 is high concern about data integration.																
5	<table border="1"> <thead> <tr> <th>#</th> <th>Risk Statements</th> <th>Score (1 agree to 3 disagree)</th> <th>Comments & Explanation</th> </tr> </thead> <tbody> <tr> <td>1</td> <td>There are no privacy concerns that would violate any regulations or harm trust with the public or partners.</td> <td></td> <td></td> </tr> <tr> <td>2</td> <td>I feel confident that the vendor will remain in business during the entirety of the contract or the process will be repeatable as needed throughout the 5-year investment period.</td> <td></td> <td></td> </tr> </tbody> </table>						#	Risk Statements	Score (1 agree to 3 disagree)	Comments & Explanation	1	There are no privacy concerns that would violate any regulations or harm trust with the public or partners.			2	I feel confident that the vendor will remain in business during the entirety of the contract or the process will be repeatable as needed throughout the 5-year investment period.		
#	Risk Statements	Score (1 agree to 3 disagree)	Comments & Explanation															
1	There are no privacy concerns that would violate any regulations or harm trust with the public or partners.																	
2	I feel confident that the vendor will remain in business during the entirety of the contract or the process will be repeatable as needed throughout the 5-year investment period.																	
6																		
7																		

8	3	I feel confident that the vendor will not change the methodology or formatting of the data during the entirety of the contract or the collection process will not change.		
9	4	I feel confident that the integration or use of the data investment will not compromise DOT system security , such as increasing risk of malware or data breaches, especially for personal identifiable information.		
10	5	I feel confident that the processes or methodologies involved in producing or leveraging the data are not at risk of regulatory or policy changes that could inhibit its use. This could include the use of novel technologies, like AI or cloud storage or computing.		
11	6	I feel confident that the vendor or collection process does not have unexpected fees that could cause the estimated costs to be greater. This could include vendor fees to run reports.		

12	7	I feel confident that the vendor or data collection expert will be responsive and communicative in answering questions and open about any changes in the data or process.		
13	If you can identify other risks to the data investment, please include them below.			
14	Score 1 (agree to 3 (disagree)			
15	#	Risk	Score	Comments & Explanation
16	8			
17	9			
18	10			
19	11			
20	12			