



Pennsylvania
Department of Transportation

**Three-Dimensional Modeling and
Uncertainty Quantification of
Subsurface Soils**

FINAL REPORT

May 15, 2026

By: Allen Cadden, Michael Prohaska, Ross Cutts,
Jeremy Brown, Ahmad Alhasan
Applied Research Associates, Inc.
Schnabel Engineering, LLC

COMMONWEALTH OF PENNSYLVANIA
DEPARTMENT OF TRANSPORTATION

CONTRACT # 4300805489
WORK ORDER/RFQ # 221216

1. Report No. FHWA-PA-2026-RFQ221216		2. Government Accession No.		3. Recipient's Catalog No.	
4. Title and Subtitle Three-Dimensional Modeling & Uncertainty of Subsurface Soils				5. Report Date May 2026	
				6. Performing Organization Code	
7. Author(s) Allen Cadden, Michael Prohaska, Ross Cutts; Jeremy Brown; Ahmad Alhasan				8. Performing Organization Report No.	
9. Performing Organization Name and Address Applied Research Associates, Inc. (ARA) 100 Trade Centre Dr., Suite 200 Champaign, IL 61820 Schnabel Engineering, LLC (Schnabel) 3 Dickinson Dr., Suite 200 Chadds Ford, PA 19317				10. Work Unit No. (TRAIS)	
				11. Contract or Grant No. 4300805489	
12. Sponsoring Agency Name and Address The Pennsylvania Department of Transportation Bureau of Planning and Research Commonwealth Keystone Building 400 North Street, 6 th Floor Harrisburg, PA 17120-0064				13. Type of Report and Period Covered Final Report June 2024 – June 2026	
				14. Sponsoring Agency Code	
15. Supplementary Notes					
16. Abstract This report documents the development of a framework for a unified, web-based Geotechnical Data Management System (GDMS) to modernize the management of Pennsylvania's subsurface geotechnical data. Historically, boring logs, laboratory results, field observations, and construction records have been stored in fragmented, district-level formats with no centralized statewide system, resulting in duplicated investigations, inefficient planning, and lost institutional knowledge. The GDMS framework was developed through a literature review of peer agency best practices, a system architecture and implementation plan, and a functional pilot system demonstration. The pilot system was built on an open-source software stack with DIGGS-compliant data structures and validates the technical feasibility of centralizing geotechnical data with interactive 2D/3D visualization, automated quality assurance, and integration with existing enterprise systems. A four-stage expansion roadmap guides the transition from the current research pilot to a statewide production system, ultimately supporting predictive analytics, geotechnical asset management, and digital twin capabilities. Sustainable adoption requires formalized data governance, QA/QC protocols, stakeholder engagement across more than twenty-five identified internal and external groups, and targeted training programs. The GDMS is positioned to serve as the central platform for geotechnical data stewardship across the infrastructure lifecycle.					
17. Key Words Geotechnical Engineering, Subsurface Investigation, Geotechnical Data Management, Three-Dimensional Modeling, Data Integration, Data Standards, Geographic Information Systems, Transportation Asset Management, Decision Support Systems, Infrastructure Data Governance				18. Distribution Statement No restrictions. This document is available from the National Technical Information Service, Springfield, VA 22161	
19. Security Classif. (of this report) Unclassified		20. Security Classif. (of this page) Unclassified		21. No. of Pages 152	22. Price

TABLE OF CONTENTS

LIST OF FIGURES	II
LIST OF TABLES	III
ABBREVIATIONS	IV
DEFINITIONS	VIII
EXECUTIVE SUMMARY	1
1. INTRODUCTION AND TASKS	3
1.1 BACKGROUND	3
1.2 GOALS	3
1.3 TASKS	4
2. TASK 1. LITERATURE REVIEW	6
2.1 TASK INTRODUCTION	6
2.2 LITERATURE SUMMARY	6
2.3 AGENCY INTERVIEWS	7
2.4 DATA INTEGRATION, STANDARDIZATION, AND GOVERNANCE	10
2.5 DATA INTERPRETATION AND UNCERTAINTY QUANTIFICATION	11
2.6 TASK CONCLUSIONS	14
3. TASK 2. IMPLEMENTATION PLAN	16
3.1 TASK INTRODUCTION	16
3.2 SYSTEM ARCHITECTURE	16
3.3 EXTERNAL DATA SOURCES AND INTEGRATION	18
3.4 DATA STORAGE AND MANAGEMENT	21
3.5 SYSTEM ADMINISTRATION AND MAINTENANCE	22
3.6 IMPLEMENTATION PLAN	24
3.7 TASK CONCLUSION	26
4. TASK 3. DATA INTEGRATION AND DEVELOPMENT OF PILOT 3D MODEL	27
4.1 TASK INTRODUCTION	27
4.2 SYSTEM ARCHITECTURE	27
4.3 PILOT DATA MODEL	27
4.4 FRONT-END TECHNOLOGIES	28
4.5 KEY FEATURES OF THE PILOT	28
4.6 INTERACTION WITH THE GDMS	28
4.7 TASK CONCLUSIONS	37
5. TASK 4. DEVELOP RECOMMENDATIONS	38
5.1 TASK INTRODUCTION	38
5.2 ESTABLISHING GDMS VALUE	38
5.3 PILOT RESPONSE AND EXPANSION	40
5.4 STAGED EXPANSION	41
5.5 DATA COLLECTION AND INTEGRATION	42
5.6 GDMS TOOLS AND UTILIZATION	46
5.7 TASK CONCLUSIONS	46
6. SUMMARY OF PROJECT CONCLUSIONS AND RECOMMENDATIONS	47
6.1 CONCLUSIONS	47
6.2 RECOMMENDATIONS	47
APPENDICES	54
APPENDIX A. INTERIM REPORT 1	55
APPENDIX B. REVISED INTERIM REPORT 4	56

LIST OF FIGURES

Figure 1: Proposed system architecture.	16
Figure 2: Visualization of GDMS data storage and management systems.....	18
Figure 3: Visualization of external source integration.....	19
Figure 4: Visualization of data ingestion framework.	20
Figure 5: Visualization of components, features, and visualization tools.	22
Figure 6: Proposed implementation schedule.	25
Figure 7: Opening screen as viewed from a web browser.	29
Figure 8: Borehole selection tool example for a single borehole.	30
Figure 9: Borehole selection tool example for a desired region.	30
Figure 10: Visualization tool for changes in features with depth.	31
Figure 11: Project-selection tool.....	31
Figure 12: Selected project summary page and borehole map.	32
Figure 13: Query page with common queries pre-established.....	33
Figure 14: Subsurface statistics options including SPT-N values vs. elevation and subsurface variability vs. elevation.	34
Figure 15: Geotechnical uncertainty visualization tool using a “light and shadow” presentation method.....	35
Figure 16: Karst site remediation demonstration providing a satellite overlay of the project site and borehole locations.	35
Figure 17: Karst site remediation demonstration showing records associated with each location ID (e.g., RB-03, RB-04) and object type (boring).	36
Figure 18: Karst site remediation demonstration showing a heat map of geophysical results along with boring locations.....	37
Figure 19: Karst site remediation demonstration showing a second visualization of grout volume at each boring with depth.	37
Figure 20: Example influence-commitment matrix for GDMS adoption.....	40
Figure 21: New record integration workflow.	43
Figure 22: Historical boring record integration workflow.....	44
Figure 23: Historical construction record integration workflow.	45

LIST OF TABLES

Table 1. Example internal stakeholder registry for GDMS development.	39
---	----

ABBREVIATIONS

Abbreviation	Full Name
AASHTO	American Association of State Highway and Transportation Officials
AGS	Association of Geotechnical and Geoenvironmental Specialists
AI	Artificial Intelligence
AKDOT	Alaska Department of Transportation and Public Facilities
API	Application Programming Interface
AR	Augmented Reality
ASCII	American Standard Code for Information Interchange
ASCE	American Society of Civil Engineers
ARA	Applied Research Associates, Inc.
ASCE	American Society of Civil Engineers
ASTM	American Society for Testing and Materials (now ASTM International)
BIM	Building Information Modeling
BMS	Bridge Management System
CAD	Computer-Aided Design
Caltrans	California Department of Transportation
CCDF	Conditional Cumulative Distribution Function
CGD	Canterbury Geotechnical Database
COGO	Coordinate Geometry
CRS	Coordinate Reference System
CSS	Cascading Style Sheets
CSV	Comma-Separated Values
DCNR	Pennsylvania Department of Conservation and Natural Resources
DEP	Pennsylvania Department of Environmental Protection
DGE	District Geotechnical Engineer
DIGGS	Data Interchange for Geotechnical and Geoenvironmental Specialists
DNR	Department of Natural Resources
DOM	Document Object Model
DOT	Department of Transportation
DSM	Deep Soil Mixing
DWG	Drawing (AutoCAD file format)
DXF	Drawing Exchange Format

ECMS	Engineering and Construction Management System
EDA	Exploratory Data Analysis
EPA	Environmental Protection Agency (U.S.)
EQuIS	Environmental Quality Information System
ESRI	Environmental Systems Research Institute
FEMA	Federal Emergency Management Agency
FHWA	Federal Highway Administration
GAM	Geotechnical Asset Management
GDMS	Geotechnical Data Management System
GeoJSON	Geographic JavaScript Object Notation
GIS	Geographic Information System
GiST	Generalized Search Tree (PostgreSQL index type)
GML	Geography Markup Language
GNSS	Global Navigation Satellite System
GPS	Global Positioning System
GSLIB	Geostatistical Software Library
HTML	HyperText Markup Language
IDW	Inverse Distance Weighted (interpolation method)
IFC	Industry Foundation Classes
InSAR	Interferometric Synthetic Aperture Radar
ISSMGE	International Society for Soil Mechanics and Geotechnical Engineering
IT	Information Technology
JSON	JavaScript Object Notation
LaDOTD	Louisiana Department of Transportation and Development
LiDAR	Light Detection and Ranging
LMS	Learning Management System
MAP-21	Moving Ahead for Progress in the 21st Century Act
ML	Machine Learning
MPMS	Multi-modal Project Management System
NASEM	National Academies of Sciences, Engineering, and Medicine
NCHRP	National Cooperative Highway Research Program
NGMDB	National Geologic Map Database
NOAA	National Oceanic and Atmospheric Administration
NRCS	Natural Resources Conservation Service

NZGD	New Zealand Geotechnical Database
OAuth	Open Authorization
OCR	Optical Character Recognition
OEM	Original Equipment Manufacturer
OGC	Open Geospatial Consortium
ORM	Object-Relational Mapping
OSHA	Occupational Safety and Health Administration
PAMS	Pavement Asset Management System
PDF	Portable Document Format
PennDOT	Pennsylvania Department of Transportation
PO	Purchase Order
PostGIS	Spatial database extension for PostgreSQL
PostgreSQL	Open-source relational database management system
QA/QC	Quality Assurance / Quality Control
QGIS	Open-source Geographic Information System
RDBMS	Relational Database Management System
REST	Representational State Transfer
RFQ	Request for Qualifications
RMS	Roadway Management System
ROI	Return on Investment
RQD	Rock Quality Designation
SfM	Structure from Motion
SGeMS	Stanford Geostatistical Modeling Software
SMS	Short Message Service
SPT	Standard Penetration Test
SQL	Structured Query Language
TAM	Transportation Asset Management
TAMP	Transportation Asset Management Plan
TBM	Tunnel Boring Machine
TIMS	Transportation Information Management System (Ohio DOT)
TIN	Triangulated Irregular Network
TRB	Transportation Research Board
UAV	Unmanned Aerial Vehicle
URL	Uniform Resource Locator

USDA	United States Department of Agriculture
USGS	United States Geological Survey
VBA	Visual Basic for Applications
VR	Virtual Reality
WebGL	Web Graphics Library
WSGI	Web Server Gateway Interface
XML	Extensible Markup Language

DEFINITIONS

Definitions of key terms are modified from publicly available definitions and software documentation.

Term (Category)	Definition and GDMS Use
3DEP	Three-Dimensional Elevation Program. A national initiative led by the United States Geological Survey (USGS) to systematically collect high-resolution topographic data across the United States using lidar and interferometric synthetic aperture radar (IfSAR). 3DEP provides publicly available digital elevation models, point clouds, and derived products at various resolutions. In the GDMS, 3DEP data supports terrain analysis, slope characterization, and spatial contextualization of boring locations and subsurface features. usgs.gov/3d-elevation-program
API	Application Programming Interface. A set of rules, protocols, and tools that allow different software systems to communicate and exchange data, serving as a bridge between applications. APIs enable applications to request and deliver information without knowledge of each system's internal workings. In the GDMS, RESTful APIs are the primary mechanism for data ingestion from field platforms and for integration with external systems. aws.amazon.com/what-is/restful-api
ArcGIS	A suite of geographic information system (GIS) software products developed by Esri for creating, managing, analyzing, and visualizing geospatial data. ArcGIS Online provides cloud-based GIS services, while ArcGIS Pro is the desktop application. The GDMS integrates with ArcGIS services through REST APIs for base maps, infrastructure layers, and spatial analysis. esri.com/arcgis
BIM	Building Information Modeling. A digital process for creating and managing information about a built asset throughout its lifecycle. BIM uses 3D models enriched with data on materials, properties, and relationships. The report references BIM interoperability as a long-term goal for the GDMS, particularly through Industry Foundation Classes (IFC) and AASHTO's BIM for Infrastructure Roadmap. nibs.org/bim
BMS	Bridge Management System. PennDOT's enterprise system for maintaining bridge inventory data, condition assessments, inspection reports, foundation information, maintenance history, and structural details. The GDMS is designed to integrate with BMS through API connections to correlate subsurface conditions with bridge foundation performance.
BoreDM	A geotechnical data management software platform specializing in organizing and querying large quantities of borehole and geotechnical data through a centralized database. BoreDM is one of several vendor-neutral field collection tools supported by the GDMS's RESTful API ingestion framework. boredmlogs.com
CesiumJS	An open-source JavaScript library for creating interactive 3D globes and maps in a web browser. In the GDMS pilot, CesiumJS serves as the 3D visualization engine, enabling subsurface visualization of boring data, three-dimensional representation of soil layers, and interactive exploration of underground features. cesium.com

CSS	Cascading Style Sheets. A stylesheet language used to describe the visual presentation (layout, colors, fonts, spacing) of HTML documents. Along with HTML and JavaScript, CSS is a core web technology used in the GDMS user interface. w3.org/Style/CSS
DIGGS	Data Interchange for Geotechnical and Geoenvironmental Specialists. A GML/XML-based schema and standard for exchanging geotechnical and geoenvironmental data between organizations, designed to standardize borehole, sampling, and test data for sharing and reuse. DIGGS includes built-in Schematron validation rules for automated quality assurance. It is the primary data-exchange schema recommended for the GDMS. geoinstitute.org/diggs
Django	A high-level, open-source Python web framework that follows a Model–View–Template (MVT) architecture and promotes rapid development with clean, pragmatic design. Django 5.1 serves as the GDMS back-end framework, providing the application structure, database Object-Relational Mapping (ORM), user authentication and authorization, administrative interface, and RESTful API endpoints. djangoproject.com
DOM	Document Object Model. A programming interface for web documents that represents the structure of an HTML or XML page as a tree of objects. JavaScript manipulates the DOM to create interactive and dynamic web page experiences in the GDMS user interface. developer.mozilla.org/DOM
EQuIS	Environmental Quality Information System. A comprehensive data management platform designed for both environmental and geotechnical data, offering tools for visualization, reporting, and regulatory compliance. EQuIS accommodates multidisciplinary datasets, making it suitable for projects spanning geotechnical and environmental data management. earthsoft.com/equis
GeoJSON	An open-standard file format based on JSON for representing geographic features (points, lines, polygons) and their non-spatial attributes. GeoJSON is widely used in web-based mapping and serves in the GDMS as a pathway for long-term storage and display of boring locations, CAD surfaces, and point clouds. geojson.org
gINT	A geotechnical data management software product developed by Bentley Systems, widely used for managing borehole data, laboratory test results, and producing boring log reports. gINT is being phased out in favor of Bentley’s cloud-based OpenGround platform. Legacy gINT files represent a significant source of historical data requiring parsing and migration into the GDMS. bentley.com/gINT
GeoPandas	An open-source Python library that extends Pandas to support spatial data types and geometric operations, enabling GIS workflows within a familiar DataFrame interface. GeoPandas integrates with Shapely for geometry handling and can read and write geospatial formats such as shapefiles and GeoJSON. In the GDMS, GeoPandas facilitates spatial joins, coordinate transformations, and geospatial analysis of boring locations and subsurface features. geopandas.org
GIS	Geographic Information System. A computer-based technology used to capture, store, manage, analyze, and visualize spatial or geographic data. GIS integrates location data with descriptive information, enabling spatial analysis and

	mapping. The GDMS is built on GIS principles with PostGIS providing the spatial database engine. esri.com/what-is-gis
GML	Geography Markup Language. An XML-based encoding standard published by the Open Geospatial Consortium (OGC) for expressing geographic features. DIGGS is built on GML, inheriting its capacity to represent complex spatial objects and relationships in a standardized, interoperable format. ogc.org/GML
Gunicorn	Green Unicorn. A Python Web Server Gateway Interface (WSGI) HTTP server designed for deploying Python web applications in production environments. Gunicorn manages multiple worker processes to handle concurrent requests efficiently and is commonly paired with a reverse proxy such as Nginx. In the GDMS architecture, Gunicorn serves as the application server that runs the Django back-end, bridging incoming HTTP requests to the Python application layer. gunicorn.org
HTML	HyperText Markup Language. The standard markup language for creating the structure and content of web pages. Together with CSS and JavaScript, HTML is a core web technology used in the GDMS user interface for rendering dashboards, maps, data tables, and interactive forms. developer.mozilla.org/HTML
IFC	Industry Foundation Classes. An open, international standard (ISO 16739) for describing and exchanging building and infrastructure data in BIM workflows. IFC provides a vendor-neutral data model for physical objects, spatial relationships, and properties. Recent experiments have demonstrated the feasibility of integrating DIGGS geotechnical data into IFC-compliant subsurface models. buildingsmart.org/IFC
InSAR	Interferometric Synthetic Aperture Radar. A remote sensing technique that compares successive radar images of the earth's surface to detect millimeter-scale ground deformation over time. InSAR is referenced in the report as a promising technology for geohazard monitoring (landslides, subsidence) that could feed data into the GDMS.
JavaScript	A programming language that enables web pages and applications to perform dynamic actions such as updating content, responding to user events, and controlling multimedia. Along with HTML and CSS, JavaScript is a core web technology. The GDMS front-end relies on JavaScript libraries including Leaflet.js, CesiumJS, and Bootstrap. developer.mozilla.org/JavaScript
JSON	JavaScript Object Notation. An open, language-independent data interchange format that uses human-readable text to represent structured data (objects, arrays, strings, numbers, booleans, and null values). JSON is widely used for API communication, configuration files, and data storage. GeoJSON extends JSON specifically for geographic feature representation. json.org; geojson.org
Kriging	A geostatistical interpolation method that uses a variogram model of spatial correlation to produce the best linear unbiased prediction of a variable at unsampled locations, along with an estimate of prediction uncertainty. Variants include ordinary kriging (constant mean), universal kriging (accounts for trends), block kriging (area averages), and cokriging (uses correlated secondary

	variables). The report identifies kriging as a primary analysis method for subsurface modeling.
Leaflet	A lightweight, open-source JavaScript library designed for creating interactive maps for both desktop and mobile browsers. Leaflet emphasizes simplicity, performance, and usability. In the GDMS, Leaflet provides the 2D mapping interface with base maps, boring location markers, layer controls, drawing tools, measurement utilities, and Esri GIS integration. leafletjs.com
LiDAR	Light Detection and Ranging. A remote sensing technology that uses pulsed laser light to measure distances and generate precise three-dimensional representations of terrain and objects. In the GDMS, LiDAR data provides high-resolution digital elevation models for 3D visualization, accurate boring location control, and design surface integration.
Matplotlib	A comprehensive, open-source Python plotting library that produces publication-quality static, animated, and interactive figures in a wide range of formats. Matplotlib provides fine-grained control over plot elements and serves as the foundational visualization engine in the GDMS for generating boring logs, subsurface profiles, and laboratory test-result plots. matplotlib.org
NumPy	A foundational open-source Python library that provides support for large, multidimensional arrays and matrices, along with a collection of high-performance mathematical functions for operating on them. NumPy underpins nearly all scientific computing in the Python ecosystem and serves as the computational backbone of the GDMS for tasks such as coordinate-array manipulation, interpolation, and numerical analysis of geotechnical test data. numpy.org
OAuth 2.0	An industry-standard authorization protocol that allows third-party applications to obtain limited access to a web service on behalf of a user without exposing the user's credentials. In the GDMS security architecture, OAuth 2.0 governs how external applications authenticate and receive permission to interact with the system's data and APIs. oauth.net/2
OCR	Optical Character Recognition. A computational process that analyzes the pixel structure of text within an image and converts those patterns into machine-readable characters using feature extraction and classification algorithms. Modern OCR systems apply machine-learning models to handle variable fonts, noise, and distortions. In the GDMS workflow, OCR is the first step for digitizing historical boring logs stored as scanned PDFs.
OpenGround	A cloud-connected geotechnical data management and reporting platform developed by Bentley Systems, evolved from the UK-based HoleBase/Keynetix product line. OpenGround supports collaboration from field data collection through desktop analysis, configurable log production, and laboratory reporting, with API access and integration with Excel and PowerBI. It is the designated successor to gINT. bentley.com/openground
Pandas	A powerful, open-source Python library for data manipulation and analysis, built around the DataFrame data structure for handling tabular data with labeled rows and columns. Pandas provides tools for reading, cleaning, transforming, aggregating, and exporting structured datasets, making it a core component of the GDMS data-processing pipeline for tasks such as parsing legacy borehole

	records and preparing laboratory test results for database ingestion. pandas.pydata.org
PAMS	Pavement Asset Management System. A PennDOT data system that stores pavement-related information including condition indices, material properties, construction history, maintenance schedules, and performance metrics. PAMS offers an API interface that enables integration with the GDMS for cross-referencing pavement performance data with subsurface geotechnical records, supporting analyses such as subgrade condition correlation, comprehensive performance evaluation, and material property assessment. penndot.pa.gov
Plotly	An open-source Python graphing library that produces interactive, browser-based visualizations including charts, maps, and three-dimensional surface plots. Plotly figures support hover tooltips, zooming, and panning, making them well suited for the GDMS web interface where users explore boring locations, subsurface models, and spatial datasets interactively. plotly.com
PostGIS	A spatial database extension for the PostgreSQL relational database management system. PostGIS adds support for geographic objects, spatial indexing (GiST indexes), and spatial query functions (distance calculations, area measurements, intersection tests). PostGIS serves as the GDMS's spatial database engine, enabling efficient storage and querying of boring locations, project boundaries, and 3D subsurface surfaces. postgis.net
PostgreSQL	An open-source, enterprise-grade relational database management system (RDBMS) known for its reliability, SQL compliance, and extensibility. PostgreSQL serves as the GDMS's core database engine, chosen for its flexibility, strong performance with complex queries, native support for JSON data types, and its ability to host the PostGIS spatial extension. postgresql.org
QA/QC	Quality Assurance / Quality Control. The combined set of systematic processes (QA) and specific checks (QC) used to ensure that data meets defined standards of accuracy, completeness, and consistency. In the GDMS, QA/QC operates at two levels: automated system validation (format compliance, range checks, coordinate validation, DIGGS schema conformance) and human-review processes (supervisor review, draft-to-final approval workflows).
RMS	Roadway Management System. PennDOT's enterprise system for maintaining roadway inventory data, drainage structure details, maintenance history, and traffic data. The GDMS is designed to integrate with RMS through API connections to provide infrastructure context for geotechnical features.
RSLog	A borehole log management software product developed by Rocscience, providing a streamlined interface for inputting, storing, and visualizing borehole data. RSLog integrates with other Rocscience tools (Slide3, RS2) and is one of the field collection platforms supported by the GDMS's vendor-neutral API. rocscience.com/rslog
Seaborn	A Python data-visualization library built on top of Matplotlib that provides a high-level interface for creating statistically informative and aesthetically refined graphics. Seaborn simplifies the generation of complex plot types such as heatmaps, pair plots, and distribution summaries, and is used in the GDMS for exploratory analysis and presentation-ready visualizations of geotechnical data trends. seaborn.pydata.org

Section 508	A provision of the U.S. Rehabilitation Act requiring federal agencies and organizations receiving federal funding to make their electronic and information technology accessible to people with disabilities. The report identifies Section 508 compliance as a requirement that must be resolved during GDMS standardization and governance. section508.gov
Scikit-learn	An open-source Python machine-learning library offering a consistent API for classification, regression, clustering, and dimensionality-reduction algorithms, along with utilities for model selection, preprocessing, and evaluation. Within the GDMS framework, Scikit-learn supports data-driven approaches such as soil classification prediction, spatial interpolation of subsurface properties, and anomaly detection in geotechnical datasets. scikit-learn.org
SQL	Structured Query Language. The standard programming language for storing, manipulating, and retrieving data in relational databases. In the GDMS, SQL-based queries provide transparent and flexible data retrieval, enabling searches by location, project, soil type, and test parameters. w3schools.com/sql
Stakeholder	An individual, group, or organization that has an interest in or is affected by a project, decision, or system. Stakeholders can be internal (e.g., employees, managers) or external (e.g., customers, regulators, communities). In the GDMS workflow stakeholders are identified. In the GDMS context, the report identifies over twenty-five distinct stakeholder groups and recommends structured engagement through stakeholder registries, influence–commitment matrices, and iterative feedback cycles to guide system development and ensure long-term adoption.
Survey123	An Esri form-based field data collection application that runs on mobile devices and integrates with ArcGIS Online. Survey123 allows field crews to capture geotechnical data with GPS coordinates, photos, and validation rules. It is one of the field collection tools supported by the GDMS’s RESTful API ingestion framework. survey123.arcgis.com
Three.js	A JavaScript library for creating and displaying 3D graphics in a web browser using WebGL. Three.js provides a high-level API for building 3D scenes with cameras, lights, materials, and geometries, abstracting much of the complexity of raw WebGL programming. threejs.org
Variogram	A mathematical function used in geostatistics that describes the spatial dependence of a geotechnical property by plotting the average squared difference between paired measurements against their separation distance. Key parameters include the nugget (measurement noise), sill (total variance), and range (distance beyond which measurements are uncorrelated). Variogram modeling is the foundational step for kriging and geostatistical simulation.
WebGL	Web Graphics Library. A JavaScript API for rendering high-performance, interactive 2D and 3D graphics within a web browser without requiring plugins. WebGL is the underlying rendering technology used by Three.js and CesiumJS in the GDMS. khr.org/webgl
XML	Extensible Markup Language. A markup language that defines rules for encoding documents in a format that is both human-readable and machine-readable. XML is the foundation for GML and DIGGS, providing the syntax for structured geotechnical data exchange. w3.org/XML

EXECUTIVE SUMMARY

Research project **Three-Dimensional Modeling and Uncertainty Quantification of Subsurface Soils** (RFQ 221216) was initiated on June 3, 2024, by Applied Research Associates, Inc. (ARA Inc.) and Schnabel Engineering, LLC on behalf of the Pennsylvania Department of Transportation (PennDOT). The project objective was to develop a framework for a comprehensive three-dimensional geologic model of Pennsylvania, integrating multiple data sources and advanced modeling techniques to enhance understanding of subsurface conditions throughout the state. The project scope encompasses four components: enhanced geologic mapping, integration of PennDOT's subsurface data, incorporation of additional data sources, and development of data management strategies. This approach covers both historical data and ongoing geotechnical investigations. The work was organized across five tasks: literature review and stakeholder interviews; implementation planning and system architecture; pilot development and demonstration; development of recommendations for adoption and expansion; and preparation of this final report.

Task 1, literature review, cataloged geotechnical data portals from peer DOTs and international bodies, and evaluated software across geotechnical modeling, web development, geostatistical analysis, and data management categories. Structured interviews with Caltrans, PennDOT, LaDOTD, and Ohio DOT identified recurring themes: the difficulty of inventorying geotechnical assets across districts, the value of standardized data structures such as DIGGS (Data Interchange for Geotechnical and Geoenvironmental Specialists), the critical role of centralized data governance, and cautious optimism toward advanced technologies including LiDAR, InSAR, and machine learning. PennDOT staff expressed a strong desire to consolidate data into a single system, adopt DIGGS or equivalent schema, and integrate with existing enterprise platforms (ECMS, OneMap, BMS, RMS).

Task 2 translated these findings into a three-tier system architecture: data sources (e.g., field collection, GIS data lakes, legacy archives), a central processing system (RESTful API layer, validation engine, PostGIS spatial database), and an access layer (web interface, export tools, analytics). The back-end was built on Django with PostgreSQL/PostGIS, using DIGGS-compliant schemas and a vendor-neutral API that accepts data from any standards-compliant field platform (e.g., BoreDM, RSLog, OpenGround, Survey123, or others). The plan specifies integration with PennDOT enterprise systems, DCNR geologic mapping, FEMA flood data, and federal data services (e.g., USGS, NOAA, EPA, NRCS). Data storage was organized around four core models; projects, borings, samples, and laboratory tests, with spatial indexing, version control, and change tracking. Security used Django's authentication framework with group-based permissions, OAuth 2.0, and token-based authentication.

The Task 3 pilot GDMS was deployed as a functional proof of concept independent of PennDOT's production IT. The pilot demonstrated a Leaflet-based interactive web map with PennDOT roadway and DCNR bedrock overlays; CesiumJS-3D visualization of subsurface boring data and soil layers; project management with status tracking and filtering; subsurface statistics tools displaying SPT N-values and variability by elevation; an uncertainty visualization tool using a "light and shadow" metaphor to estimate data confidence across five parameters for four in-situ test methods; an SQL-based query system; a construction data management demonstration using District 6 karst remediation data (low-mobility grouting); and role-based access control. The data

model comprises five entities (i.e., GDMS_Points, SPT_Blows, SPT_Lith, SPT_Lab, and Project) designed for expansion to construction, maintenance, and other geospatially referenced data. The pilot dataset is incomplete and not intended for engineering use.

Feedback from PennDOT engineers and IT staff (Task 4) was positive, particularly regarding 3D visualization and database flexibility. Participants emphasized that 2D cross-sections remain essential for design, that PostgreSQL/PostGIS is preferred over ArcGIS Online for complex queries, and that districts are eager to integrate historic datasets, GAM program data, and project files. Concerns about data interpolation, liability, and workflow consistency reinforced the need for formalized quality assurance / quality control (QA/QC) governance, Section 508 accessibility compliance, and standardized procedures. IT staff confirmed there are no fundamental barriers to enterprise integration. Task 4 identifies a broad stakeholder landscape spanning fourteen internal groups (from executive leadership and the geotechnical engineering division through IT, data governors, and training) and thirteen external categories (consulting firms, contractors, utilities, state and federal agencies, academia, and others), with a registry framework, influence–commitment matrices, and structured engagement strategies.

A four-stage expansion is recommended. Stage 1 (one to two years) establishes DIGGS-compliant schemas, formalizes QA/QC workflows, resolves accessibility and licensing, and revises PennDOT publications. Stage 2 (one to two years) expands through district-level pilots in Districts 4 and 11, testing automated georeferencing and GAM-type data integration. Stage 3 (two to five years) embeds the GDMS into enterprise systems for statewide asset coverage, automated project tracking, and centralized geotechnical standards. Stage 4 (five or more years) positions the GDMS within a digital twin environment supporting predictive analytics, BIM interoperability, and resilience planning. Detailed workflows are provided for three record types: new field records (eight-step process from pre-planning through real-time GPS capture, inline validation, and supervisor approval), historical boring records (nine steps including OCR parsing, expert review, and DIGGS-aligned modeling), and historical construction records (ten steps covering CAD-to-GIS conversion, relational linkage to geotechnical entities, and lifecycle tracking). The report frames GDMS utilization across three maturity levels: simple archive and retrieval, moderate schema-compliant system with QA/QC and 3D visualization, and an extensive lifecycle digital twin. The report also identifies future tools including real-time sensor integration, hazard early-warning, equipment tracking, and construction quality assurance.

The GDMS pilot has validated the technical feasibility of a unified geotechnical data platform for PennDOT. The transition from pilot to production requires formalizing data governance and QA/QC protocols, resolving accessibility and licensing requirements, expanding district data contributions and stakeholder engagement, developing training and change management programs, revising publications to incorporate digital data delivery, and establishing clear maintenance responsibilities. The staged expansion strategy provides a disciplined path from standardization through district pilots, statewide integration, and digital twin capabilities. The GDMS serves as PennDOT’s central system for geotechnical data stewardship.

1. INTRODUCTION AND TASKS

1.1 Background

PennDOT's geotechnical data, including boring logs, lab results, field observations, construction records, and geophysical surveys, has historically been collected and stored at the district level in a mix of paper documents, scanned PDFs, gINT databases, spreadsheets, and legacy reports. No centralized, statewide system existed for managing this information, and compliance with data submission requirements varies across districts. ECMS prompts data uploads during project closeout, but without standardized geotechnical data formats, consistent collection and integration remain limited. As a result, subsurface information is often difficult to locate, inconsistent in format, and inaccessible for reuse, leading to duplicated field investigations, inefficient planning, and lost institutional knowledge. At the same time, national guidance from FHWA, AASHTO, and TRB has increasingly emphasized integrating geotechnical data into digital delivery workflows, asset management, and risk-based decision-making, creating both an imperative and an opportunity for PennDOT to modernize.

This project **Three-Dimensional Modeling and Uncertainty Quantification of Subsurface Soils** (RFQ 221216) was initiated on June 3, 2024, following a kickoff meeting on May 15, 2024, and is performed by Applied Research Associates, Inc. (ARA) and Schnabel Engineering, LLC (Schnabel) on behalf of PennDOT. The scope covers four components: enhancement of existing geologic mapping, integration of PennDOT's subsurface data, incorporation of additional data sources, and development of data management strategies. Work is organized across five tasks: a literature review and stakeholder interviews; an implementation plan defining system architecture; development of a pilot Geotechnical Data Management System (GDMS); recommendations for organizational adoption and staged expansion; and this final report. The project builds on experience from peer DOTs including Caltrans, Ohio DOT, LaDOTD, and Florida DOT that have pursued geotechnical data portals and standardized schemas such as DIGGS, as well as international examples from the New Zealand Geotechnical Database and the British Geological Survey.

1.2 Goals

The central goal of this project is to develop a plan for the GDMS as a standards-based platform that unifies geologic mapping, historic and current geotechnical records, and authoritative external datasets. The system pursues three primary technical objectives: enhanced data integration through a centralized, DIGGS-compliant repository connected to PennDOT enterprise systems (e.g., BMS, RMS, PAMS, ECMS, OneMap) and external sources; workflow efficiency through digital field collection, automated point-of-entry validation, integrated two-dimensional (2D) and three-dimensional (3D) visualization, and streamlined draft-to-final approval; and data quality assurance through automated validation, structured human review, uncertainty quantification, and comprehensive audit trails. The underlying methodology is to enter data once, validate it at the source, and store it in one place.

Beyond the core platform, the project aims to make historical data discoverable and query-able to reduce duplicated investigations; replace paper and PDF workflows with structured, machine-readable data in support of PennDOT's digital delivery strategy; provide a data foundation for geotechnical asset management including risk-based inventorying of slopes, walls, embankments,

and hazardous sites; and establish a platform for future capabilities such as geostatistical modeling, machine learning, and digital twin integration across the infrastructure lifecycle.

The project also focuses on long-term GDMS sustainability, recognizing that technical capability alone is insufficient for long-term success. A central goal is to build governance, stakeholder engagement, and change management structures needed for sustainable adoption — including formalized data standards and QA/QC protocols, role-based training, and a staged transition from pilot to enterprise system aligned with FHWA and AASHTO asset management guidance.

1.3 Tasks

The following sections summarize the scope and methods of each task. Individual tasks are reported under their respective section headers. Task sections have been condensed from their original interim deliverables to improve readability and reduce redundant information within this consolidated report.

Task 1. Literature Review and Stakeholder Interviews

The team completed a literature review and comparative analysis of national and international geologic and geotechnical data systems, cataloging best practices, software tools, and data standards. Structured interviews with Caltrans, PennDOT, LaDOTD, and Ohio DOT captured operational practices, common challenges, and lessons learned about inventorying geotechnical assets, adopting standardized schemas, and applying remote sensing and Machine Learning (ML). Task 1 produced a written literature review, interview summaries, software comparisons, and appendices on geostatistical and uncertainty methods to inform system design.

Task 2. Implementation Plan and System Architecture

Task 2 translated Task 1 findings into a practical implementation plan and system architecture. The plan specifies a web-based GDMS built on a Django back-end with PostgreSQL/PostGIS spatial storage, DIGGS-compliant data structures, API-first ingestion, and phased deployment steps from proof of concept to rollout. Detailed components include data ingestion and legacy conversion strategies, QA/QC procedures, access control, performance optimization, and integration points with PennDOT enterprise systems and authoritative external services.

Task 3. Data Integration and Development of 3D Model

The pilot implemented the core stack and data model (borings, SPT records, lithology, lab tests, projects), provides interactive 2D/3D visualization (Leaflet, Three.js), supports project grouping and queries, and includes an uncertainty visualization tool to show data coverage and confidence. The pilot also demonstrated vendor-neutral ingestion, basic legacy conversion, and role-based access control. The developed GDMS remains a demonstration system, not yet integrated into PennDOT production IT, and the pilot dataset is incomplete and neither the data nor the interactive tools are intended for engineering use.

Task 4. Develop Recommendations

Task 4 developed recommendations for transitioning the GDMS from a research pilot to an operational tool that can be adopted across PennDOT. The task produced stakeholder identification and engagement strategies, a staged expansion plan progressing from standardization through district pilots to statewide integration, and detailed workflow guidance for three record types: new field records, historical boring records, and historical construction records. Task 4 also addresses how to establish GDMS value with stakeholders inside and outside PennDOT, and identifies potential future tools and capabilities at progressive levels of system maturity.

Task 5. Final Report

Task 5 consolidates findings, technical outputs, and recommendations from Tasks 1 through 4 into this final report and provides a summary of conclusions and recommendations. The report is intended to serve as both a comprehensive record of the work performed and a reference document for PennDOT as the Department evaluates next steps for transitioning the GDMS from a research pilot to a production enterprise system.

2. TASK 1. LITERATURE REVIEW

2.1 Task Introduction

The research team conducted a literature review and structured agency interviews to establish the state of practice in geotechnical data management and identify requirements for the GDMS. The literature review cataloged geotechnical data storage systems and portals from peer DOTs and international bodies, evaluated software tools across geotechnical modeling, web development, geostatistical analysis, and data management categories, and identified key data sources from Pennsylvania open data portals, federal agencies, and field collection platforms. Structured interviews with Caltrans, PennDOT, LaDOTD, and Ohio DOT captured operational practices, common challenges, and lessons learned regarding data inventorying, standardization, governance, and the integration of advanced technologies.

An approach to data interpretation and uncertainty quantification was also developed, covering variogram modeling, kriging interpolation, and simulation techniques. The complete literature review, software evaluations, and geostatistical analysis methods are documented in the Task 1 interim report, attached as Appendix A. The following sections summarize key findings.

2.2 Literature Summary

2.2.1 *Geologic and Geotechnical Data Storage Systems*

The literature review identified several relevant geotechnical data storage and display systems:

- **USGS National Geologic Map Database (NGMDB)** (web portal): Online tools to store and display geologic data, including EarthExplorer and the State Geologic Map Compilation Geodatabase.
- **New Zealand Geotechnical Database (NZGD)** (web portal): An online database compiling geotechnical information from historical and new sources, developed following the Christchurch earthquakes. Operates on a voluntary data-sharing model.
- **Pennsylvania DCNR GEologic Data Explorer (PaGEODE)** (web portal): GIS portal providing access to geology, quarry, well, and elevation data.
- **Washington State DNR Geologic Information Portal** (web portal): Provides access to boreholes, geophysics, hazards, seismic classification, and geologic information.
- **Utah Geologic Survey Interactive Map** (web portal): 3D terrain map with geologic maps at varying scales.

Key publications reviewed include Dasenbrock et al. (2021) on UAV-based LiDAR for geotechnical data, Pavlis and Mason (2017) on 3D geologic mapping technologies, NCHRP Synthesis 446 (Olsen et al., 2013) on geospatial technologies in transportation, and the USGS State Geologic Map Compilation Geodatabase. Complete citations are provided in the References section.

2.2.2 *Geotechnical Data Management Systems Across DOTs*

Several state DOTs have implemented systems for managing geotechnical data, with varying levels of complexity:

- **Publicly Accessible ArcGIS Layers:** Florida DOT, New Jersey DOT, Washington State DOT, and Vermont DOT provide ArcGIS layers with access to scanned PDF boring logs.

These layers offer searchable interfaces but the underlying data within PDFs is not directly queryable.

- **Custom GIS Websites:** Caltrans (GeoDOG platform), Kentucky DOT, Missouri Geologic Survey, and British Geological Survey (GeoIndex) provide interactive map-based access to geotechnical data with varying levels of data integration.
- **Alaska DOT GAM Portal:** Provides visual tracking of geotechnical assets and geohazard events with ratings reports and photographs.
- **Internal Data Management:** Maryland DOT's Boring Request Tool stores raw geotechnical data in a database and dynamically generates gINT files, incorporating machine learning for prediction in untested areas.

2.2.3 Software Tools for 3D Geologic Mapping

Evaluated software tools cover four categories: geotechnical modeling (Leapfrog Geo, ArcGIS, QGIS), web application development (Django, Leaflet.js, Three.js, CesiumJS, MapBox GL JS), geostatistical analysis (SGeMS, ArcGIS Geostatistical Analyst, Surfer), and geotechnical data management (gINT, OpenGround, RSLog, BoreDM, EQuIS, Tablogs). Virtual reality platforms were also reviewed as a future consideration.

It is important to note that GDMS serves a different purpose than commercial 3D geological modeling tools such as Leapfrog Geo. While these tools create interpretive 3D geological models from drillhole data, GDMS is a data management platform focused on collecting, validating, storing, and sharing raw geotechnical data. These approaches are complementary: data managed in the GDMS can be exported to commercial modeling tools for interpretive analysis.

2.3 Agency Interviews

The research team conducted structured interviews with representatives from agencies known to have active geotechnical data management initiatives.

2.3.1 Caltrans Meeting Summary (July 24, 2024)

The research team met with Sharid Amiri and Jacqueline Martin from Caltrans to discuss various aspects of geotechnical data management. Amiri introduced the GeoDOG platform ([link](#)), a public repository for geotechnical data related to assets such as slopes, bridges, and walls. Currently, data on this platform is stored in PDF format, which may not fully align with conventional data definitions. Amiri also mentioned that Caltrans is transitioning from gINT to a new platform, potentially RocScience, indicating an ongoing modernization effort for data storage and accessibility.

Amiri highlighted Caltrans' innovation portal for submitting research and innovation ideas, focusing on data analytics, visualization, and potential future integration of AI and machine learning. In terms of GIS integration, Amiri and Martin discussed ongoing efforts to integrate GIS into their system to improve data management and accessibility. Martin provided insights into the Vision system, a SQL-based tool for managing project delivery aspects like schedule, cost, and resources.

Both Martin and Amiri emphasized the need for improved data management practices, particularly in accessing and utilizing geotechnical data for projects. They discussed the use of GeoDOG for

archiving final documents and the associated challenges. Amiri also mentioned a management directive on AI, focusing on safety, privacy, and alignment with state policy, indicating a strategic approach to AI adoption within Caltrans.

It should be noted that California was a part of the early DIGGS development and developed an introductory document to geotechnical data management in 2008, (Turner et al) which discussed the COSMOS/PEER-LL system used for exchange and dissemination of interagency data.

2.3.2 PennDOT Meeting Summary (July 23, 2024)

The research team met with Sara Mullaney and Beverly Miller from PennDOT to discuss various aspects of geotechnical data management. PennDOT staff expressed a preference for the current use of ECMS and the evaluation of BoreDM and Autodesk software for 3D subsurface modeling, expressing a desire to consolidate data into one system, and to adopt DIGGS for standard data formats. Other systems currently used included PennDOT OneMap, Bridge Management System (BMS), and Roadway Management System (RMS). It was noted that iPads are currently used in the field by construction personnel. Miller and Mullaney indicated there was a Strike off Letter developed for AI and Machine learning. Mullaney emphasized the need for interoperability between software tools and PennDOT's data management system, highlighting the importance of adopting DIGGS for data standardization.

Miller discussed the challenges of collecting and managing data across districts, noting the centralized system for bridge database management and the varying levels of compliance with geotechnical data submission requirements among districts. It was noted that some Districts are beginning to develop inventories of geotechnical assets (e.g., slopes). Miller and Mullaney would prefer a centralized data management system, potentially managed by approximately two staff members in each District. Mullaney and Miller also explored the potential of 3D modeling and visualization, including the use of Autodesk for bridge inspection and the exploration of virtual reality applications. They indicated that top of rock and geological features would be helpful to include in the model.

The team also discussed current data validation practices and the potential for predictive modeling and AI to improve data reliability. While discussing software mandates, Miller favored standardizing on a single software for data collection, whereas Mullaney advocated for flexibility in using various tools. Additionally, Mullaney discussed the benefits of a statewide 3D geotechnical model for planning and hazard management, recognizing its value in refining project-specific data and improving the efficiency of data requests.

A separate conversation was held with Sarah McInnes that helped shape the questions and general understanding of PennDOT needs. The information shared was similar to the above discussion.

2.3.3 LaDOTD Meeting Summary (July 22, 2024)

The research team met with Jesse Rauser from LaDOTD to discuss various challenges and advancements in geotechnical data management. Rauser outlined the difficulties in assembling a diverse team of experts for data management and coordinating their skill sets. He detailed the extensive use of GIS for managing and visualizing data, including LiDAR and imagery data, which support asset location data for bridges and roadways. Rauser highlighted the agency's reliance on

GIS for various tasks, particularly in maintaining road and bridge networks. Although current GIS usage is mostly 2D, there is significant interest in transitioning to 3D GIS applications.

Rauser confirmed the use of ArcGIS Pro, indicating a transition to more advanced GIS software. He explained that data collection is conducted at the project level, utilizing systems like Headlight for construction data and tablets for borehole logging. Rauser expressed concerns about the interoperability between different data systems and emphasized the need for tools that encourage user engagement. Ongoing efforts to convert historical geotechnical data for compatibility with current systems were also discussed. Rauser mentioned a decade-long effort to reformat data schemas to be DIGGS-friendly, aiming for standardized data formats. Additionally, he acknowledged the challenges of requiring consultants to adopt new data formats like DIGGS.

2.3.4 Ohio DOT Meeting Summary (July 30, 2024)

The research team met with Stephen Taliaferro and Christopher Merklin from the Ohio Department of Transportation (DOT) to discuss various aspects of geotechnical data management. Taliaferro elaborated on the use of a field data collector tool for data collection across various geotechnical assets, emphasizing the standardization efforts and the desire for enhanced geohazard visibility. Merklin highlighted the integration of geotechnical data into a central system known as TIMS, which facilitates the association of geohazards with other assets like culverts for improved management and decision-making. Additionally, Merklin mentioned the existence of an office of data governance, underscoring its critical role in enterprise data-driven decision-making and the importance of uniform data collection and integration practices.

The discussion also explored new technologies such as LiDAR and InSAR for geohazard monitoring. Both Taliaferro and Merklin acknowledged the challenges and potential benefits of integrating these technologies into their current systems. Taliaferro expressed concerns about resource limitations affecting the reinspection of geotechnical assets, highlighting the need for efficient data management and potential savings through remote sensing technologies. Looking ahead, Merklin mentioned the potential of machine learning and predictive models to enhance data analysis and decision-making, indicating a cautious yet optimistic approach toward adopting these technologies within the agency.

2.3.5 Key Takeaways Across All Meetings

Throughout the discussions with Caltrans, PennDOT, LaDOTD, and Ohio DOT, several common themes emerged:

- **Challenges in Inventorying Geotechnical Data and Assets:** All groups acknowledged the difficulties in maintaining comprehensive inventories of geotechnical data and assets. Agencies often have requirements to send geotechnical data to a central storage program, but compliance varies across districts. However, advancements in tools and technologies are gradually improving this process.
- **Importance of Asset and Hazard Inventories for Funding:** Having a detailed inventory of assets and hazards has been shown to facilitate easier access to funding. This underscores the value of maintaining accurate and comprehensive geotechnical data records.
- **Value of Uniform Data Structures:** All groups recognized the benefits of adopting standardized data structures such as DIGGS. Uniform data formats improve data integration, accessibility, and reliability across different systems and stakeholders.

- **Integration of Advanced Technologies:** The integration of new technologies like LiDAR, InSAR, and AI holds promise for enhancing geotechnical data management. These technologies can improve data accuracy, monitoring capabilities, and predictive modeling.
- **Data Governance and Centralization:** The establishment of centralized systems and offices of data governance is critical for effective data management. Uniform data collection and integration practices ensure consistency and support enterprise-level decision-making.
- **Resource Limitations and Efficiency:** Concerns about resource limitations were common, particularly regarding the reinspection of geotechnical assets. Efficient data management practices and the use of remote sensing technologies can help mitigate these challenges.
- **Future Directions:** There is cautious optimism about the adoption of machine learning and predictive models. These technologies have the potential to enhance data analysis and decision-making, although their implementation requires careful consideration of resources and capabilities.

2.4 Data Integration, Standardization, and Governance

The literature review identified data integration and standardization as critical to GDMS success. PennDOT's historic geotechnical data primarily exists as paper and PDF boring logs and supplemental laboratory data, typically stored at the district level, with variable success in gathering digital gINT files at the Central Office. Two key standards were identified for adoption:

- **DIGGS (Data Interchange for Geotechnical and Geoenvironmental Specialists):** A GML/XML-based schema for standardized geotechnical data exchange, with growing adoption among consulting organizations and software vendors. DIGGS offers vendor-neutral data exchange, built-in schematron validation rules for automated QA/QC, and alignment with OGC standards for spatial data interoperability.
- **AGS (Association of Geotechnical & Geoenvironmental Specialists) Format:** Well-established in Europe and Asia, providing another standardized approach to geotechnical data management.

Recent interoperability experiments between OGC and IFC have demonstrated the feasibility of integrating both DIGGS and AGS into IFC-compliant subsurface data models.

2.4.1 Recommended Key Data Sources

The following data sources are recommended for use with the GDMS as they are readily available and can contribute to the comprehensiveness and accuracy of the GDMS:

- **Pennsylvania DOT Open Data**
 - URL: data-pennshare.opendata.arcgis.com
 - Content: Geological data, Private Property, Right of Way, and Structural geology.
- **Pennsylvania DCNR Open Data**
 - URL: newdata-dcnr.opendata.arcgis.com
 - Content: Bedrock geology, surficial geology, and structural geology.
- **Pennsylvania Department of Environmental Protection Open Data Portal**
 - URL: newdata-padep-1.opendata.arcgis.com
 - Content: Environmental data, including mining activities, oil and gas wells, and water resources.
- **FEMA Flood Maps**

- URL: hazards-fema.maps.arcgis.com
 - Content: Flood hazard data, including flood zones and floodplain boundaries.
- **USGS GeoLog Locator**
 - URL: [USGS - GeoLog Locator](https://geologlocator.usgs.gov/)
 - Content: Borehole geophysical logs and related data
- **National Geologic Map Database:** Interactive Database for Geologic Maps of the United States [MapView \(beta\) | NGMDB \(usgs.gov\)](https://ngmdb.usgs.gov/)
- **GEOSETTA**
 - URL: <https://geosetta.org/>
 - Content: Public Geotechnical Data sets, subsurface predictions

2.4.2 Data Collection and Integration Tools

Integrating field data collection tools such as Survey123 and ESRI Field Maps and Vendor specific collectors enables field personnel to submit data in real time. These tools allow for the collection and submission of geotechnical data, including photos, GPS coordinates, and descriptive information, directly from the field.

- **Survey123 Integration:**
 - Custom Forms can be tailored to PennDOT's data collection needs.
 - Data Submission is provided by real-time synchronization with the central database where internet access is available and stored locally until access is restored.
 - GIS integration is seamless where data element collection is referenced with coordinates and elevations.
- **ESRI Field Maps Integration:**
 - Field Data Collection: Spatial data, geotechnical observations, and infrastructure locations.
 - Offline Capabilities: Data collection in areas without network coverage, with automatic synchronization once reconnected.
- **Vendor Supported Field Collection Tools:**
 - Most geotechnical software available today has focused on including a field collector application. These perform in various ways, but are generally designed to quickly populate the parent software geotechnical database, provide some quick visualization of the data in the field, and provide backup protection should internet access not be readily available.

A key element of field collection and vendor platforms available today is the flexibility and accessibility of data through REST API services. This compatibility will support integration with the GDMS platform. Using these sources and tools, the GDMS can remain current, authoritative, and scalable as new data becomes available.

2.5 Data Interpretation and Uncertainty Quantification

Effective geotechnical modeling requires integrating diverse data sets and applying defensible methods for data interpretation and uncertainty quantification. This section summarizes methodologies, tools, and strategies using open-source Python libraries (as an example) for data analysis, highlighting geostatistical methods. Detailed formulations and examples are provided in the Task 1 interim report attached in Appendix A.

2.5.1 Data Interpretation Using Python

Python is frequently utilized for its robust ecosystem of libraries that facilitate data analysis, visualization, and geospatial processing. These capabilities make Python a prominent tool for geotechnical data interpretation and uncertainty quantification. Several key advantages of Python include:

- **Data Manipulation and Analysis:** Libraries such as NumPy and Pandas are tools often utilized due to their efficiency in handling large datasets, such as borehole logs and soil properties. These tools enable researchers to perform complex statistical analyses and data transformations critical for geotechnical design (Harris et al., 2020; McKinney, 2010).
- **Geospatial Processing:** GeoPandas extends Pandas' capabilities to geospatial data, facilitating spatial queries and visualizations. This is crucial in geotechnical engineering, where the spatial distribution of soil types and groundwater levels plays a significant role in risk assessment and site investigations (Jordahl et al., 2020).
- **Machine Learning:** Scikit-learn provides machine learning algorithms applied to predictive modeling and pattern recognition in geotechnical contexts, such as predicting soil properties or classifying soil types (Pedregosa et al., 2011).
- **Visualization:** The use of visualization libraries such as Matplotlib, Seaborn, and Plotly are well-documented in the literature and allow the creation of detailed borehole logs, cross-sectional views, and 3D models. Visualization is essential for conveying complex subsurface conditions to stakeholders (Hunter, 2007; Waskom, 2021).
- **Integration:** Python's flexibility in integrating with various back-end systems, such as Django, is a key advantage in developing web-based applications that streamline data analysis and decision-making processes (Holovaty & Kaplan-Moss, 2009).

These tools and methodologies have been foundational in recent geotechnical research, underscoring Python's versatility and effectiveness in geotechnical modeling.

2.5.2 Uncertainty Quantification and Geostatistical Analysis

Uncertainty quantification is essential in subsurface modeling due to the inherent variability of geological materials. As Curran and Hammah (2006) noted, "uncertainty is king; make room for it." Geological materials, formed under diverse thermo-physio-chemical conditions, have histories that are often incomplete or unknown. Consequently, their properties can vary significantly with location. A geotechnical parameter predicted from a model based on a single observation may be inaccurate, making it crucial to account for geospatial uncertainty.

Variogram Modeling

- **Definition:** The variogram measures the relationship between two locations based on the variance of differences between them, providing insight into spatial continuity. (Barnes, 2003).
- **Importance:** Variograms reveal spatial patterns that are not evident in standard statistical summaries, allowing for more accurate predictions.
- **Components:** Key components include the nugget effect, sill, and range, each representing different aspects of spatial variability.

- **Modeling:** Analysts use various models, including spherical, exponential, and Gaussian models, to fit experimental data, guiding geostatistical interpolation and improving model accuracy. These models help in understanding the degree of spatial correlation, which is essential for making reliable predictions in areas where direct data may be sparse (Cressie, 1993).

Geostatistical Interpolation

- **Kriging:** A statistical interpolation method that uses variograms to account for spatial correlation between sample points, providing the best linear unbiased prediction of unknown values. (Matheron, 1963)
- **IDW (Inverse Distance Weighted):** This deterministic method assigns weights inversely proportional to the distance between sample points, providing a simpler but less statistically robust approach than kriging. It is often used when a quick estimation is needed, though it may not account for spatial correlation as effectively as kriging (Shepard, 1968).

Types of Kriging:

- **Ordinary Kriging:** Assumes constant mean and variance across the spatial field, making it suitable for datasets without apparent trends.
- **Universal Kriging:** Accounts for trends in the data by incorporating a drift function, useful in areas where soil properties change systematically with location.
- **Block Kriging:** Estimates average values over a block of grids, reducing prediction error and making it ideal for regional-scale assessments.
- **Cokriging:** Enhances accuracy using correlated variables, beneficial when multiple related geotechnical properties are measured.

Limitations of Kriging

- The accuracy of kriging heavily depends on the quality of the variogram model, and poor variogram fitting can lead to unreliable predictions. Additionally, in cases where sample data is sparse or highly variable, kriging may result in "over-smoothing," masking important local variations (Webster & Oliver, 2007).

Geostatistical Simulation

Geostatistical simulation offers an alternative to interpolation by generating multiple realizations that reflect the full range of possible outcomes. This approach is valuable for risk assessment and decision-making in environments with significant uncertainty (Goovaerts, 1997).

- **Sequential Indicator Simulation:** Used for categorical data, such as soil classification, to produce multiple realizations that honor the data's spatial distribution.
- **Sequential Gaussian Simulation:** Applied to continuous data, allowing for the creation of multiple equiprobable realizations that capture the uncertainty in the predictions.

Simulations provide not only a single value for unknown locations but also the uncertainty associated with the estimated value through the conditional cumulative distribution function (CCDF), making them a powerful tool for uncertainty quantification.

2.5.3 Case Studies and Applications

Several case studies demonstrate the practical application of these methodologies in real-world scenarios:

- **Case Study 1: 3D Geotechnical Modeling in Turin, Italy (Rienzo et al., 2008)**
 - Created a geological map using data from 380 boreholes covering approximately 50 km².
 - Developed a 3D GIS model to manage spatial data for a geological Turin model.
 - Utilized deterministic triangular network (TIN) method for spatial interpolation.
 - Conducted uniaxial compression and shear tests on soil samples to examine geotechnical properties.
 - Correlated geotechnical parameters with cementation maps.
- **Case Study 2: Integration of BIM and GIS for Geotechnical Zoning in Peshawar, Pakistan (Khan et al., 2021)**
 - Created a geodatabase incorporating borehole information for geostatistical interpolation through kriging.
 - Produced digital maps illustrating geotechnical parameters like SPT-N values, plasticity index, and unconfined compressive strength.
 - Employed Autodesk Civil 3D for BIM to handle coordinate systems, visualize borehole data, and create geotechnical surfaces and solids.
 - Integrated BIM-GIS approach facilitated storage and management of geotechnical information in IFC format.
- **Case Study 3: AI-Based Site Characterization by the University of Dayton (Wang et al., 2020)**
 - Developed a methodology combining AI-based machine learning algorithms with other disciplinary principles to interpret geotechnical site investigation data.
 - Used a machine learning module to analyze original site characterization data and deduce subsurface stratification models.
 - Incorporated stochastic simulation techniques to estimate subsurface conditions at unexplored locations with uncertainty quantification.
 - Demonstrated benefits of the methodology through three case studies.

These studies highlight the effectiveness of combining geostatistical analysis with advanced visualization and modeling techniques to enhance the reliability of geotechnical models.

2.6 Task Conclusions

The GDMS should integrate current modeling technology with centralized data management. Such a system will significantly enhance PennDOT's capabilities in infrastructure planning, design, and management by offering an advanced tool for managing, visualizing and analyzing subsurface conditions. The following conclusions developed during Task 1 span six themes: decision-making and cost efficiency (items 1, 6), data standardization and management (items 2, 9), collaboration and training (items 3, 5), scalability (item 4), historical data integration (items 8, 11), and analytical capabilities (items 10, 12). The benefits of this project are far-reaching and include the following:

1. **Informed Decision Making:** The interactive 3D model will provide PennDOT with a detailed understanding of subsurface conditions, leading to more informed and confident decisions in infrastructure planning and design. This will help mitigate risks and optimize

project costs by reducing the uncertainties typically associated with subsurface investigations.

2. **Streamlined Data Management:** By integrating diverse data sources—ranging from historic geotechnical records to new data collections—into a unified model, the project will streamline data management processes, making it easier to analyze, store, and retrieve critical geotechnical information.
3. **Enhanced Collaboration:** The web-based interface, along with potential VR/AR integration, will facilitate better communication, collaboration and cognition among various stakeholders, including engineers, geologists, policymakers, and the public. This improved collaboration will support more cohesive and effective project execution.
4. **Scalability and Flexibility:** The use of open-source technologies and a modular design approach ensures that the model can easily be expanded and adapted to accommodate future data sources, technological advancements, and evolving project needs.
5. **Educational and Training Value:** The 3D model will serve as an educational resource for training new professionals in geotechnical engineering and public infrastructure management. It will also enhance public understanding of subsurface conditions and their impact on infrastructure.
6. **Cost Savings in Site Investigations:** By providing detailed preliminary insights into subsurface conditions, the model will help optimize site investigation strategies, potentially leading to significant cost savings on field investigations and testing.
7. **Advancement in Public Infrastructure Management:** This project would position PennDOT as a leader in the application of advanced technologies for infrastructure management, setting a benchmark for other state agencies to follow.
8. **Historical Data Integration:** By leveraging historic geotechnical data and integrating it with new data through tools like Geosetta, the model will offer a more complete and accurate representation of Pennsylvania’s subsurface conditions, improving the reliability of geotechnical assessments.
9. **Standardized Data Practices:** The adoption of industry-standard data formats like DIGGS ensures consistent data collection, storage, and exchange, which will enhance collaboration with external partners and future-proof the database for technological advancements.
10. **Predictive Capabilities:** Incorporating machine learning and geostatistical models for predicting SPT, strength, plasticity, etc. values and other geotechnical parameters will allow for comprehensive preliminary assessments in areas with limited physical testing data, further reducing uncertainty.
11. **Wider Data Coverage:** By integrating Geosetta's API for accessing nearby historical data, the model will provide a more holistic view of subsurface conditions, enhancing PennDOT’s ability to plan and execute projects effectively.
12. **Quality Assurance for New Data:** The model will serve as a platform for validating and ensuring the quality of new geotechnical data submissions, thereby improving the overall consistency and reliability of data used in future infrastructure projects.

The conclusions above reflect consistent themes across the literature, software evaluation, and agency interviews. These findings directly informed the system architecture developed in Task 2.

3. TASK 2. IMPLEMENTATION PLAN

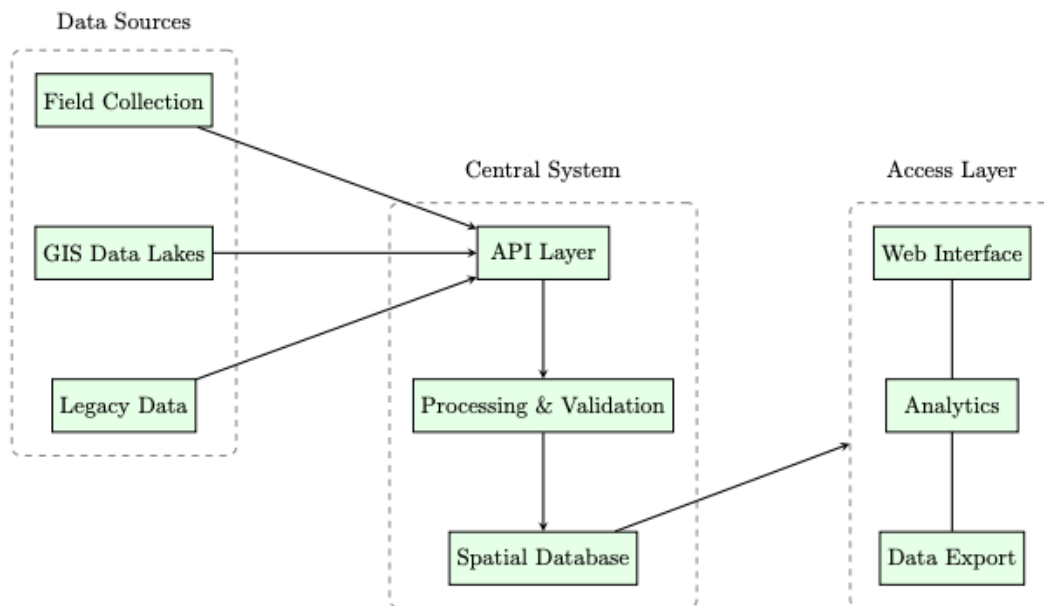
3.1 Task Introduction

Building on the findings from Task 1, Task 2 developed a detailed implementation framework and system architecture for PennDOT's GDMS. The proposed design addresses both immediate data management needs and future analytical capabilities, including 2D and 3D visualization and uncertainty quantification. The architecture specifies a three-tier, web-based system encompassing data sources, a central processing and storage layer, and a user access layer; built on open-source technologies to eliminate vendor lock-in and reduce long-term licensing costs.

Key components developed in this task include a core data model organized around projects, borings, samples, and laboratory tests; integration strategies for PennDOT enterprise systems and authoritative external data sources; data ingestion and legacy conversion procedures; QA/QC validation frameworks; user interface and visualization requirements; system administration and maintenance responsibilities; and a phased implementation plan from proof of concept through full deployment. The following sections summarize key architectural decisions and system components.

3.2 System Architecture

The GDMS should implement a three-tier, web-based architecture (Figure 1) built on open-source, non-proprietary technologies to eliminate vendor lock-in and reduce long-term licensing costs.



System Benefits:
Ease of Use • Security & Control • Flexibility & Growth

Figure 1: Proposed system architecture.

The system architecture is recommended to consist of these three primary components that work together to collect, process, and deliver geotechnical data.

- **Data Sources:** Field data collection from any standards-compliant platform (BoreDM, RSLog, OpenGround, Survey123, or others) through RESTful API integration with PennDOT's enterprise data systems; along with legacy data from historic reports, geotechnical investigations, and geological records.
- **Central System:** A Django 5.1 back-end with PostgreSQL/PostGIS spatial database, implementing DIGGS-compliant schemas for standardized data exchange. The central system provides API-based data ingestion with format validation, quality control checks, and spatial reference validation.
- **Access Layer:** A web interface with interactive mapping (Leaflet.js), 3D visualization (Three.js), data export in multiple formats (CSV, Excel, DIGGS, Shapefile, GeoJSON, DXF/DWG, PDF), and SQL-based query capabilities.

The GDMS should provide standardized RESTful API endpoints so that any compliant field platform can connect. The responsibility for API integration should lie with the field tool vendor, not PennDOT, consistent with how DIGGS adoption works industry-wide.

It should be noted that the software tools described here (Django, PostgreSQL/PostGIS, Leaflet, Three.js) are open-source components assembled to build the custom GDMS platform. Other similar tools are available and should be considered to meet specific needs as the GDMS is developed. The project team evaluated commercial off-the-shelf geotechnical data management systems during Task 1 and found that a custom platform is likely to better serve PennDOT's specific integration, visualization, and scalability requirements. Commercial tools remain complementary for field data collection and specialized analysis.

The overall data storage and management system supporting a core database module, that in itself could be connected to other core PennDOT systems (Figure 2). GDMS should implement PostGIS as its primary database with spatial indexing, version control, change tracking (including identification of who collected data, who made modifications, and who approved the final record), and comprehensive metadata management.

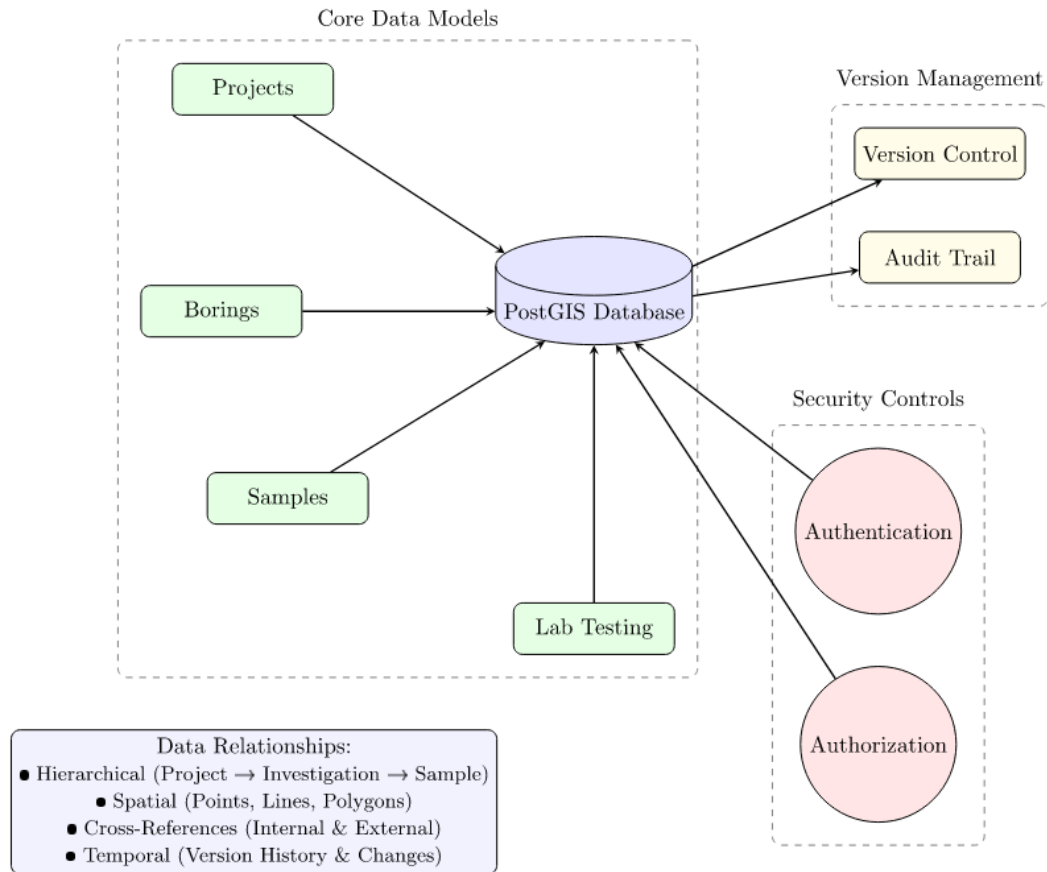


Figure 2: Visualization of GDMS data storage and management systems.

3.2.1 Core Data Model

The database architecture is organized around four primary data models: **Projects** (organizational structure, metadata, boundaries, timelines), **Borings** (location, elevation, drilling details, equipment, temporal data), **Samples** (physical properties, recovery, chain of custody, testing assignments), and **Laboratory Tests** (results, methodologies, quality indicators, DIGGS validation). Spatial data management uses PostGIS for point, linear, polygonal, and three-dimensional geometry types with spatial indexing. Data relationships follow the DIGGS standard hierarchy, ensuring industry-wide compatibility. The data model is designed to accommodate future expansion such as geophysical data, construction records, foundation elements, and maintenance data objects.

3.3 External Data Sources and Integration

The GDMS connects directly to authoritative external data sources rather than duplicating their contents (Figure 3). This enables PennDOT to leverage data maintained by specialized organizations while focusing GDMS resources on geotechnical-specific capabilities. Integration targets include:

- **PennDOT Enterprise Systems:** Bridge Management System (BMS), Roadway Management System (RMS), Pavement Asset Management System (PAMS), Survey and

Mapping (including LiDAR via USGS 3DEP), and Design Standards (Publication 293, Publication 222, Publication 408, and others).

- **Geographic Information Systems:** ArcGIS Online Services for base maps, infrastructure layers, and spatial context
- **Pennsylvania-Specific Sources:** DCNR geologic mapping (bedrock geology, surficial deposits, structural features, mineral resources, groundwater), DEP environmental data. These PA-specific sources directly support PennDOT's unique subsurface challenges including karst, mining, and complex geology.
- **Additional Data Sources:** USGS data services (geologic mapping, stream gauges, LiDAR/3DEP), FEMA flood data, NOAA/EPA climate and environmental data, NRCS/USDA Web Soil Survey.

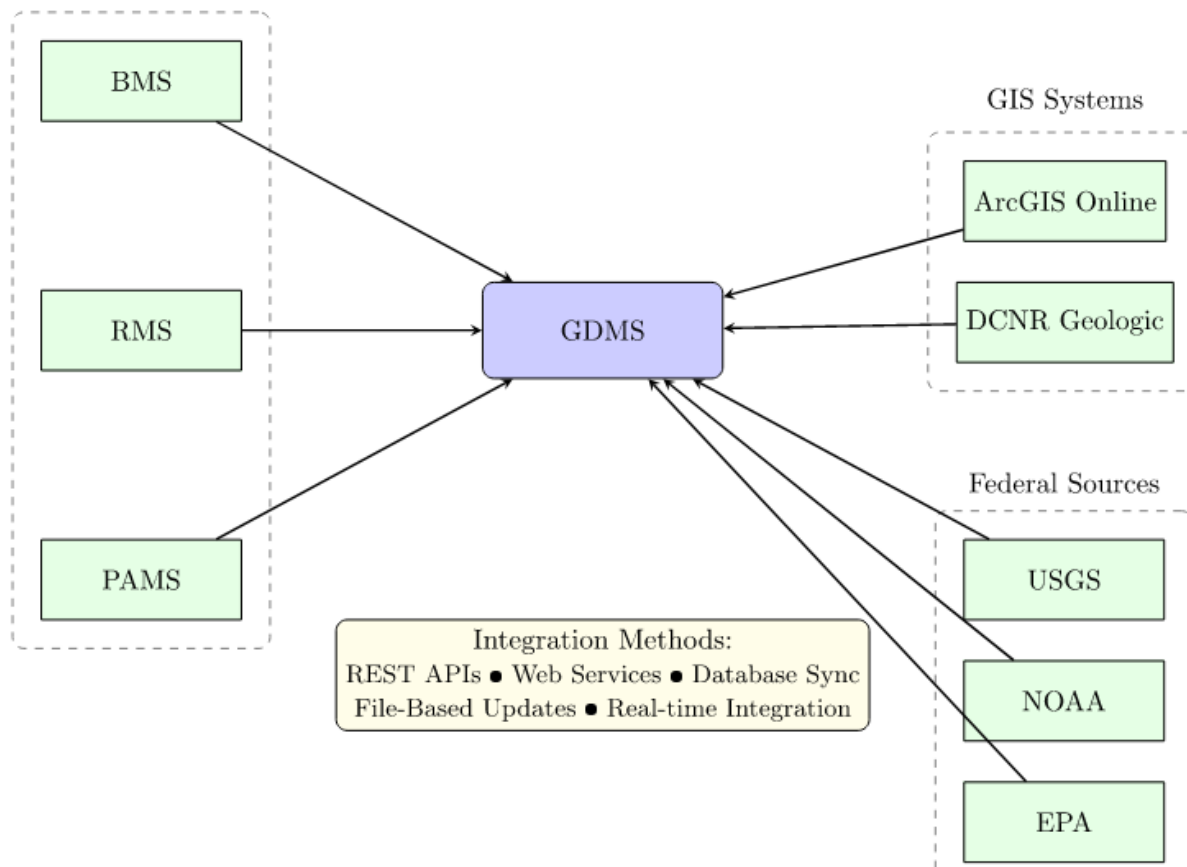


Figure 3: Visualization of external source integration.

Integration is achieved through REST API connections and tiered synchronization based on data criticality and update frequency. Pennsylvania LiDAR data is available through the USGS 3D Elevation Program (3DEP), which provides statewide coverage. Much of this data is maintained by authoritative federal and state agencies, ensuring currency and reliability.

Integration security should be maintained through:

- **Robust API key management:** Unique cryptographic keys can be assigned to each application or service accessing the system, allowing precise tracking and control of all data requests while enabling immediate access revocation if needed.

- **OAuth 2.0 authorization:** This industry-standard protocol provides secure delegated access, allowing third-party applications to obtain limited access to the system on behalf of users without exposing their credentials.
- **Token-based authentication:** Short-lived access tokens can be used to validate each request, ensuring that even if a token is compromised, it can only be used for a limited time before requiring re-authentication.
- **IP restrictions for database connections:** Access to critical databases can be limited to specific IP addresses or ranges, creating an additional security barrier that helps prevent unauthorized access attempts from unknown locations.

This framework enables the GDMS to provide reliable access to current, authoritative data while minimizing data management overhead and maximizing system efficiency. Wherever possible the GDMS system should maintain direct connections to authoritative sources rather than duplicating data so that users have access to the most current information while reducing maintenance burden and improving overall system sustainability.

3.3.1 Data Ingestion

The GDMS should handle both modern field collection data and legacy geotechnical information through a unified ingestion framework (Figure 4). Field data collection supporting any modern platform through standardized REST APIs can provide real-time data submission with GPS location tracking, photo integration, offline capability, and integrated validation. The system’s vendor-neutral approach would ensure field teams and consultants can continue using preferred tools while maintaining consistent data integration and quality.

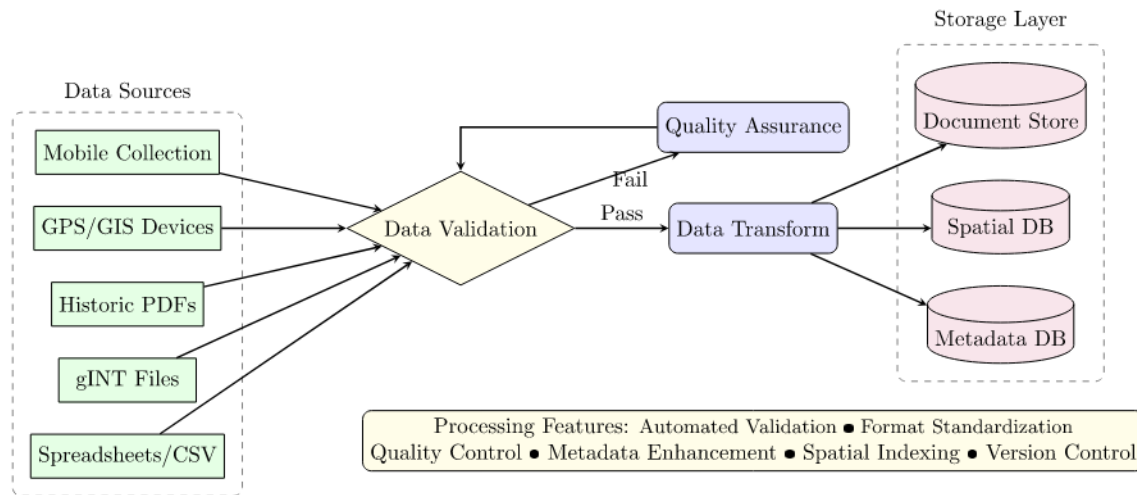


Figure 4: Visualization of data ingestion framework.

Legacy data processing addresses historical PDFs (OCR extraction with expert review), gINT files (automated parsing with quality verification), and spreadsheet/CSV data (standardized transformation pipelines). Quality assurance combines automated validation (format compliance, value ranges, spatial reference accuracy, DIGGS schema compliance) with user review processes (data entry review, supervisor approval, configurable workflows). The framework for this data ingestion is further developed in Task 4.

The DIGGS data model is designed to accommodate future expansion. An example of this is how raw geophysical data files (e.g., seismic, resistivity, ground-penetrating radar) can be stored as binary objects with metadata, while interpreted results can be ingested as georeferenced data points or profile images linked to project and location records.

3.3.2 Data Quality Assurance

Maintaining high data quality standards will be essential for long term system success. The platform is recommended to implement comprehensive validation procedures and data standards that ensure consistency and reliability across all geotechnical information. A key concept to achieving this is to enter data only once (where it can be validated at the time of entry) and store it in only one place. This can be supplemented with uncertainty quantification methods to gain a clear understanding of data reliability and guidance for the need of subsequent exploration, while robust maintenance procedures shall ensure ongoing data integrity.

3.4 Data Storage and Management

The GDMS should implement a data storage architecture centered on PostGIS as its primary spatial database. Spatial data management should handle multiple geometry types; point features for borings and sample locations, linear features for cross-sections and geophysical profiles, polygonal features for project boundaries and geological units, and three-dimensional surfaces for subsurface modeling. Spatial relationships should be managed through indexing and topological rules, enabling efficient proximity calculations and spatial analysis across all data types. The pilot (Task 3) demonstrated a subset of these capabilities using five core entities; the production system should expand to support the full range of geometry types and spatial relationships. Data relationships should follow the DIGGS standard hierarchy, maintaining DIGGS-compliant connections between projects, investigations, borings, samples, and test results, with both internal referencing (test correlations, stratigraphic relationships) and external referencing (persistent identifiers and connections to other systems).

Change tracking should utilize Django's built-in capabilities to maintain modification history, including user identification, timestamps for key data changes, and audit logging to track critical operations affecting data integrity and security. These capabilities were not implemented in the pilot but are recommended for the production system.

Security should engage Django's authentication framework with group-based permissions, OAuth 2.0, and token-based authentication; the pilot demonstrated basic Django authentication with group-based permissions, and production deployment should add OAuth 2.0 and token-based authentication for external application access. Performance optimization should leverage Django's ORM, including proper model indexing, query result caching, database connection pooling, and lazy loading of related objects.

3.4.1 User Interface and Access

The GDMS user interface should provide integrated dashboards with project overviews, activity feeds, and system notifications (Figure 5). Key interface components should include at a minimum:

- **Map Interface:** Leaflet-based interactive mapping with multiple base maps, layer control, drawing tools, and measurement utilities

- **Visualization:** Plotly-based boring log visualization, cross-section development, property correlation charts, and time-series displays for 2D interpretation and Three.js -powered subsurface visualization with interactive point cloud display, surface modeling, and profile generation for 3D visualization potentially integrated with the structural and roadway models.
- **Query System:** Full-text search, advanced query builder with spatial selection, property-based filters, and saved query templates
- **Data Export:** CSV, Excel, DIGGS, Shapefile, GeoJSON, CAD formats, PDF reports, and shareable secure project links
- **Analytical Tools:** custom built data aggregation and analytical tools such as subsurface uncertainty models.

The interface should implement responsive design for desktop, tablet, and mobile platforms. The GDMS visualization framework can be developed to support template-based reporting, allowing PennDOT to define preferred display formats and data presentation standards.

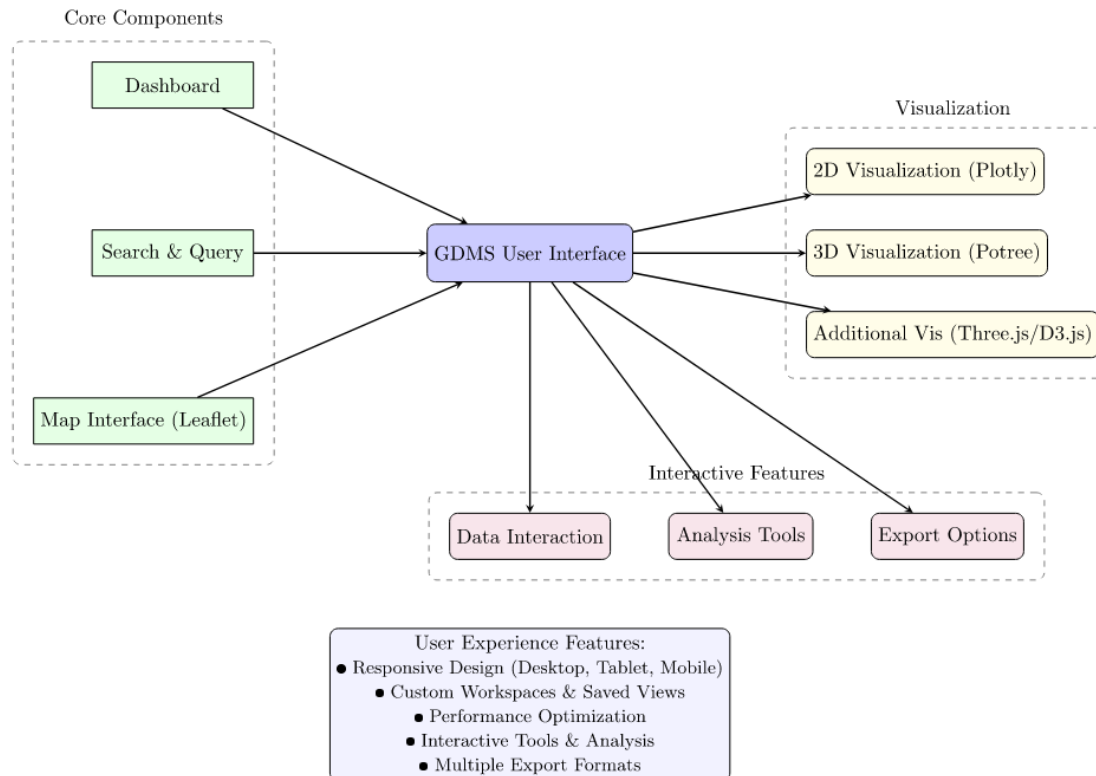


Figure 5: Visualization of components, features, and visualization tools.

3.5 System Administration and Maintenance

System administration and maintenance are necessary for ongoing system operations, ensuring reliability, security, and performance.

3.5.1 System Monitoring

Performance Monitoring

The system is recommended to leverage Django's built-in monitoring and debugging tools to track critical components and ensure optimal performance. Initial monitoring will utilize Django Debug Toolbar for development insights and Django's logging framework for production monitoring. This practical approach will provide essential metrics including: server performance metrics, database query analysis, application logging, session tracking, and basic resource utilization monitoring.

3.5.2 User Administration

User management is recommended to implement Django's built-in User model and authentication back-end, providing robust account lifecycle management and access control. Account administration should leverage Django's admin interface for user management, role assignments, and permission control. Activity monitoring will utilize Django's logging framework to track key user interactions including: authentication events, resource access, data modifications, administrative operations, and session management.

3.5.3 System Maintenance

The system maintenance responsibilities are divided between internal PennDOT staff and external development or support resources.

PennDOT Maintenance Responsibilities

The following maintenance procedures should be implemented by PennDOT's IT team for day-to-day system operations: regular database backups using standard PostgreSQL utilities, basic system monitoring and health checks, geotechnical group admin account management through Django's admin interface, routine security updates and patches, and basic performance monitoring and optimization.

Development Team Support

The external development team should provide specialized technical support for: major version upgrades, complex database optimizations, custom feature development, architecture modifications, and performance troubleshooting for complex issues.

3.5.4 Support Structure

User and system support should also be divided into two camps. The first line of support should be developed internally within PennDOT so that users can experience efficient response times. More complex support may require input from the development team to address more complex issues.

PennDOT Internal Support

PennDOT should establish an internal support structure to handle: first-line user support and basic troubleshooting, basic system administration tasks, internal user training and documentation, and standard operational procedures.

External Technical Support

The development team should provide secondary support for: complex technical issues beyond internal support capabilities, system optimization recommendations, feature enhancement consulting, internal and external technical training, and critical issue resolution.

3.5.5 Security Management

PennDOT internal IT security management team should be responsible for establishing security procedures, basic monitoring, account management policies, and compliance metrics. While the external development team's security support would develop security patch recommendations, resolve complex security issues, consult on security architecture, and implement best practice features and guidance.

3.5.6 Professional Support Services

Long-term system success relies on collaboration between PennDOT's internal teams and the development team. While PennDOT manages day-to-day operations, the development team remains available for:

- Complex technical consulting
- Major system enhancements
- Performance optimization recommendations
- Architecture evolution guidance
- Advanced feature development

This collaborative approach ensures that PennDOT maintains operational control of the system while having access to specialized technical expertise when needed. The development team serves as a technical resource for complex issues and system evolution rather than handling routine maintenance and support.

3.6 Implementation Plan

The GDMS implementation follows two semi-independent plans, one for development of the GDMS system and initial piloting, detailed in this section, and a second for PennDOT expansion, detailed in Task 4. Considering the GDMS system and initial piloting a four-phase approach as outlined in Figure 6 is recommended. Additional details beyond initial system development and pilot are detailed in Task 4.

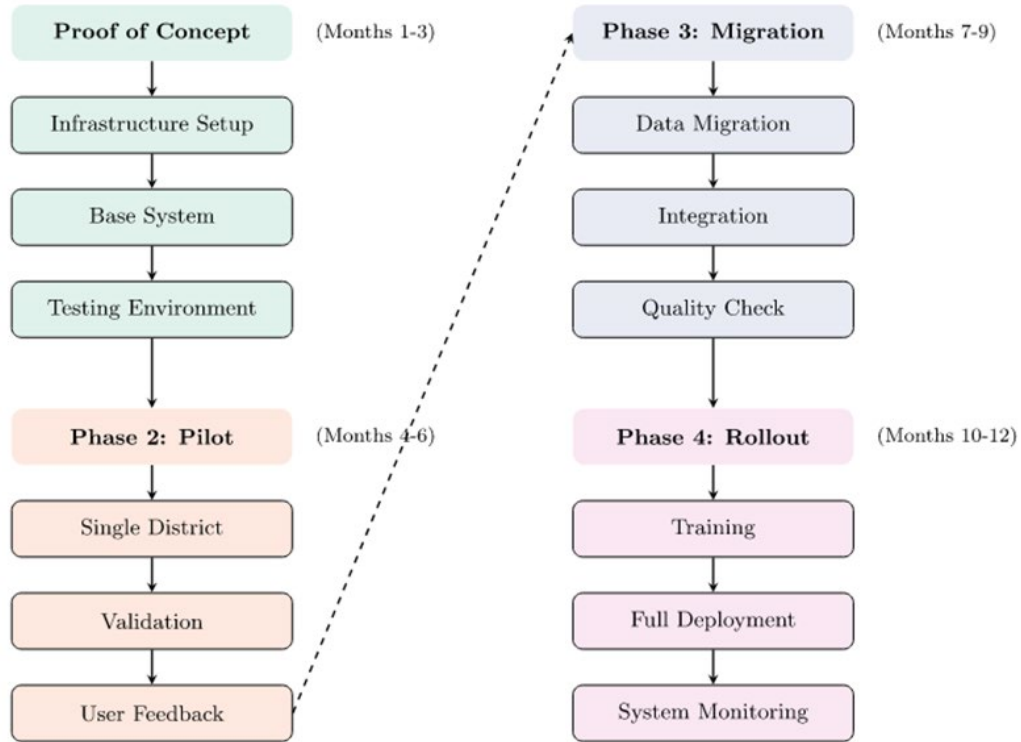


Figure 6: Proposed implementation schedule.

Each phase of the four phases focuses on three key tasks summarized below:

- **Phase 1: Proof of Concept (Pilot):** Establish basic Django deployment with PostgreSQL/PostGIS, implement core data models with DIGGS compliance, create essential user interfaces, and demonstrate integration with external services. The pilot project also included a construction data management demonstration using District 6 karst remediation data, including low-mobility grouting volumes and subsurface treatment visualization (see Task 3).
- **Phase 2: Pilot Implementation Program (Future):** The Pilot set up in Phase 1 should be implemented in a shared way so that a selected representative PennDOT district can initiate a full deployment, perform data migration and validation, conduct integration testing with external systems, collect user feedback, and evaluate geophysical and additional data integration including storage of raw binary files and incorporation of interpreted results.
- **Phase 3: Data Migration (Future):** Systematic transfer of historical and current data across districts with comprehensive quality verification, implementation of standardized APIs, and parallel data cleanup operations.
- **Phase 4: Full Rollout (Future):** Multi-district deployment with coordinated rollout scheduling, comprehensive training programs, local system optimization, and establishment of support structures.

Success should be measured with each phase. Success metric to consider should span technical performance (response times, validation rates, integration reliability, uptime) and user adoption

(adoption rates, feature usage, satisfaction scores, time savings). The research team also recommends employing a proactive risk management strategy addressing both technical and operational challenges during each phase which should include comprehensive testing protocols, robust backup and recovery procedures, continuous performance monitoring and optimization, regular security validation and penetration testing, and system redundancy and failover capabilities for critical functions.

These metrics and risk management procedures will help guide the initial proof of concept implementation and inform the path toward full system deployment. By carefully monitoring both technical performance and user adoption, the research team can identify and address potential issues early while ensuring the system meets PennDOT's needs effectively.

3.7 Task Conclusion

The proposed GDMS architecture provides a standards-based, scalable platform for geotechnical data management. By leveraging open-source technologies and DIGGS-compliant schemas, the system avoids vendor lock-in while enabling integration with PennDOT's enterprise systems and authoritative external data sources. The implementation plan defines a core data model, data ingestion and legacy conversion strategies, QA/QC validation procedures, user interface and visualization requirements, and system administration responsibilities. The phased implementation approach, from proof of concept through pilot, data migration, and full rollout, is designed to validate core capabilities incrementally before expanding to statewide deployment. These components provided the technical specification for the pilot system developed in Task 3.

4. TASK 3. DATA INTEGRATION AND DEVELOPMENT OF PILOT 3D MODEL

4.1 Task Introduction

Task 3 implemented the architecture and data model defined in Task 2 as a pilot GDMS. The pilot was developed as a proof-of-concept demonstration, independent of PennDOT's IT infrastructure, to validate the technical approach and illustrate how a centralized geotechnical data platform could function at a statewide scale. The pilot is not a production system; it is a working model intended to inform future development and integration into PennDOT's enterprise environment. The pilot dataset is incomplete and not intended for engineering use.

The pilot is a Django-based web application built on a PostgreSQL/PostGIS database. The data model organizes borings, Standard Penetration Test results, lithology, laboratory data, and project metadata into a relational framework designed for future expansion. Front-end capabilities include an interactive 2D web map (Leaflet.js) with PennDOT roadway and DCNR bedrock overlays, 3D subsurface visualization (Three.js/CesiumJS), project management with status tracking and filtering, subsurface statistics tools, an uncertainty visualization tool using a "light and shadow" metaphor to estimate data confidence, an SQL-based query builder with stored query support, and role-based access control.

The pilot effort was also extended to demonstrate construction data management capabilities in coordination with District 6, incorporating subsurface data and grouting records from an ongoing karst remediation project using low-mobility grouting. This demonstration illustrates how the GDMS can link geotechnical investigation data with construction records within a single platform. The resulting system, which remains available for evaluation, is described in the following sections.

4.2 System Architecture

Back-end Framework

The GDMS pilot is built on Django 5.1 with a PostgreSQL/PostGIS spatial database back-end enabling efficient geospatial data storage, complex spatial queries, and integration with other GIS systems. Core Python libraries include NumPy (numerical computing), Pandas (data manipulation), and Gunicorn (production WSGI server).

4.3 Pilot Data Model

The GDMS pilot data model represents geotechnical information through a geospatial and project-based framework:

- **GDMS_Points:** Central model for boring locations with spatial coordinates, elevation, source information, unique identifier, and metadata (rock depth, completion date, etc.)
- **SPT_Blow:** Standard Penetration Test data linked to GDMS_Points, including sample depths, blow counts, and penetration measurements
- **SPT_Lith:** Lithology (soil/rock type) information with depth intervals and descriptive information
- **SPT_Lab:** Laboratory test results with test identifiers, types, units, and test-specific data

- **Project:** Organizes GDMS_Points into logical groupings with project name, location, status, and date information

The data model includes foreign keys, many-to-many relationships, and validation constraints. The production GDMS can expand to include construction data (e.g., grouting volumes, subsurface treatments), foundation elements, geophysical surveys, ground modification records, deep foundation data, maintenance data, and other geospatially referenced objects.

4.4 Front-end Technologies

The GDMS pilot employs Leaflet.js for interactive 2D mapping (with extensions for drawing, measurement, geocoding, and Esri GIS integration), Three.js visualization of subsurface boring data and soil layers, and HTML5/CSS3 with Bootstrap for responsive design.

4.5 Key Features of the Pilot

Key features of the GDMS incorporated into the pilot include:

- **Interactive Web Map:** Visualization of boring locations, integration with PennDOT roadway data and PA DCNR bedrock geology, drawing tools for area selection, search functionality, and detailed data popups.
- **Project Management:** Organization of borings into project groupings with status tracking, metadata, and filtering.
- **Analysis Tools:** Subsurface statistics (soil properties within selected areas), uncertainty visualization (“light and shadow” metaphor for test coverage and data confidence), Three.js-powered 3D views, and SQL-based query tools with stored query support.
- **Data Integration:** Standardized data formats consistent with the DIGGS schema, support for multiple sources including geophysical surveys, ground modification records, and deep foundation data.
- **User Access Control:** Django-based authentication with secure login, group-based permissions, granular access control, and role-based administrative access

4.6 Interaction with the GDMS

The pilot tool is available for testing at PennDOT Geotechnical Data Management System (GDMS) (174.138.85.143) through April 2026, after which it will be archived and provided to PennDOT. The following screen captures illustrate the pilot capabilities at the time of this report.

The Opening Screen is the landing page for the GDMS provides a summary of the purpose and goals of the tool along with a disclaimer that this is not to be used for engineering purposes (Figure 7). A menu bar appears in the top header bar.

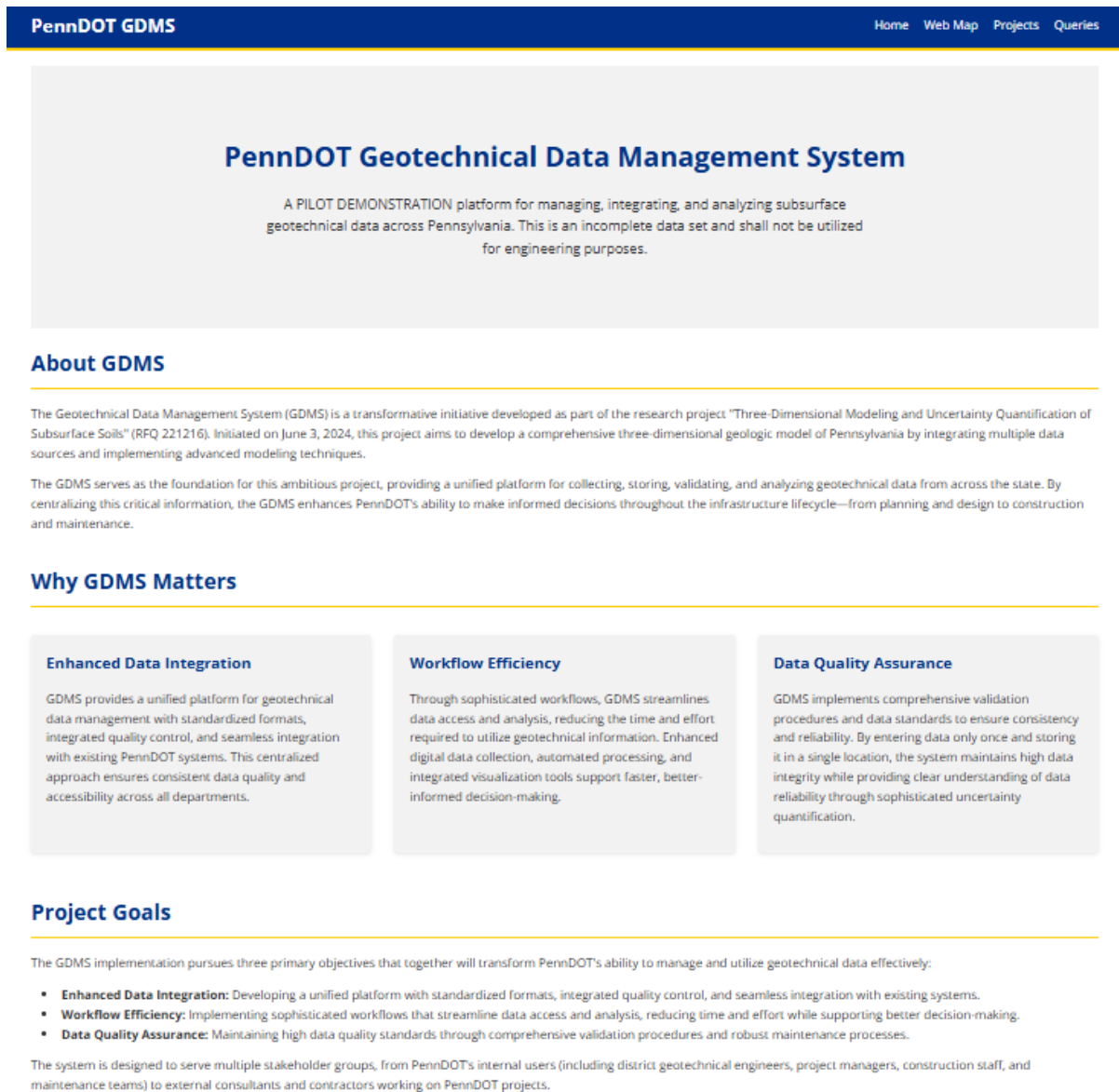


Figure 7: Opening screen as viewed from a web browser.

A Web Map feature provides a GIS based location visualization tool and a variety of base map options (note some of the base map options are scale dependent). This portal provides drawing and selection tools as well as the ability to quickly view individual point data (Figure 8).

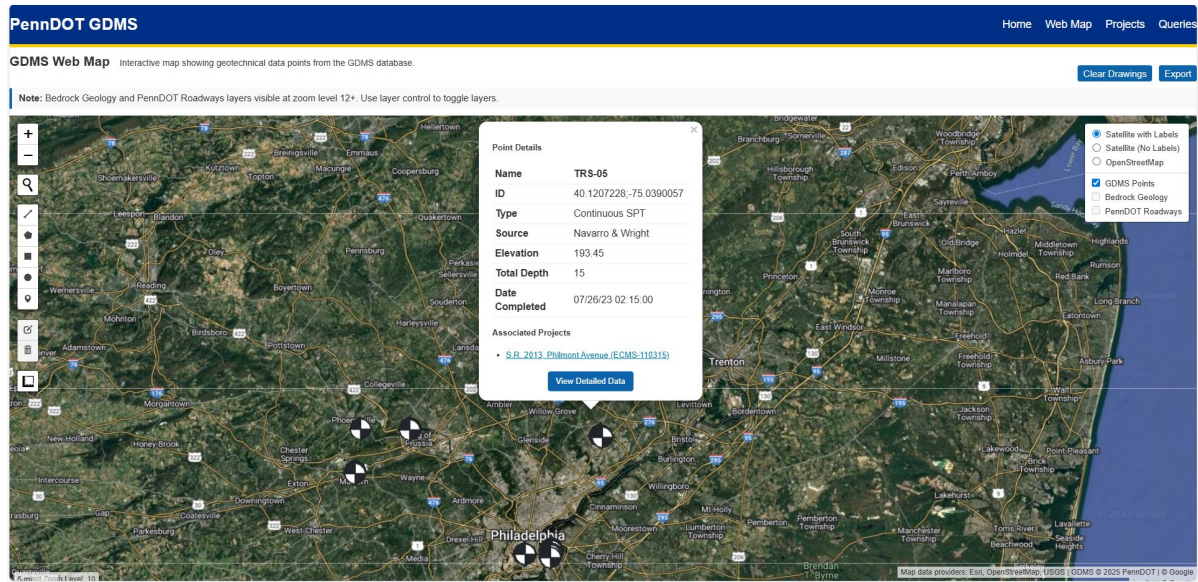


Figure 8: Borehole selection tool example for a single borehole.

This can be further extended to visualize data in a 3D format by selecting points in a box and right clicking on the box (Figure 9) or as expanded data (Figure 10).



Figure 9: Borehole selection tool example for a desired region.



Figure 10: Visualization tool for changes in features with depth.

The Projects page provides a simple means to view all data related to an individual project (Figure 11). Further the user can select the project and access advanced tools for reviewing the data statistics, uncertainty evaluation and future data conversion opportunities (Figure 12).

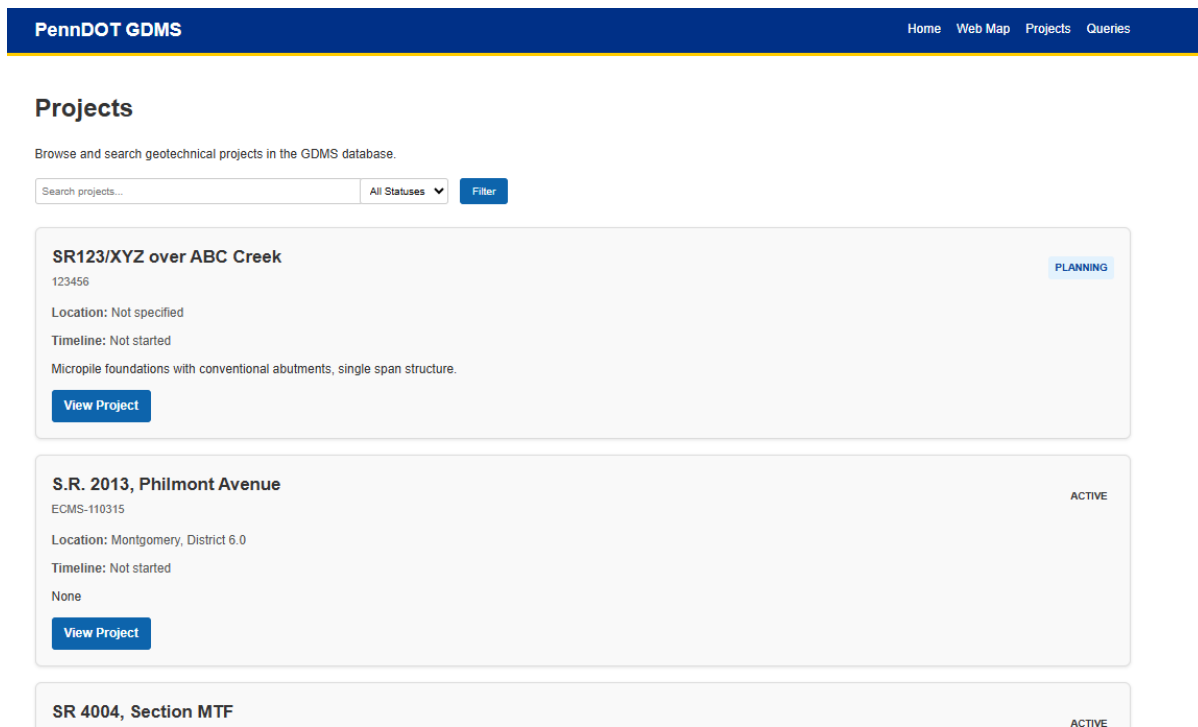


Figure 11: Project-selection tool.

S.R. 2013, Philmont Avenue

ECMS-110315

ACTIVE

Project Details

Location: Montgomery, District 6.0

Start Date: Not specified

End Date: Not specified

Status: Active

Description

No description provided.

Project Points Map

View in 3D

? Estimate Uncertainty

Subsurface Statistics



Project Points (14)

 Upload DIGGS File

 Open Project Files

Download All Project Data (DIGGS)

ID	Type	Source	Elevation	Total Depth	Date Completed	Actions
HV-01	Continuous SPT	Navarro & Wright	128.30	15.00	07/18/23 10:00:00	🔍 📄
HV-02	Continuous SPT	Navarro & Wright	131.77	15.00	07/17/23 15:45:00	🔍 📄
HV-03	Continuous SPT	Navarro & Wright	129.70	15.00	07/17/23 11:45:00	🔍 📄

Figure 12: Selected project summary page and borehole map.

The Queries page provides examples of standard queries that can be set up and used repeatedly as well as a means to build queries to search the GDMS for specific information (Figure 13).

PennDOT GDMS

HomeWeb MapProjectsQueries

Predefined Queries

Run predefined queries to analyze geotechnical data in the GDMS database.

All Types

Filter

View Execution History

Add Example Queries

All Points

BASIC QUERY

Get all geotechnical points in the database

Run Query

Average Elevation by Type

STATISTICAL QUERY

Calculate average elevation grouped by point type

Run Query

Points by Date Range

ADVANCED QUERY

Get points completed within a specific date range

Run Query

Points by Elevation Range

ADVANCED QUERY

Get points within a specific elevation range

Run Query

Points by Geographic Area

SPATIAL QUERY

Figure 13: Query page with common queries pre-established.

A demonstration of the abilities of analytics applied to the database are demonstrated on the Subsurface Statistics page which incorporates a tool to visually represent subsurface borings data (Figure 14). Example tools include SPT N by elevation and variability by elevation for user selected individual borings or region. This tool can be readily expanded to display geotechnical data in a PennDOT defined format.

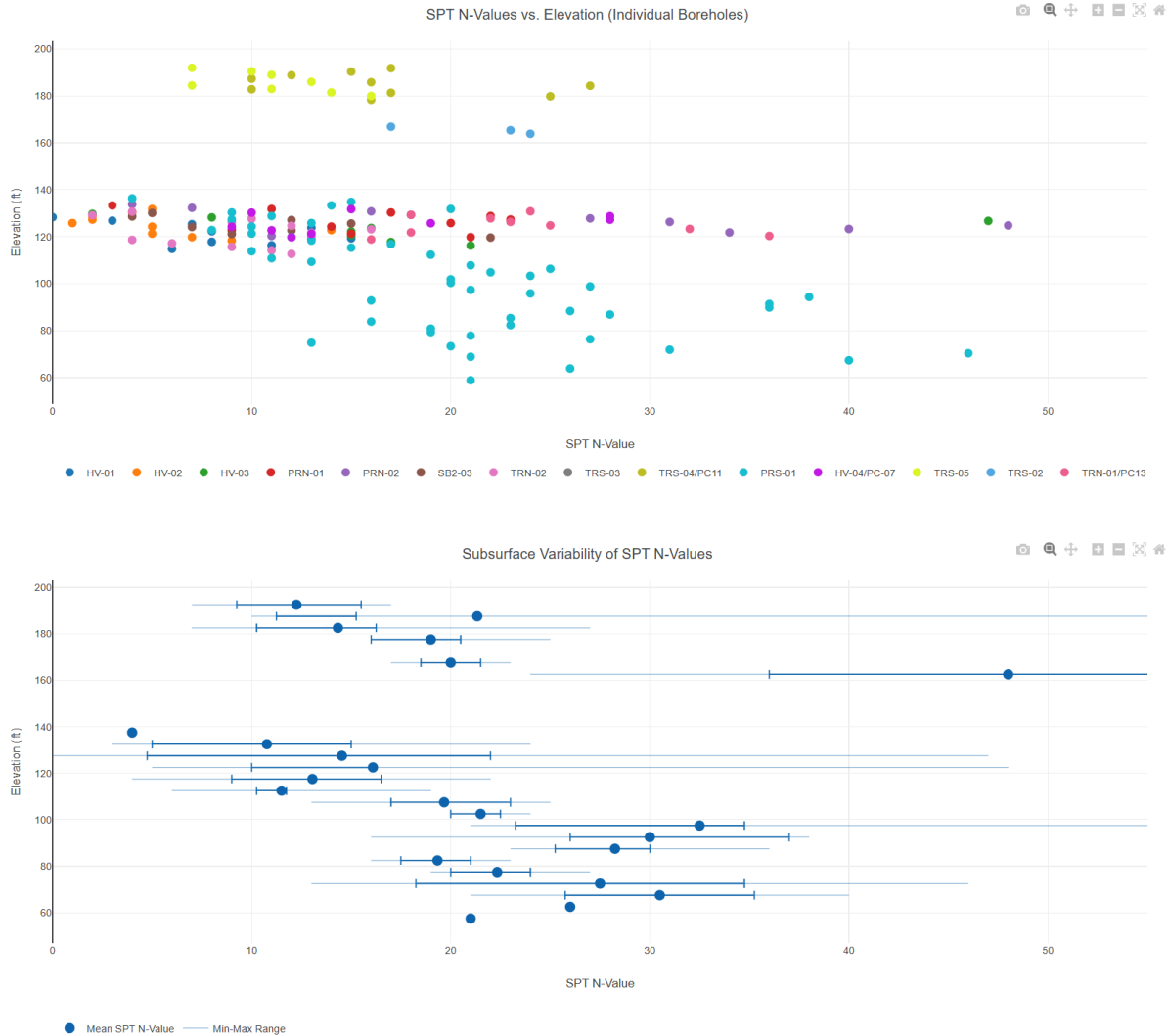


Figure 14: Subsurface statistics options including SPT-N values vs. elevation and subsurface variability vs. elevation.

An extension of the analytics includes the demonstration Geotechnical Uncertainty tool that provides estimated information on site knowledge, strength, modulus, classification, and groundwater for four in-situ testing methods based on user-defined testing method accuracy (Figure 15). By default, these tools use values defined by the research team based on professional experience. For a fully developed GDMS this tool can provide an insights into the quality and uncertainty of data within project limits with customization of general uncertainty and test accuracy based on established studies and PennDOT experience.

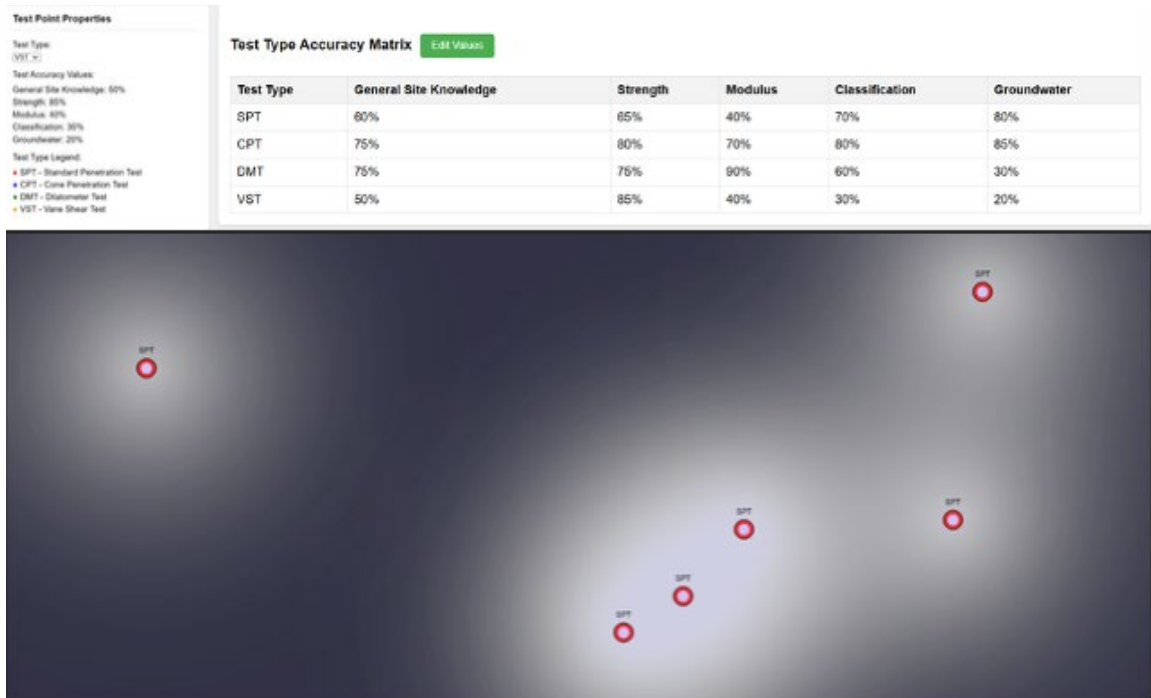


Figure 15: Geotechnical uncertainty visualization tool using a “light and shadow” presentation method.

The pilot effort was extended to demonstrate integration with Construction Data Management. The research team has been working with District 6 on a karst site with sinkhole issues being remediated with low mobility grouting. As a demonstration, the subsurface data has been included in this GDMS pilot and a link to the visualization tools developed for this project outside of this research effort was provided in the GDMS as an example of future development potential leveraging the project, query, as well as 2D and 3D tools (Figure 16 through Figure 19).

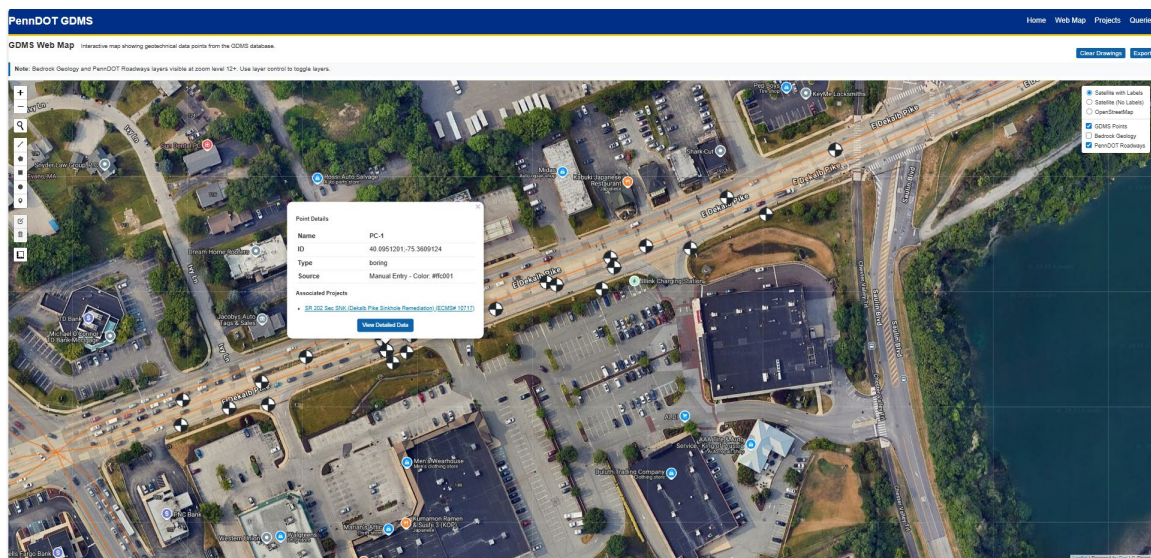


Figure 16: Karst site remediation demonstration providing a satellite overlay of the project site and borehole locations.

The construction information and more detailed subsurface data is accessible through links on the project page.

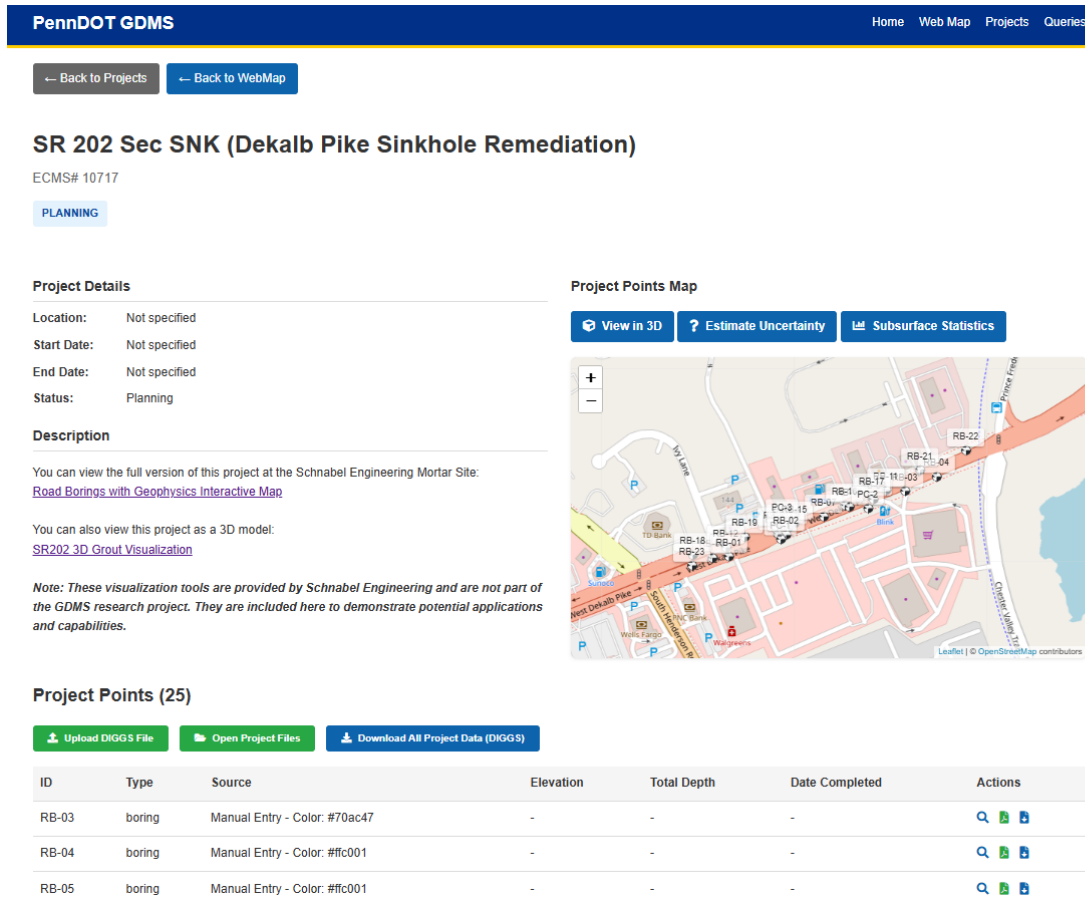


Figure 17: Karst site remediation demonstration showing records associated with each location ID (e.g., RB-03, RB-04) and object type (boring).

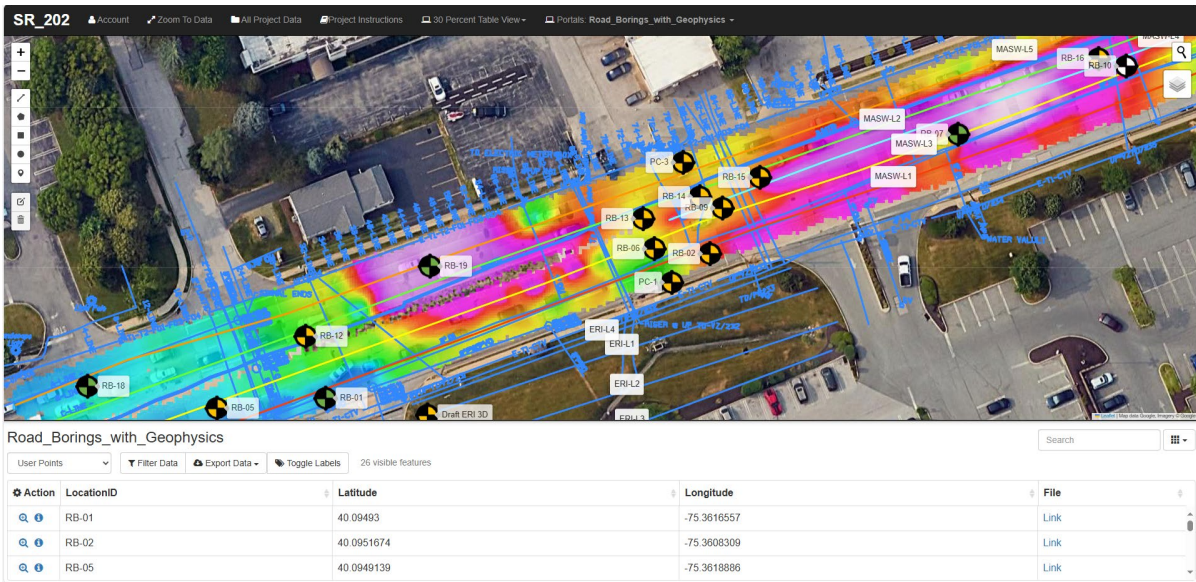


Figure 18: Karst site remediation demonstration showing a heat map of geophysical results along with boring locations.

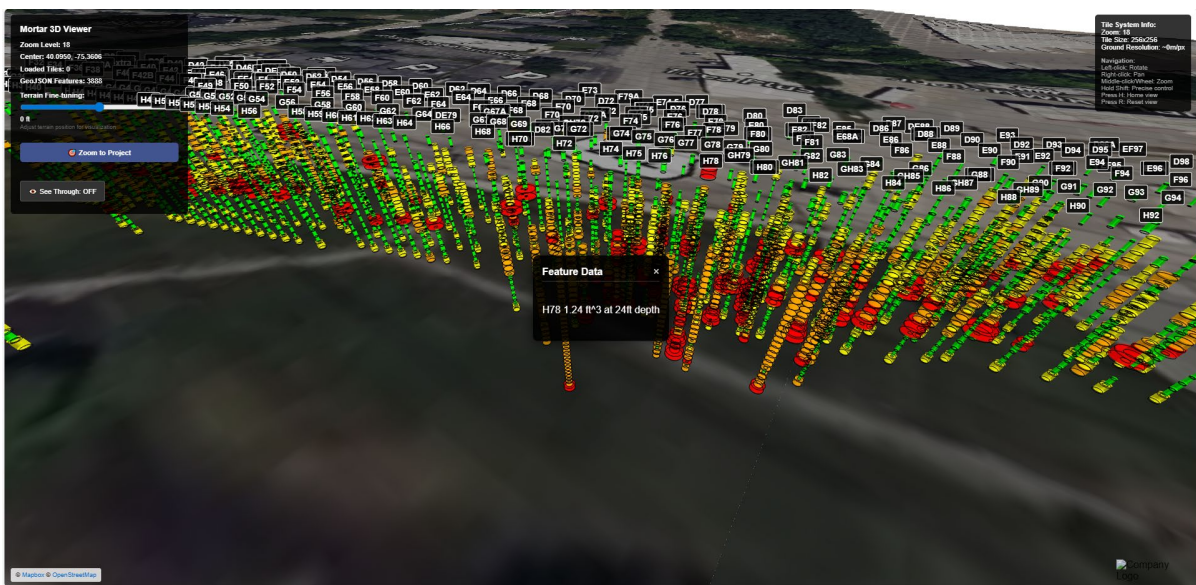


Figure 19: Karst site remediation demonstration showing a second visualization of grout volume at each boring with depth.

4.7 Task Conclusions

The pilot GDMS demonstrates the feasibility of compiling subsurface geologic and geotechnical data in a single platform with capabilities to link investigation, construction, maintenance, and management information. The pilot validated interactive 2D and 3D visualization, subsurface statistical analysis, uncertainty quantification, SQL-based querying, and role-based access control. The District 6 karst remediation demonstration further illustrated how construction records can be integrated alongside geotechnical investigation data within the same system. These results confirm the core technical approach and provide a foundation for the recommendations developed in Task 4.

5. TASK 4. DEVELOP RECOMMENDATIONS

5.1 Task Introduction

Task 4 developed guidance for transitioning the GDMS from a proof-of-concept demonstration to an operational, enterprise-ready system. This task documents pilot feedback from PennDOT engineers and IT staff, addressing concerns raised regarding data interpolation, workflow consistency, QA/QC governance, and accessibility compliance. A stakeholder identification framework covering fourteen internal groups and thirteen external categories was developed alongside structured engagement strategies.

Task 4 also presents a four-stage expansion strategy progressing from standardization and data governance through district-level pilots, statewide enterprise integration, and digital twin optimization. Step-by-step workflow guidance is provided for integrating three categories of data into the GDMS: new field boring records, historical boring records, and historical construction as-builts. A maturity model and supporting use cases illustrate practical application across implementation tiers. Key elements are summarized below; the detailed Task 4 interim report is attached as Appendix B.

5.2 Establishing GDMS Value

The success of GDMS development depends on demonstrating value to stakeholders. Task 4 developed a stakeholder identification framework covering fourteen internal groups (executive leadership, geotechnical engineering division, district personnel, design engineers, construction operations, maintenance and asset managers, surveyors, spatial analysts, IT management, traffic and safety, right of way, data governors, and training) and thirteen external categories (consulting firms, contractors, utilities, municipalities, state and federal agencies, academia, industry associations, data providers, professional societies, funding partners, public stakeholders, and software vendors).

An example internal stakeholder registry is provided in Table 1, mapping each group's interests, decision authority, communication channels, and interaction frequency. Engagement strategies include stakeholder workshops, influence-commitment matrices, and structured feedback cycles to further expand such a registry. Figure 20 demonstrates a commitment vs influence graph providing examples of how different stakeholders may be engaged in the process.

To recognize the return on investment, we consider evaluation of direct financial savings (reduced duplicate investigations, consolidated data management), productivity gains (faster data access, standardized workflows), risk reduction (uncertainty visualization, QA/QC workflows reducing design errors), and strategic benefits (improved coordination, research partnerships, standards adoption). Given the results of the pilot study, significant enhancements in all categories can be realized and should be measured as the GDMS is implemented.

Table 1. Example internal stakeholder registry for GDMS development.

Stakeholder	Interest / Responsibilities	Decision Authority	Communication Channels	Interaction Frequency
Executive Leadership	Set strategy; approve funding; sponsor program.	Final approval on budget, policy, and statewide rollout.	Executive briefings; steering committee meetings; email.	Quarterly or as-needed
Geotechnical Engineering Division	Define technical requirements; validate geotechnical content; lead QA.	Approve technical standards and acceptance criteria.	Technical workshops; review meetings; GDMS notifications.	Weekly to monthly
District Personnel and Project Managers	Submit exploration requests; implement local workflows.	Local approval of project scopes, schedules, and priorities.	District meetings; project portals; GDMS tickets.	Weekly to monthly
Design and Review Engineers	Use GDMS outputs for design and foundation decisions and QA/QC; perform peer review.	Approve design use of GDMS data; request additional investigations.	Project meetings; design reviews; GDMS queries.	Project-driven
Construction and Field Operations	Provide as-built and construction data; perform field collection.	Approve field data submissions and schedules.	Field briefings; mobile app updates; daily logs.	Daily to weekly during active work
Maintenance and Asset Management	Inventory geotechnical assets; prioritize inspections; plan interventions.	Prioritize maintenance actions and allocate funds.	Asset dashboards; reports; coordination meetings.	Monthly
Surveyors and Mappers	Provide control, LiDAR, and topographic data; ensure georeferencing and spatial databases.	Approve spatial control and coordinate transforms.	GIS services and workshops; technical calls.	Project-driven
Spatial Analysts	Manage spatial layers, mapping services, and visualization components.	Approve data integration and representation methods.	Data transfers; API documentation; GIS services and workshops; technical calls.	Weekly to monthly
IT and Enterprise Management	Host systems; enforce security; manage API integrations; system performance; alignment with PennDOT IT policies.	Approve infrastructure, security posture, and integrations. Access control and feature implementation.	Architecture reviews; control and access; ticketing system.	Weekly to monthly
Traffic and Safety Operations	Coordinate traffic control for field operations.	Approve traffic control plans and schedules.	Traffic coordination meetings; permit systems.	Project-driven
Right of Way and Permitting	Manage property access and utility clearances.	Approve access permits and utility coordination.	Permit workflows; coordination meetings; GDMS links.	Project-driven
Data Governors and Records Management	Define data standards, metadata, retention, and DIGGS compliance.	Enforce data policies and retention rules.	Governance meetings; policy documents; compliance reports.	Monthly
Training and Workforce Development	Develop role-based training and adoption materials.	Approve training curriculum and delivery approach.	Training sessions; LMS; webinars.	Quarterly or as-needed

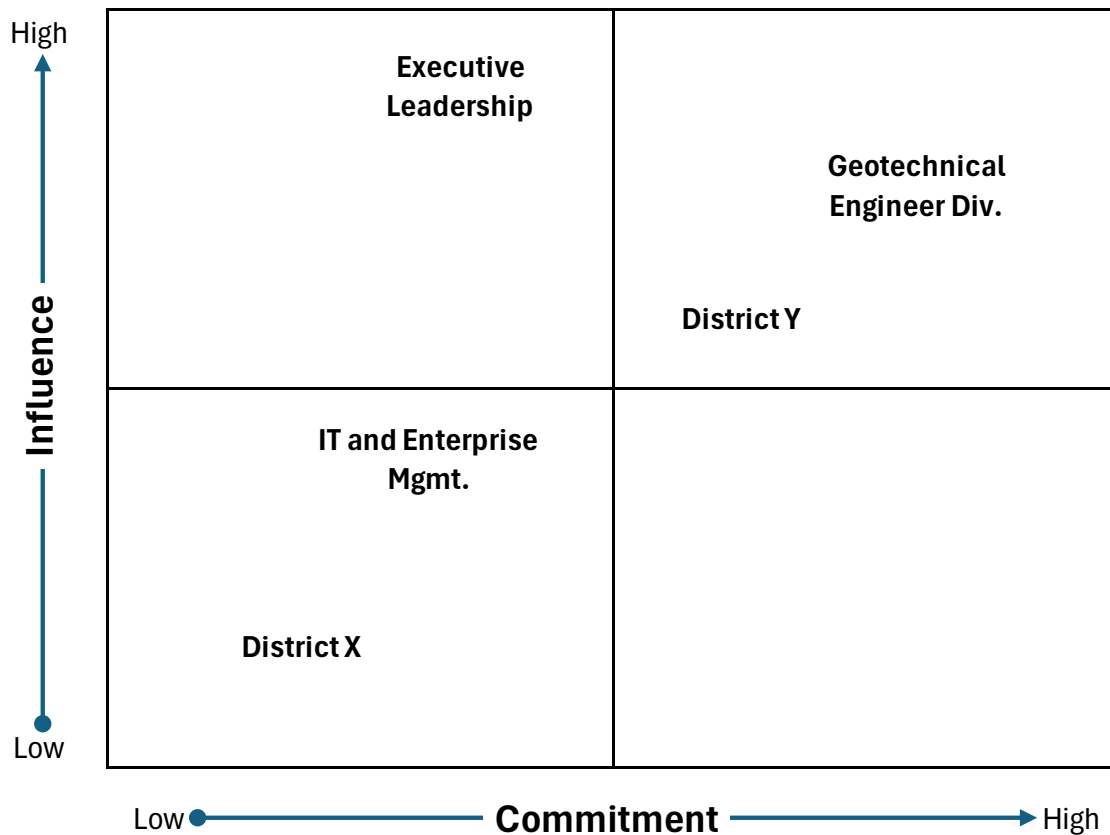


Figure 20: Example influence-commitment matrix for GDMS adoption.

5.3 Pilot Response and Expansion

Response to the GDMS pilot has been generally positive. Across several months of discussions, district engineers and project staff have consistently indicated that the demo showcased strong visualization capabilities, particularly in 3D modeling of subsurface data. The ability to generate subsurface statistics, visualize borehole variability, and display grout models in 3D was seen as a useful and powerful tool for exploration and communication of project progression and results. At the same time, participants emphasized that while 3D visualization adds significant value, traditional two-dimensional cross-sections remain essential for engineering practice including design. Several concerns were also raised about the risks of interpolating data for visualization, underscoring the need to balance accessibility and tool development with technical flexibility and liability considerations.

Database integration emerged as a recurring consideration with the project team highlighting the advantages of PostgreSQL with PostGIS extensions over ArcGIS Online, citing greater flexibility, fewer constraints, and stronger performance for complex queries. Queries within the GDMS are SQL-based, offering transparency and flexibility, although caution must be practiced against open query builders due to security concerns. This approach was preferred by PennDOT staff, based on the importance of maintaining control over database structures and ensuring compatibility with existing enterprise systems. The GDMS's ability to connect to multiple databases was seen as a

key strength, enabling broader integration across PennDOT's digital delivery ecosystem as detailed in the Task 2 report.

Feedback also focused on data incorporation with districts expressing interest in integrating historic datasets, geotechnical asset management (GAM) and GAM-type data, and ongoing project files (e.g., borings and CAD surfaces) into the GDMS. JSON and geoJSON formats were identified as promising pathways for long-term storage and display, while automated methods for georeferencing logs and reports are being developed to streamline the inclusion of legacy data. These capabilities were viewed as critical for demonstrating GDMS's value in real-world projects.

Workflow and quality assurance considerations have also been extensively discussed across the project team and PennDOT. PennDOT emphasized the importance of retaining control over QA/QC processes, similar to the draft-to-final workflows already in place for the demonstration grouting data.

Integration with PennDOT's broader digital delivery plan remains a priority and meetings with IT staff confirmed that neither security nor ESRI tools are expected to pose major hurdles. Preliminary delivery methods should be confirmed for compatibility with PennDOT's current geospatial data systems including those such as PennDOT Multi-modal Project Management System (MPMS) presented as OneMap. District-level pilot rollouts (possibly in Districts 4 and 11) should be considered to further advance and demonstrate the GDMS's potential benefits before statewide implementation.

Based on the pilot response, there is significant potential in expansion, development and statewide implementation of the GDMS. This includes functionality to include automated project tracking, corridor risk flagging, and improved visibility of geotechnical data across districts and the state as a whole. The pilot has established a strong foundation, and feedback reflects both satisfaction with its current capabilities and a roadmap for enhancements that will make GDMS a valuable part of PennDOT's geotechnical data management and digital delivery strategy.

However, future expansion requires deliberate attention to several factors; the system must continue to balance visualization with engineering utility, maintain database flexibility, and establish robust workflows for new and historic and statewide data integration. QA/QC governance, licensing, and accessibility must be formalized to ensure sustainable adoption. Functional enhancements (e.g., project tracking, corridor risk flagging, and a centralized hub for geotechnical standards and training) may further increase the GDMS's value. Equally important, PennDOT must invest in training and change management to encourage district-level adoption and institutionalize GDMS within agency workflows.

5.4 Staged Expansion

Following additional GDMS development detailed in Task 2, a four-stage expansion strategy is recommended. While the pilot GDMS is technically validated, it is not yet integrated into PennDOT's production IT environment and requires formalized governance, data standards, and organizational alignment before broader deployment. Direct transition from a proof-of-concept to statewide implementation risks inconsistent workflows across districts, unresolved QA/QC and accessibility requirements, and insufficient stakeholder engagement. The four-stage approach recommended includes:

- **Stage 1: Standardization and Data Governance (Near-term, 1-2 years):** Implement DIGGS-compliant schemas, formalize QA/QC workflows, resolve accessibility and licensing requirements (Section 508, Survey123 seats), create governance framework, and revise applicable PennDOT publications including Publication 222.
- **Stage 2: District-Level Pilots (Near-term, 1-2 years):** Implement the GDMS system through limited channels (Districts 4 and 11), test automated georeferencing for historic datasets and construction history, demonstrate 3D visualization alongside 2D cross-sections for engineering use. Begin evaluation of additional data set integration such as geophysical and foundation construction records.
- **Stage 3: Statewide Expansion and Enterprise Integration (Mid-term, 2-5 years):** Embed GDMS into PennDOT's enterprise systems, expand to statewide geotechnical assets (slopes, embankments, retaining walls, hazardous sites), deploy automated project tracking, develop centralized geotechnical standards hub, implement training and change management programs, and finalize consultant contract language for digital data delivery. Integrate with PennDOT's GAM program or the inclusion of GAM-type data for risk-based scoring.
- **Stage 4: Long-Term Optimization and Digital Twin Integration (5+ years):** Position GDMS within a digital twin environment supporting predictive analytics, BIM interoperability, and resilience planning. If digital twin integration becomes feasible earlier, it should follow a parallel pilot-to-statewide progression.

This progression manages implementation risk incrementally and aligns with FHWA Transportation Asset Management Plan guidance. The staged approach also provides PennDOT a sustainable development path during periods of constrained staffing and uncertain long-term funding.

5.5 Data Collection and Integration

The development schedule outlined above provides a framework for GDMS expansion; however, the success of each stage depends on the availability of quality data to populate and validate the system. Task 4 developed workflows for collecting and integrating new field records, historical boring records, and historical construction as-builts into the GDMS. These workflows are summarized below, visualized in Figure 21 through Figure 23 and detailed in Appendix B.

Each workflow uses a consistent three-color coding system to distinguish the nature of each step: blue indicates data collection and processing activities, yellow indicates review or QA/QC steps, and green denotes outputs and integration activities. These workflows are iterative and should be revised as needed to meet PennDOT's evolving requirements as the GDMS expands in scope, data volume, and user base.

1. **New Field (Boring) Records:** This considers an eight-step workflow from pre-planning through archiving: project setup with existing data review, digital form preparation (DIGGS-compliant), real-time field capture with GPS, inline validation and QA/QC at the source, synchronization with GDMS through RESTful APIs, supervisor review and approval, integration with external reference layers, and metadata management for provenance tracking.

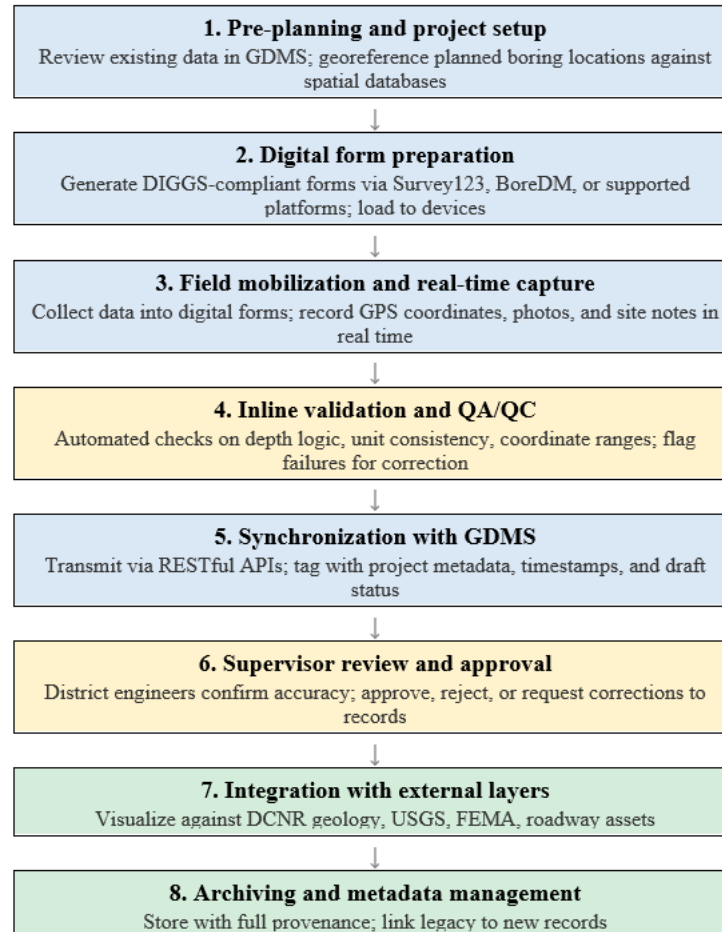


Figure 21: New record integration workflow.

2. **Historical Boring Records:** A nine-step workflow was developed for converting legacy data (PDFs, gINT files, spreadsheets) into validated, spatially indexed, DIGGS-aligned records: source discovery, document intake and cataloging, automated extraction and normalization, spatial validation and enrichment, expert QA review with provenance retention, DIGGS-aligned data modeling, database loading with spatial indexing, verification and publication, and ongoing maintenance.

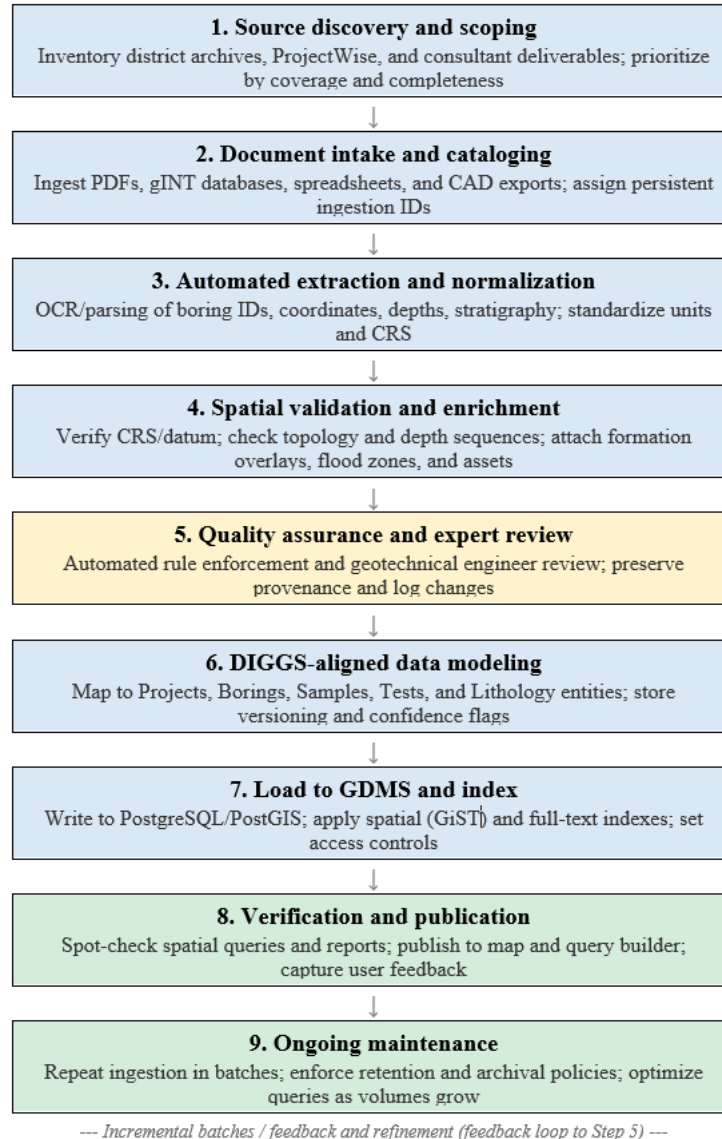


Figure 22: Historical boring record integration workflow.

3. **Historical Construction Records:** Converting as-built records such as pile driving logs or grouting records, and instrumentation data into structured GDMS entities includes: source identification, intake and metadata capture, geospatial conversion from CAD, relational linkage to geotechnical entities, QA/QC and approval, data modeling, database loading, verification and visualization, integration with monitoring and maintenance, and continuous improvement.

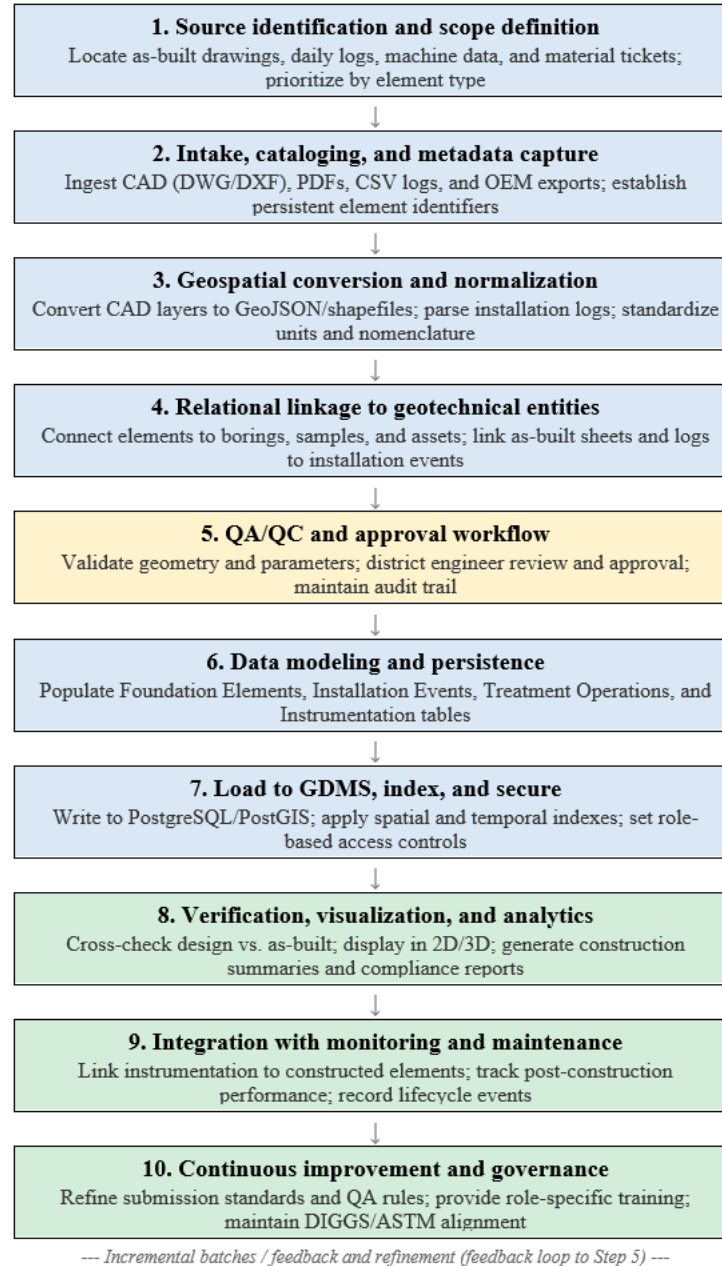


Figure 23: Historical construction record integration workflow.

5.6 GDMS Tools and Utilization

GDMS utilization is framed across three maturity levels:

- **Simple (Archive and Retrieval):** A centralized repository for historical boring logs and reports. Engineers can query by project, location, or date. Visualization is limited to 2D maps.
- **Moderate (Structured Data and Visualization):** A schema-compliant, validated system with digitized field collection, automated QA/QC, interactive maps, and 3D subsurface models overlaying geology, flood zones, and assets.
- **Extensive (Lifecycle Digital Twin):** Fully integrated platform linking investigations, construction records, and maintenance data with geostatistical analysis, uncertainty quantification, and real-time monitoring.

Example use cases are further documented in the Task 4 report, including: locating historical borings for a culvert project (simple), planning a bridge replacement with integrated QA/QC (moderate), karst remediation with lifecycle integration (complex), and highway slope validation with risk screening (moderate-complex).

Future GDMS capabilities identified in Task 4 include laboratory information management, equipment inventory and calibration tracking, real-time sensor integration, hazard early-warning systems, field productivity analysis, cross-disciplinary data correlation, site condition validation with automated consistency checks and historical benchmarking, and sustainability and resilience metrics.

5.7 Task Conclusions

The GDMS represents a significant step toward modernizing subsurface data practices and advancing risk-informed decision-making in geotechnical engineering and infrastructure management. The pilot has validated technical feasibility and earned positive feedback from PennDOT engineers and IT staff. Moving forward, success depends on disciplined governance, sustained stakeholder engagement, and careful management of the transition from pilot to production. The staged expansion strategy developed in this task provides a disciplined path from standardization through district pilots to statewide integration, with detailed workflows and stakeholder engagement strategies to support each stage.

6. SUMMARY OF PROJECT CONCLUSIONS AND RECOMMENDATIONS

6.1 Conclusions

This project established the technical feasibility, organizational framework, and expansion strategy for a centralized Geotechnical Data Management System (GDMS) capable of transforming how PennDOT collects, manages, and leverages subsurface data statewide. The Task 1 literature review and agency interviews confirmed that geotechnical data management is a recognized challenge across state DOTs, that standardized schemas such as DIGGS are gaining adoption, and that centralized governance and digital workflows are prerequisites for long-term success. Task 2 translated those findings into a three-tier, open-source system architecture built on Django and PostgreSQL/PostGIS with DIGGS-compliant data structures, vendor-neutral API ingestion, and integration points with PennDOT enterprise systems and authoritative external data sources. The architecture balances immediate data management needs with a path toward advanced analytical capabilities.

The Task 3 pilot validated core technical approaches including interactive 2D and 3D visualization, subsurface statistical analysis, uncertainty quantification, construction data management, and role-based access control. The pilot demonstrated that the proposed architecture can function as a unified geotechnical data platform at a statewide scale. Task 4 confirmed the value of the approach through structured feedback from PennDOT engineers and IT staff, who responded positively to the pilot's visualization capabilities and database flexibility while reinforcing the need for formalized QA/QC governance, accessibility compliance, and standardized workflows. Task 4 also produced a four-stage expansion strategy, detailed data integration workflows, and a stakeholder engagement framework to guide the transition from pilot to production.

Taken together, these results demonstrate that a centralized GDMS is both technically viable and organizationally achievable. The system provides a foundational platform for geotechnical data stewardship, supporting PennDOT's broader goals for digital delivery and infrastructure data management.

6.2 Recommendations

Based on the findings across all four tasks, the following recommendations are provided to guide the transition from pilot to production and the long-term institutionalization of the GDMS.

1. **Adopt a staged expansion strategy.** The project team recommends a four-stage approach to GDMS deployment.

- 1.1. Stage 1 (one to two years) should establish DIGGS-compliant schemas, formalize QA/QC workflows, resolve accessibility and licensing requirements, and revise applicable PennDOT publications including Publication 222. This stage should address the following:

- 1.1.1 Coordination with external software vendors (e.g., boring log, laboratory testing, and instrumentation data management platforms), including establishing automated data transfer workflows for future boring and laboratory data.

- 1.1.2 Legacy data migration into the GDMS.

- 1.1.3 Development of as-built pile table data fields within the GDMS (currently maintained in gINT), including field entry (e.g., Survey123) integration for data collection.
 - 1.1.4 As-needed support to complete the pilot digital representation of structure boring data for Digital Delivery.
- 1.1. Stage 2 (one to two years) should expand through district-level pilots in at least two districts or regions representing diverse geology and GDMS-targeted project types, testing automated georeferencing and GAM-type data integration.
- 1.2. Stage 3 (two to five years) should embed the GDMS into enterprise systems for statewide asset coverage, automated project tracking, and centralized geotechnical standards.
- 1.3. Stage 4 (five or more years) should position the GDMS within a digital twin environment supporting predictive analytics, BIM interoperability, and resilience planning. If digital twin integration becomes feasible earlier than anticipated, it should follow a parallel pilot-to-statewide progression.
- 2. **Formalize data governance and QA/QC protocols.** A formalized governance framework is essential before expanding beyond the pilot. This should include draft-to-final approval workflows for boring logs, construction records, and field data; automated validation rules enforcing required fields, range checks, unit consistency, and schema compliance; expert review processes for stratigraphy, coordinate confidence, and special conditions; and provenance retention preserving original values alongside normalized data. Clear roles and responsibilities for data stewardship should be established across districts and central office.
- 3. **Maintain vendor-neutral, standards-based architecture.** The GDMS should continue to accept data from any DIGGS-compliant or standards-compliant field platform, including BoreDM, RSLog, OpenGround, and Survey123, through its RESTful API. This vendor-neutral approach maximizes flexibility for PennDOT's internal staff, consultants, and contractors while ensuring data quality through schema enforcement at the point of ingestion.
- 4. **Prioritize legacy data conversion.** PennDOT's extensive archives of historical boring logs, laboratory reports, and construction records represent a significant asset. A systematic program should be established to digitize, parse, validate, and ingest these records using a combination of OCR, gINT/CSV parsing, expert review, and DIGGS-aligned data modeling. Historical construction records, including as-builts, pile driving logs, and grouting records, could also be linked relationally to corresponding geotechnical data.
- 5. **Integrate with PennDOT enterprise systems.** The GDMS should be connected to existing enterprise platforms including ECMS, OneMap, BMS, and RMS. IT staff confirmed during pilot demonstrations that no fundamental barriers to enterprise integration exist. Integration should proceed incrementally, beginning with systems that offer the greatest immediate value such as OneMap for spatial data and ECMS for project lifecycle data.
- 6. **Resolve accessibility and licensing requirements.** Section 508 accessibility compliance and licensing constraints, including Survey123 seat policies, should be addressed in the near term to ensure equitable access for all GDMS users. These requirements should be resolved prior to or concurrently with Stage 1 deployment.

7. **Develop training and change management programs.** Sustained adoption requires investment in training programs tailored to different user groups, from district geotechnical engineers and project managers to consultants and contractors. Change management strategies should be developed to institutionalize GDMS workflows across districts, supported by revised publications and standard operating procedures.
8. **Expand stakeholder engagement.** The broad stakeholder landscape identified in Task 4 should be actively engaged through the structured strategies developed in this project, including stakeholder registries, influence–commitment matrices, and role-specific communication plans. Early and continued engagement is critical to sustain interest, investment, and adoption.
9. **Pursue expanded GDMS capabilities over time.** As the system matures, PennDOT should explore expanded use cases including laboratory information management, equipment inventory and calibration tracking, real-time sensor integration, hazard early-warning systems, field productivity analysis, cross-disciplinary data correlation, site condition validation with automated consistency checks and historical benchmarking, and sustainability and resilience metrics. These capabilities align with emerging best practices in digital delivery and geotechnical asset management.
10. **Establish clear maintenance and sustainability responsibilities.** Long-term success requires defined responsibilities for system maintenance, database administration, performance tuning, schema updates, and incremental data ingestion. Staffing and funding plans should be developed to sustain GDMS operations beyond the initial deployment period, particularly given current conditions of reduced staffing and uncertain long-term funding.

The staged expansion strategy provides a disciplined path from standardization through district pilots, statewide integration, and digital twin capabilities, delivering both immediate benefits and long-term value. By following these recommendations, PennDOT can institutionalize the GDMS as a durable platform that enhances engineering rigor, supports operational efficiency, and strengthens long-term resilience of Pennsylvania’s transportation infrastructure.

REFERENCES

- Bachus, R., Machairas, N., & Cadden, A. (2020). Standardizing Geodata Transfer and Storage. *Deep Foundations*, 97–100.
- Barnes, R. (2003). *Variogram Tutorial*. Golden, Colorado: Golden Software, Inc.
- Bentley Systems. (2021). *gINT Connect Edition*. Retrieved from <https://www.bentley.com/wp-content/uploads/PFB-gINT-LTR-EN-LR.pdf>
- Burns, S. E., Torres, J., Jung, Y., Liu, C., & Tsai, Y. (2023). *Towards the Implementation of a Geotechnical Asset Management Program in the State of Georgia*. Atlanta, Georgia: Georgia Department of Transportation. Retrieved from <https://rosap.nhl.bts.gov/view/dot/67653>
- Columbia University. (2024). Kriging Interpolation. Retrieved from Columbia University Mailman School of Public Health: <https://www.publichealth.columbia.edu/research/population-health-methods/kriging-interpolation>
- Cressie, N. A. C. (1993). *Statistics for Spatial Data* (Rev. ed.). New York: John Wiley & Sons.
- Curran, J., & Hammah, R. (2006). Seven Lessons of Geomechanics Software Development. *The 41st U.S. Symposium on Rock Mechanics (USRMS)*. Golden, CO: American Rock Mechanics Association.
- Dasenbrock, D., Gray, J., Rivers, B., Ortiz, T., Kuhne, J., & Pelham, K. (2021). Looking to the Sky for Geotechnical Data. *Public Roads*, Spring 2021. Retrieved from <https://highways.dot.gov/public-roads/spring-2021/04>
- Data Interchange for Geotechnical and Geoenvironmental Specialists (DIGGS). (n.d.). *DIGGSML Official Website*. Retrieved from <https://diggsml.org/>
- De Rienzo, F., Oreste, P., & Pelizza, S. (2008). Subsurface Geological-Geotechnical Modelling to Sustain Underground Civil Planning. *Engineering Geology*, 96(3–4), 187–204.
- Deutsch, C. V., & Journel, A. G. (1998). *GSLIB: Geostatistical Software Library and User's Guide* (2nd ed.). New York: Oxford University Press.
- Esri Inc. (2024a). *ArcGIS Pro Help*. Retrieved from <https://pro.arcgis.com/en/pro-app/latest/help/main/welcome-to-the-arcgis-pro-app-help.htm>
- Esri Inc. (2024b). How Kriging Works. *ArcGIS Pro Tool Reference*. Retrieved from <https://pro.arcgis.com/en/pro-app/latest/tool-reference/3d-analyst/how-kriging-works.htm>
- Federal Highway Administration. (2013, April). *Implementation of Geotechnical Asset Management at State DOTs* (FHWA-HRT-13-031). Washington, D.C.: Federal Highway Administration.
- Federal Highway Administration. (2025, April 8). *Asset Management*. Retrieved from U.S. Department of Transportation, Federal Highway Administration: <https://www.fhwa.dot.gov/asset/guidance.cfm>

- Fish, M., & Lane, R. (2012). *GIS and the New Hampshire Rock Cut Management System* (Report FHWA-NH-RD-12323V). Concord: New Hampshire Department of Transportation. Retrieved from http://www.nh.gov/dot/org/projectdevelopment/materials/research/projects/documents/12323v_report.pdf
- Goovaerts, P. (1997). *Geostatistics for Natural Resources Evaluation*. Oxford University Press.
- Gransberg, D. D., & Loulakis, M. C. (2012). *NCHRP Synthesis 429: Geotechnical Information Practices in Design-Build Projects*. National Cooperative Highway Research Program, Transportation Research Board. Washington, D.C.: National Academies Press.
- Harris, C. R., Millman, K. J., van der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., Oliphant, T. E. (2020). Array Programming with NumPy. *Nature*, 585(7825), 357–362.
- Holovaty, A., & Kaplan-Moss, J. (2009). *The Definitive Guide to Django: Web Development Done Right*. Apress.
- Hunter, J. D. (2007). Matplotlib: A 2D Graphics Environment. *Computing in Science & Engineering*, 9(3), 90–95.
- Jebelli, H., Asadi, S., Mutis, I., Lui, R., & Cheng, J. (2024). *Digital Twins in Construction and the Built Environment*. Reston, Virginia: American Society of Civil Engineers. doi: <https://dx.doi.org/10.1061/9780784485606>
- Jordahl, K., den Bossche, J. V., Fleischmann, M., Wasserman, J., Gerard, J., Tratner, J., & Cochran, M. (2020). *GeoPandas: Python Tools for Geographic Data*. GitHub.
- Khan, M. S., Park, J., & Seo, J. (2021). Geotechnical Property Modeling and Construction Safety Zoning Based on GIS and BIM Integration. *Applied Sciences*, 11(9), 4004.
- Kringe, K., & Chatterjee, S. (2020). Uncertainty Quantification of Structural and Geotechnical Parameter by Geostatistical Simulations Applied to a Stability Analysis Case Study with Limited Exploration Data. *International Journal of Rock Mechanics and Mining Sciences*, 125, 104157.
- Li, Z. (2021). An Enhanced Dual IDW Method for High-Quality Geospatial Interpolation. *Scientific Reports*, 9903.
- Matheron, G. (1963). Principles of Geostatistics. *Economic Geology*, 58(8), 1246–1266.
- Mayne, P. W., Christopher, B. R., & DeJong, J. (2002). *Subsurface Investigations – Geotechnical Site Characterization: Reference Manual*. National Highway Institute. Washington, D.C.: Federal Highway Administration. Retrieved from <https://rosap.nhl.bts.gov/view/dot/41566>
- McKinney, W. (2010). Data Structures for Statistical Computing in Python. *Proceedings of the 9th Python in Science Conference*, 51–56.
- National Academies of Sciences, Engineering, and Medicine. (2024). *Consensus Study Report: Foundational Research Gaps and Future Directions for Digital Twins*. Washington, D.C.: National Academies Press. doi: 10.17226/26894

- New York Department of Transportation. (2025). *Digital Delivery – Interim Guidance Document*. New York Department of Transportation. Retrieved from <https://www.dot.ny.gov/divisions/engineering/design/dqab/hdm/hdm-repository/Digital%20Delivery%20Guidance.pdf>
- Ohio Department of Transportation. (n.d.). Transportation Information Mapping System (TIMS). Retrieved from https://gis.dot.state.oh.us/tims_classic
- Ohio Department of Transportation. (2005). *Geotechnical Data Management System (GDMS): Case Study 1*. Ohio Department of Transportation. Retrieved from https://www.tsaadvet.com/CaseStudies/1_CaseStudies_StateAgency.html
- Olsen, M. J., Raugust, J. D., & Roe, G. V. (2013). *NCHRP Synthesis 446: Use of Advanced Geospatial Data, Tools, Technologies, and Information in Department of Transportation Projects*. National Cooperative Highway Research Program, Transportation Research Board. Washington, D.C.: National Academies Press.
- Pavlis, T. L., & Mason, R. (2017). New World of 3D Geologic Mapping. *GSA Today*, 27(8), GSATG313A. Retrieved from <https://rock.geosociety.org/net/gsatoday/science/G313A/GSATG313A.pdf>
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... Duchesnay, E. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, 12, 2825–2830.
- Pyrzcz, M. J., & Deutsch, C. V. (2014). *Geostatistical Reservoir Modeling*. New York: Oxford University Press.
- Remy, N., Boucher, A., & Wu, J. (2009). *Applied Geostatistics with SGeMS: A User's Guide*. Stanford: Cambridge University Press.
- Rose, G., Kirk, P., Gibbons, C., & Lander, A. (2018). Three Dimensional Geological Models in Ground Engineering: When to Use, How to Build and Review, Benefits and Potential Pitfalls. *Australian Geomechanics*, 53(3), 79–88.
- Shepard, D. (1968). A Two-Dimensional Interpolation Function for Irregularly-Spaced Data. *Proceedings of the 1968 ACM National Conference*, 517–524.
- Shishlov, A. S., Costin, A. M., & Davidson, M. T. (2023). Integration of Building Information Modeling Interoperability into Nonlinear Finite Element Analysis of Bridge Substructures. *Transportation Research Record: Journal of the Transportation Research Board*, 9–21. doi: 10.1177/03611981231160172
- Supergeo. (2017). *Spatial Statistical Analyst Extension Help*. Supergeo Technologies Inc. Retrieved from https://www.supergeotek.com/Spatial_Statistical_ENG_HTML/isotropy_and_anisotropy.htm
- Svensson, M., & Friberg, O. (2024). Digital Geotechnical Information Management in a Life Cycle Perspective. *Proceedings of the 18th European Conference on Soil Mechanics and Geotechnical Engineering (ECSMGE 2024)* (pp. 871–874). Lisbon, Portugal: Taylor & Francis.

- Thibaut, R., & Vandekerckhove, G. (2021). *pysgems – Use SGeMS (Stanford Geostatistical Modeling Software) within Python*. Retrieved from <https://github.com/robinthibaut/pysgems>
- Turner, L. L., Saito, T., & Grimes, P. (2008). *End-User Interest in Geotechnical Data Management Systems* (Report No. CA08-0957). Sacramento, California: California Department of Transportation. Retrieved from <https://dot.ca.gov/-/media/dot-media/programs/research-innovation-system-information/documents/07-0057-a11y.pdf>
- Transportation Research Board. (2025, November 3). *NCHRP Synthesis 20-05/Topic 56-23*. Retrieved from <https://apps.trb.org/cmsfeed/TRBNetProjectDisplay.asp?ProjectID=5880>
- Unkefer, D., Saad, T., Mavandadi, S., Machairas, N., Collins, B., Hiel, L., ... Jang, D. (2023). The Present and Future of 3D and 4D Geotechnical Data in BIM and Digital Twins. *2023 GI Web Conference*. American Society of Civil Engineers. Available from <https://www.asce.org/education-and-events/explore-education/on-demand-webinars/the-present-and-future-of-3d-and-4d-geotechnical-data-in-bim-and-digital-twins/>
- Vessely, M., Robert, W., Richrath, S., Schaefer, V. R., Smadi, O., Loehr, E., & Boeckmann, A. (2019). *NCHRP Research Report 903: Geotechnical Asset Management for Transportation Agencies*. National Cooperative Highway Research Program, Transportation Research Board. Washington, D.C.: National Academies Press.
- Wang, H., Wang, X., & Liang, R. (2020). *Study of AI Based Methods for Characterization of Geotechnical Site Investigation Data*. Dayton, Ohio: University of Dayton. Prepared for the Ohio Department of Transportation and the U.S. Department of Transportation, Federal Highway Administration.
- Waskom, M. L. (2021). seaborn: statistical data visualization. *Journal of Open Source Software*, 6(60), 3021. doi: 10.21105/joss.03021
- Webster, R., & Oliver, M. A. (2007). *Geostatistics for Environmental Scientists* (2nd ed.). Chichester, England: John Wiley & Sons.

APPENDICES

APPENDIX A. INTERIM REPORT 1

Three-Dimensional Modeling and Uncertainty Quantification of Subsurface Soils

September 5, 2024

RFQ 221216

Purchase Order (PO) 4300805489

Task 1: Literature Review
Interim Report

Prepared for:
Pennsylvania Department of Transportation

Prepared by:
Applied Research Associates Inc. (ARA Inc.)

TABLE OF CONTENTS

1. INTRODUCTION.....	1
2. PROJECT GOAL	2
3. LITERATURE REVIEW	3
3.1 LITERATURE SUMMARY	3
GEOLOGIC AND GEOTECHNICAL DATA STORAGE SYSTEMS	3
GEOTECHNICAL DATA MANAGEMENT SYSTEMS ACROSS DOTS.....	4
4. AGENCY INTERVIEWS.....	7
4.1 CALTRANS MEETING SUMMARY (JULY 24, 2024)	7
4.2 PENNDOT MEETING SUMMARY (JULY 23, 2024)	7
4.3 LADOTD MEETING SUMMARY (JULY 22, 2024)	8
4.4 OHIO DOT MEETING SUMMARY (JULY 30, 2024)	9
4.5 KEY TAKEAWAYS ACROSS ALL MEETINGS	9
5. SOFTWARE TOOLS AND PLATFORMS	10
5.1 GEOTECHNICAL MODELING SOFTWARE.....	10
5.2 WEB APPLICATION DEVELOPMENT	10
5.3 VIRTUAL REALITY (VR) IMPLEMENTATION.....	11
5.4 GEOSTATISTICAL ANALYSIS AND MODELING SOFTWARE.....	12
6. DATA INTEGRATION, STANDARDIZATION AND GOVERNANCE	15
6.1 DATA INTEGRATION AND STANDARDIZATION	15
6.2 EXISTING DATA SOURCES AND DATA GOVERNANCE.....	16
7. DATA INTERPRETATION, UNCERTAINTY QUANTIFICATION AND DATA ANALYSIS	18
7.1 DATA INTERPRETATION USING PYTHON.....	18
7.2 UNCERTAINTY QUANTIFICATION AND GEOSTATISTICAL ANALYSIS	18
7.3 CASE STUDIES AND APPLICATIONS.....	20
8. CONCLUSION AND PROJECT BENEFITS	22
9. LITERATURE REVIEW REFERENCES	24
APPENDIX A -AGENCY INTERVIEW QUESTIONS ON 3D GEOTECHNICAL MODELING AND ADVANCED DATA INTEGRATION	27
APPENDIX B – SGEMS AND ARCGIS PRO DETAILED COMPARISON	30
APPENDIX C - DETAILED TECHNICAL CONCEPTS AND ANALYSIS	32

1. INTRODUCTION

Research project **Three-Dimensional Modeling and Uncertainty Quantification of Subsurface Soils** was initiated on June 3, 2024, to investigate the opportunities and challenges with development of a three-dimensional geologic model of Pennsylvania. Following the kickoff meeting held on May 15, 2024, clarification of this effort discussed the need to enhance existing geologic mapping, incorporate other subsurface geotechnical data held within PennDOT files as well as other potentially available data sets that could enhance the model. The overall purpose of this tool will be to improve the agency's understanding of available data as it will apply to future project needs. No software or tool development will be produced in this research; however, pilot demonstrations may be developed for proof of concept if needed.

This initial report was developed as a summary of our literature search related to this topic. Throughout this project, we aim to provide context to the following points of interest:

- Who is the audience and the user profile?
- What is the level of detail?
- How much internal vs external data content is desired?
- How do you envision this tool being used?
- How extensive do you want analytics to be integrated?
- What level of maintenance and support will be available?

Following PennDOT's review of this document, we suggest scheduling a meeting to discuss of topics of interest for further refinement.

2. PROJECT GOAL

This project's goal is to provide guidance on the development of a three-dimensional (3D) geologic model for subsurface conditions across Pennsylvania. We understand the primary goal will be to enhance and improve the agencies workflow related to geologic and geotechnical issues as well as provide a tool to consultants working for the agency. It is likely that this tool (or a portion of it) would also be available to other agencies as well as the public in general. This tool may serve as a data management tool becoming a key component of future geotechnical design and construction projects recognizing that we must go beyond the current state of practice to incorporate interoperability of data which will be a critical factor as PennDOT implements its digital delivery goals. A component of this effort will be to explore capabilities for data transfer such as Data Interchange for Geotechnical and Geoenvironmental Services (DIGGS).

Consideration for planning of this model include:

- Integrate various data sources such as geologic mapping, associated relevant and available GIS layers, borehole data, seismic velocities, and digital elevation models.
- Explore fundamental guidance for future data collection and delivery to the Department so that it can add to the model value.
- Define workflows for data collection and integration as project data becomes available.
- Development of procedures for traditional test boring data conversion from digital files and PDF documents for incorporation into the model.
- Analytical tools for data validation, geotechnical parameter evaluation and probability distribution will also be explored for future implementation.
- Perform uncertainty quantification to characterize spatial variability of subsurface soils.
- Consideration of differences in the spatial and depth distribution of data.
- Consider the use of machine learning tools to interpret the data sets and provide guidance for users.

3. LITERATURE REVIEW

A literature search was performed to gather information relative to available 3D geologic models, experiences in developing such models and recommendations for future development methods and tools. Limited references were identified in these searches specific to geologic modeling. Agencies such as USGS and DCNR have been developing tools and models for various locals which have led to standardized schema formats for bedrock geologic representation. These efforts have added significantly to the understanding of geologic modeling standards and have laid the groundwork for more sophisticated models of the future.

To supplement this literature search, we conducted interviews with representatives from various Department of Transportation (DOT) agencies familiar with geotechnical data management and conversion efforts with their organizations.

The following sections summarize the findings from the literature review and the key insights from these interviews.

3.1 Literature Summary

Geologic and Geotechnical Data Storage Systems

- **United States Geological Survey (USGS):** The USGS has a number of online tools to store and display data and search for available literature. This includes the [National Geologic Map Database \(usgs.gov\)](#) and [EarthExplorer \(usgs.gov\)](#).
- **New Zealand Geotechnical Database (NZGD):** The NZGD is an online database that compiles geotechnical information from both historical and new sources. Developed on the success of the Canterbury Geotechnical Database (CGD) following the Christchurch earthquakes of 2010/2011, the NZGD provides efficient access to geotechnical information for routine functions, emergency response, disaster recovery, resiliency, and research. The platform operates on a voluntary basis, where users agree to upload data to enhance the tool in exchange for access. The overall costs of the system are covered by the Ministry of Business, Innovation, and Employment, and the Earthquake Commission. [NZGD Link](#)
- **Pennsylvania DCNR GEOlogic Data Explorer ([PaGEODE \(state.pa.us\)](#))** provides a GIS portal to access geology, quarry, well, and elevation data among other data sets.
- **Washington State Department of Natural Resources, Geologic Information Portal** puts complex geologic and hazards information into the hands of everyone. The data ranges from boreholes to geophysics, hazards, seismic classification as well as surface and subsurface geologic information. [Geologic Information Portal | WA - DNR](#)
- **Utah Geologic Survey [Interactive Geologic Map Portal \(utah.gov\)](#)** provides a 3D terrain map with geologic maps of varying scales draped over the terrain.

- **Looking to the Sky for Geotechnical Data (Dasenbrock et al., 2021):** This study explores the use of Unmanned Aerial Vehicles (UAVs) equipped with LiDAR and photogrammetry for rapid responses to embankment stability issues. While the integration of additional geotechnical data is discussed, the study does not provide specific models for such integration.
- **The New World of 3D Geologic Mapping (Pavlis and Mason, 2017):** This paper contrasts 2.5D imagery, like that found in Google Earth, with true 3D datasets such as LiDAR and Structure from Motion (SfM) photogrammetry. The authors highlight the high costs and complexities associated with LiDAR and advocate for the use of SfM technology to support geologic interpretations, especially when coupled with UAV systems.
- **Use of Advanced Geospatial Data, Tools, Technologies, and Information in Department of Transportation Projects (NCHRP Synthesis 446, Olsen et al., 2013):** This synthesis provides a comprehensive survey of U.S. transportation agencies and service providers regarding their use of advanced geospatial technologies. Key challenges identified include managing large data sets, the need for specialized technical expertise, rapid technological changes, and the lack of standardized workflows. However, opportunities such as improved productivity, safety benefits, cost savings, and better decision-making were also highlighted. The document does note the lack of geotechnical data integration into geospatial systems but mentions specific applications, like New Hampshire's hazard rating system for rock cuts along highways.
- **State Geologic Map Compilation Geodatabase of the Conterminous United States ([The State Geologic Map Compilation \(SGMC\) Geodatabase of the Conterminous United States - ScienceBase-Catalog](#))** The USGS compiled a seamless, spatial database of 48 State geologic maps that range from 1:50,000 to 1:1,000,000 scale. This is a national digital geologic map database developed by integrating state data and new data. Inconsistencies have been corrected, however discontinuities between state geologic interpretations have not been reconciled. This dataset can also be accessed through the ArcGIS Hub. This also searchable at <https://ngmdb.usgs.gov/mapview/>

Geotechnical Data Management Systems Across DOTs

Several state DOTs have implemented innovative systems for managing and storing geotechnical data. These systems, though varied in complexity and functionality, share the goal of preserving geotechnical history and supporting ongoing engineering projects:

- **Publicly Accessible ArcGIS Layers:**
 - **Florida DOT ([Public Soil Boring Viewer \(fdot.gov\)](#)), New Jersey DOT([Geotechnical Data Management System \(GDMS\) Viewer \(arcgis.com\)](#)), Washington State DOT ([Geologic Information Portal | WA - DNR](#)), and Vermont DOT (<https://geodata.vermont.gov/datasets/VTrans::vt-subsurface-boring-projects>)**: These agencies have developed ArcGIS layers that provide access to scanned PDF geotechnical boring logs. While these layers offer a visual and searchable interface,

they are limited in that the underlying geotechnical data within the PDFs is not directly searchable or filterable. [ArcGIS service Link Example](#)

- **Custom GIS Websites:**

- **California DOT (Caltrans):** Caltrans' GeoDOG platform allows users to explore and retrieve geotechnical information through an interactive GIS interface. [GeoDOG website](#).
- **Kentucky DOT:** Kentucky offers a custom website that allows users to search for borehole information efficiently. [Kentucky Example](#).
- **Missouri Geologic Survey** provides an interactive map of a variety of geologic information from formations to mines and other hazards. Well core information is included but there does not seem to be boring log information. [GeoSTRAT \(arcgis.com\)](#) More detailed well log information is available at [Well Drilling Logs \(arcgis.com\)](#).
- **British Geological Survey GeoIndex OnShore.** Is a component of their OpenGeoScience tools where a wide range of datasets are available including over a million borehole scans. [GeoIndex - British Geological Survey \(bgs.ac.uk\)](#)
- **Alaska (AKDOT) geotechnical asset management portal.** Provides a visual tracking of assets and geohazard events. This public facing portal includes ratings reports and photographs, but does not seem to include geotechnical subsurface data. [AKDOT&PF GAM Program Overview \(arcgis.com\)](#)
- **Conetec** provides a portal for their clients to view all of the in-situ data collected for a client within a gis portal.

- **Project-Specific Data Management:**

- **Vermont DOT:** Vermont integrates their ArcGIS layer of scanned PDF documents with a ProjectWise server, providing links to more detailed project-specific data.

- **Internal Geotechnical Data Management Tools:**

- **Maryland DOT:** Maryland DOT's Boring Request Tool stores raw geotechnical data in a database and dynamically generates gINT files. It is equipped with machine learning capabilities for predicting geotechnical conditions in untested areas and has been recognized for its efficiency and cost-saving benefits.

Software Tools for 3D Geologic Mapping

- **Tools and Techniques for 3D Geologic Mapping in ArcScene (Carroll, 2014):** This paper discusses the use of Esri's ArcScene for 3D geologic mapping at the Illinois State Geological Survey. While ArcScene offers valuable tools for 3D mapping, it also has

significant limitations, including the need for custom programming and potential memory resource constraints. The development of detailed 3D models requires complex procedures to compile and interpret datasets into fence or block models.

4. AGENCY INTERVIEWS

As part of our research efforts, we sought to understand similar geotechnical data management initiatives in other agencies. To facilitate these discussions, we developed a list of topics (see Appendix A) that guided our conversations with agency representatives known to have interest and active projects related to geotechnical data management. Additionally, we conducted interviews with PennDOT geotechnical personnel to gain a broader understanding of processes and procedures within the state. The questions used in these interviews were reviewed with the Research Department prior to initiating discussions.

4.1 Caltrans Meeting Summary (July 24, 2024)

Our research team met with Sharid Amiri and Jacqueline Martin from Caltrans to discuss various aspects of geotechnical data management. Sharid introduced the GeoDOG platform ([link](#)), a public repository for geotechnical data related to assets such as slopes, bridges, and walls. Currently, data on this platform is stored in PDF format, which may not fully align with conventional data definitions. Sharid also mentioned that Caltrans is transitioning from gINT to a new platform, potentially RocScience, indicating an ongoing modernization effort for data storage and accessibility.

Sharid highlighted Caltrans' innovation portal for submitting research and innovation ideas, focusing on data analytics, visualization, and potential future integration of AI and machine learning. In terms of GIS integration, Sharid and Jacqueline discussed ongoing efforts to integrate GIS into their system to improve data management and accessibility. Jacqueline provided insights into the Vision system, a SQL-based tool for managing project delivery aspects like schedule, cost, and resources.

Both Jacqueline and Sharid emphasized the need for improved data management practices, particularly in accessing and utilizing geotechnical data for projects. They discussed the use of GeoDOG for archiving final documents and the associated challenges. Sharid also mentioned a management directive on AI, focusing on safety, privacy, and alignment with state policy, indicating a strategic approach to AI adoption within Caltrans.

It should be noted that California was a part of the early DIGGS development and developed an introductory document to geotechnical data management in 2008, (Turner et al) [Microsoft Word - FINAL REPORT Task_0057_102108.doc \(ca.gov\)](#) which discussed the COSMOS/PEER-LL system used for exchange and dissemination of interagency data.

4.2 PennDOT Meeting Summary (July 23, 2024)

Our research team met with Sara Mullaney and Beverly Miller from PennDOT to discuss various aspects of geotechnical data management. Sara and Beverly discussed the current use of ECMS and the evaluation of BoreDM and Autodesk software for 3D subsurface modeling, expressing a desire to consolidate data into one system, Maintenance IQ, and to adopt DIGGS for standard

data formats. Other systems currently used included PennDOT OneMap, Bridge Management System (BMS), and Roadway Management System (RMS). It was noted that iPads are currently used in the field by construction personnel. Beverly and Sara indicated there was a Strike off Letter developed for AI and Machine learning. Sara emphasized the importance of seamless integration between software tools and PennDOT's data management system, highlighting the critical need for adopting DIGGS for data standardization.

Allen Cadden and Beverly discussed the challenges of collecting and managing data across districts, noting the centralized system for bridge database management and the varying levels of compliance with geotechnical data submission requirements among districts. It was noted that some Districts are beginning to develop inventories of geotechnical assets (e.g., slopes). Beverly and Sara would prefer a centralized data management system, potentially managed by say two staff in each District. Sara and Beverly also explored the potential of 3D modeling and visualization, including the use of Autodesk for bridge inspection and the exploration of virtual reality applications. They indicated that top or rock and geological features would be helpful to include in the model.

The team also discussed current data validation practices and the potential for predictive modeling and AI to improve data reliability. While discussing software mandates, Beverly favored standardizing on a single software for data collection, whereas Sara advocated for flexibility in using various tools. Additionally, Sara and Ross Cutts discussed the benefits of a statewide 3D geotechnical model for planning and hazard management, recognizing its value in refining project-specific data and improving the efficiency of data requests.

A separate conversation was held with Sara McInnes that helped shape the questions and general understanding of PennDOT needs. The information shared was similar to the above discussion.

4.3 LaDOTD Meeting Summary (July 22, 2024)

Our research team met with Jesse Rauser from LaDOTD to discuss various challenges and advancements in geotechnical data management. Jesse outlined the difficulties in assembling a diverse team of experts for data management and coordinating their skill sets. He detailed the extensive use of GIS for managing and visualizing data, including LiDAR and imagery data, which support asset location data for bridges and roadways. Jesse highlighted the agency's reliance on GIS for various tasks, particularly in maintaining road and bridge networks. Although current GIS usage is mostly 2D, there is significant interest in transitioning to 3D GIS applications.

Jesse confirmed the use of ArcGIS Pro, indicating a transition to more advanced GIS software. He explained that data collection is conducted at the project level, utilizing systems like Headlight for construction data and tablets for borehole logging. Jesse expressed concerns about the interoperability between different data systems and emphasized the need for tools that encourage user engagement. Ongoing efforts to convert historical geotechnical data for compatibility with current systems were also discussed. Jesse mentioned a decade-long effort to reformat data schemas to be DIGGS-friendly, aiming for standardized data formats.

Additionally, he acknowledged the challenges of requiring consultants to adopt new data formats like DIGGS.

4.4 Ohio DOT Meeting Summary (July 30, 2024)

Our research team met with Stephen Taliaferro and Christopher Merklin from the Ohio Department of Transportation (DOT) to discuss various aspects of geotechnical data management. Stephen elaborated on the use of the Collector tool for data collection across various geotechnical assets, emphasizing the standardization efforts and the desire for enhanced geohazard visibility. Christopher highlighted the integration of geotechnical data into a central system known as TIMS ([link](#)), which facilitates the association of geohazards with other assets like culverts for improved management and decision-making. Additionally, Christopher mentioned the existence of an office of data governance, underscoring its critical role in enterprise data-driven decision-making and the importance of uniform data collection and integration practices.

The discussion also explored new technologies such as LiDAR and InSAR for geohazard monitoring. Both Stephen and Christopher acknowledged the challenges and potential benefits of integrating these technologies into their current systems. Stephen expressed concerns about resource limitations affecting the reinspection of geotechnical assets, highlighting the need for efficient data management and potential savings through remote sensing technologies. Looking ahead, Christopher mentioned the potential of machine learning and predictive models to enhance data analysis and decision-making, indicating a cautious yet optimistic approach toward adopting these technologies within the agency.

4.5 Key Takeaways Across All Meetings

Throughout our discussions with Caltrans, PennDOT, LaDOTD, and Ohio DOT, several common themes emerged:

1. **Challenges in Inventorying Geotechnical Data and Assets:** All groups acknowledged the difficulties in maintaining comprehensive inventories of geotechnical data and assets. Agencies often have requirements to send geotechnical data to a central storage program, but compliance varies across districts. However, advancements in tools and technologies are gradually improving this process.
2. **Importance of Asset and Hazard Inventories for Funding:** Having a detailed inventory of assets and hazards has been shown to facilitate easier access to funding. This underscores the value of maintaining accurate and comprehensive geotechnical data records.
3. **Value of Uniform Data Structures:** All groups recognized the benefits of adopting standardized data structures such as DIGGS. Uniform data formats improve data integration, accessibility, and reliability across different systems and stakeholders.
4. **Integration of Advanced Technologies:** The integration of new technologies like LiDAR, InSAR, and AI holds promise for enhancing geotechnical data management. These technologies can improve data accuracy, monitoring capabilities, and predictive modeling.

5. **Data Governance and Centralization:** The establishment of centralized systems and offices of data governance is critical for effective data management. Uniform data collection and integration practices ensure consistency and support enterprise-level decision-making.
6. **Resource Limitations and Efficiency:** Concerns about resource limitations were common, particularly regarding the reinspection of geotechnical assets. Efficient data management practices and the use of remote sensing technologies can help mitigate these challenges.
7. **Future Directions:** There is cautious optimism about the adoption of machine learning and predictive models. These technologies have the potential to enhance data analysis and decision-making, although their implementation requires careful consideration of resources and capabilities.

5. SOFTWARE TOOLS AND PLATFORMS

This section reviews the software tools and platforms considered for developing the 3D geotechnical model for Pennsylvania subsurface soils. The tools are categorized based on their primary function, such as geotechnical modeling, web application development, virtual reality integration, and geostatistical analysis.

5.1 Geotechnical Modeling Software

These software packages are evaluated for their ability to create detailed 3D geological models and integrate various geotechnical data types. Key considerations included data integration capabilities, export options, and cost.

1. **Leapfrog Geo**
 - **Highlights:** Excellent for creating detailed 3D geological models; easily incorporates various data types.
 - **Export Options:** Various formats (.dxf, .obj, .fbx), convertible for web use.
2. **ArcGIS**
 - **HighlightsPros:** Strong GIS data integration; good for 3D subsurface modeling.
 - **Export Options:** Various 3D formats; good web integration.
3. **QGIS (with QGIS2threejs plugin)**
 - **Pros:** Open-source, free; plugin available for web-ready 3D visualizations.
 - **Export Options:** Direct export to web-usable formats.

5.2 Web Application Development

To create an interactive platform for visualizing the 3D geotechnical model, the following technologies were evaluated for their ability to deliver robust and scalable web-based solutions.

1. Django

- **Function:** Python web framework for backend development.
- **Use Case:** Suitable for developing the underlying data management system and APIs.

2. Leaflet.js

- **Function:** JavaScript library for interactive maps.
- **Use Case:** Ideal for integrating 2D geospatial data with basic interactive features.

3. Three.js

- **Function:** JavaScript library for 3D graphics in browsers.
- **Use Case:** Provides high-performance 3D visualization for complex geotechnical data.

4. Cesium

- **Function:** JavaScript library for 3D globes and maps.
- **Use Case:** Enables detailed 3D terrain visualization and exploration of subsurface conditions.

5. MapBox GL JS

- **Function:** Library for 3D terrain visualization.
- **Use Case:** Offers flexible and efficient base map interfaces for 2D and 3D geospatial data.

5.3 Virtual Reality (VR) Implementation

To enhance user interaction with the 3D geotechnical model, VR and AR platforms may be considered for their potential to provide immersive experiences and advanced data visualization.

- **Key Considerations for VR Platforms:**

- Standalone capability
- High-resolution displays
- Advanced motion tracking
- Comfortable design
- Wide developer support
- BIM and CAD compatibility
- Multi-user collaboration
- Measurement and annotation tools
- Cross-section views

While these tools are beyond the current scope of the project, exploring their future integration could add significant value, especially for training and stakeholder engagement.

5.4 Geostatistical Analysis and Modeling Software

This subsection compares the geostatistical analysis capabilities.

1. Stanford Geostatistical Modeling Software (SGeMS)
 - Pros: Open-source software, supports various kriging and simulation algorithms, multiple-point simulation, and Python integration.
 - Cons: Limited to specific file formats; requires technical expertise for effective use.
 - Link: [SGeMS User Guide](#)
 -
2. Geostatistical Analyst Extension of ArcGIS Pro
 - Pros: Supports a wide array of data formats, rich user-friendly built-in libraries for variogram modeling and interpolation.
 - Cons: Limited option for simulation algorithm.
 - Link: [Geostatistical Analyst in ArcGIS Pro](#)
3. Surfer
 - Pros: Simple and easy to learn interface, use of simple algorithms for geostatistical analysis, custom algorithms can be incorporated using Visual Basic for Applications (VBA).
 - Cons: Cannot consider multiple realizations through simulation, cannot import variogram models from other software.
 - Link: [Surfer](#)

A table of comparisons between the SGeMS and the ArcGIS Pro package is included as Appendix B.

5.5 Additional **Geotechnical Data Management Tools**

This subsection provides an overview of relevant tools for geotechnical data management and visualization, highlighting their unique features, strengths, and areas of application.

1. **gINT (Bentley Systems)**

- **Focus:** gINT is widely used for geotechnical data management, offering tools to automate repetitive tasks, reduce redundant data entry, and integrate with CAD, GIS, and other database applications. It supports extensive data reporting and visualization capabilities, making it a comprehensive solution for managing borehole data, lab test results, and other geotechnical information. However, gINT is expected to be phased out and replaced by Bentley's OpenGround platform soon, which offers enhanced cloud-based data management and collaboration features.
- **Link:** [Bently gINT](#)

2. **OpenGround (Bentley Systems)**

- **Focus:** OpenGround is an evolution from HoleBase developed in the UK by Keynetix. This has been ported to be a cloud-connected geotechnical data management and reporting system. It provides for collaboration from field data collection devices to desktop analysis and configurable log production and laboratory reporting. The data is accessible via API connections and can integrate with Excel and PowerBI.
- **Link:** [OpenGround](#)

3. **RocScience RSLog**

- **Focus:** RSLog is specialized in borehole log management and provides a streamlined interface for inputting, storing, and visualizing borehole data. One of its key strengths is its integration with other RocScience tools like Slide3 and RS2, which makes it particularly suitable for users already within the RocScience ecosystem. However, RSLog is more specialized and offers fewer options for complex data manipulation and reporting outside of borehole log management compared to gINT.
- **Link:** [RockScience RSLog](#)

4. Tablogs

- **Focus:** Tablogs is designed as a mobile-first solution for geotechnical data collection and management. Its primary strength lies in its field data collection capabilities, allowing users to enter borehole data directly from the field using tablets or smartphones. Tablogs emphasizes ease of use and accessibility, making it ideal for teams that need to capture data on the go. Unlike gINT, which offers extensive post-processing and analysis capabilities, Tablogs focuses more on real-time data entry and basic reporting, with less emphasis on detailed data integration with other systems like CAD or GIS.
- **Link:** [Tablogs](#)

5. BoreDM

- **Focus:** BoreDM specializes in managing and organizing large quantities of borehole and geotechnical data. It offers a centralized database for storing and querying data, making it an option for organizations that manage extensive data sets across multiple projects. While BoreDM provides robust data management capabilities, its integration with other software tools for analysis and modeling is less developed compared to gINT, which offers broader compatibility with a range of engineering applications.
- **Link:** [BoreDM](#)

6. EQuIS (Environmental Quality Information System)

- **Focus:** EQuIS is a comprehensive data management system designed for both environmental and geotechnical data. It manages complex data sets and offers powerful tools for data visualization, reporting, and regulatory compliance. EQuIS provides a more holistic approach to data management, accommodating not only geotechnical data but also environmental data, making it a versatile platform for multidisciplinary projects. However, its complexity and broader scope may make it less specialized for purely geotechnical projects than gINT or the other tools mentioned.
- **Link:** [EQuIS](#)

7. Other tools marketed to the industry include LogitEasy.com, Geosystems LogDraft, GAEA WinLoG, etc. which do not seem to have significant interest or following within the US DOT communities.

Summary:

While gINT (and its successor, OpenGround) offers comprehensive geotechnical data management with strong integration capabilities, tools like RSLog, Tablogs, BoreDM, and EQuIS cater to specific needs within the geotechnical and environmental data management space. The choice of any of the above-mentioned tools should be based on the specific requirements of the user.

6. DATA INTEGRATION, STANDARDIZATION AND GOVERNANCE

The development of a 3D geotechnical model for Pennsylvania subsurface soils requires robust data integration, adherence to standardized formats, and effective data governance. This section outlines the strategies for integrating historic and new data, adopting industry-standard formats, and leveraging existing data sources while implementing strong governance practices.

6.1 Data Integration and Standardization

The 3D geotechnical model will be built on a GIS system to provide necessary geolocated information. Numerous agencies and private sources offer significant GIS data layers to form the basis for this tool. The initial focus will be on integrating existing and new geotechnical exploration, laboratory tests, and asset records.

6.1.1 Integration of Historic Geotechnical Data

PennDOT's historic geotechnical data primarily exists as paper and PDF versions of boring logs and laboratory data, typically stored at the district level. Recent efforts to gather digital gINT files at the Central Office have shown variable success. The ECMS system prompts projects to upload data during closeout, but the lack of perceived IFC standards for geotechnical data has hindered policy development for data integration.

Leveraging public and private geotechnical data portals like Geosetta, Boreholes database by SAALG Geomechanics, and NZGD could enhance the integration of historic data. Collaborating with these platforms may allow for the incorporation of additional data elements as well as state and national geologic and geohazard mapping sources will contribute to a more comprehensive geotechnical model.

6.1.2 Adoption of Geotechnical Standards

To maximize the model's value and accuracy, it is crucial to adopt industry-standard data formats for future data collection and exchange. Currently, no US DOT agency has fully implemented a standardized geotechnical data management schema. However, ongoing research and data conversion efforts are paving the way for future adoption. Two notable standards are:

1. **DIGGS (Data Interchange for Geotechnical and Geoenvironmental Specialists):** Despite its slow evolution, DIGGS is gaining traction, with several consulting organizations and software vendors developing DIGGS-compliant tools.
2. **AGS (Association of Geotechnical & Geoenvironmental Specialists) Format:** Well-established in Europe and Asia, the AGS system provides another standardized approach to geotechnical data management.

Recent interoperability experiments between OGC and IFC have demonstrated the feasibility of integrating both DIGGS and AGS into IFC-compliant subsurface data models, suggesting promising avenues for future standardization efforts.

6.2 Existing Data Sources and Data Governance

A critical aspect of the 3D geotechnical model is the use of existing data sources and the implementation of robust data governance practices. By linking directly to authoritative data sources, the model can maintain data currency, reduce redundancy, and rely on vetted information. This strategy offers several key advantages, including up-to-date information, reduced duplication, efficient updates, and scalability.

6.2.1 Key Data Sources

The following data sources are readily available and can contribute to the comprehensiveness and accuracy of the 3D geotechnical model:

1. **Pennsylvania DOT Open Data**
 - URL: data-pennshare.opendata.arcgis.com
 - Content: Geological data, Private Property, Right of Way, and Structural geology.
2. **Pennsylvania DCNR Open Data**
 - URL: newdata-dcnr.opendata.arcgis.com
 - Content: Bedrock geology, surficial geology, and structural geology.
3. **Pennsylvania Department of Environmental Protection Open Data Portal**
 - URL: newdata-padep-1.opendata.arcgis.com
 - Content: Environmental data, including mining activities, oil and gas wells, and water resources.
4. **FEMA Flood Maps**
 - URL: hazards-fema.maps.arcgis.com
 - Content: Flood hazard data, including flood zones and floodplain boundaries.
5. **USGS GeoLog Locator**
 - URL: [USGS - GeoLog Locator](https://www.usgs.gov/geo-log-locator)
 - Content: Borehole geophysical logs and related data
6. AGI Interactive Database for Geologic Maps of the United States [MapView \(beta\) | NGMDB \(usgs.gov\)](https://ngmdb.usgs.gov)

6.2.2 Data Collection and Integration Tools

Integrating data collection tools such as Survey123 and ESRI Field Maps will further enhance the model by facilitating real-time data acquisition from field personnel. These tools allow for the collection and submission of geotechnical data, including photos, GPS coordinates, and descriptive information, directly from the field.

- **Survey123 Integration:**
 - Custom Forms: Tailored to PennDOT's data collection needs.
 - Data Submission: Real-time synchronization with the central database.
- **ESRI Field Maps Integration:**
 - Field Data Collection: Spatial data, geotechnical observations, and infrastructure locations.

- Offline Capabilities: Data collection in areas without network coverage, with automatic synchronization once reconnected.

6.2.3 Governance and Scalability

A key advantage of the above-mentioned data sources and tools is their accessibility through ESRI's REST services. This compatibility allows for seamless integration with tools such as Django and Leaflet applications. By leveraging these sources and tools, the 3D geotechnical model for Pennsylvania can become a living tool—always current, authoritative, and scalable for future data integration.

7. DATA INTERPRETATION, UNCERTAINTY QUANTIFICATION AND DATA ANALYSIS

Effective geotechnical modeling requires not only the integration of diverse data sets but also a robust approach to data interpretation and uncertainty quantification. This section outlines the methodologies, tools, and strategies recommended for comprehensive data analysis, highlighting the importance of geostatistical methods and the use of Python programming for performing these tasks.

7.1 Data Interpretation Using Python

Python is frequently cited in the literature for its robust ecosystem of libraries that facilitate data analysis, visualization, and geospatial processing. These capabilities make Python a prominent tool for geotechnical data interpretation and uncertainty quantification.

Key Advantages of Python

- **Data Manipulation and Analysis:** Libraries such as NumPy and Pandas are frequently cited for their efficiency in handling large geotechnical datasets, such as borehole logs and soil properties. These tools enable researchers to perform complex statistical analyses and data transformations critical for geotechnical design (Harris et al., 2020; McKinney, 2010).
- **Geospatial Processing:** GeoPandas extends Pandas' capabilities to geospatial data, facilitating spatial queries and visualizations. This is crucial in geotechnical engineering, where the spatial distribution of soil types and groundwater levels plays a significant role in risk assessment and site investigations (Jordahl et al., 2020).
- **Machine Learning:** Scikit-learn's comprehensive suite of machine learning algorithms is often applied to predictive modeling and pattern recognition in geotechnical contexts, such as predicting soil properties or classifying soil types (Pedregosa et al., 2011).
- **Visualization:** The use of visualization libraries such as Matplotlib, Seaborn, and Plotly is well-documented in the literature, allowing for the creation of detailed borehole logs, cross-sectional views, and 3D models. Visualization is essential for conveying complex subsurface conditions to stakeholders (Hunter, 2007; Waskom, 2021).
- **Integration:** Python's flexibility in integrating with various backend systems, such as Django, is highlighted as a key advantage in developing web-based geotechnical applications that streamline data analysis and decision-making processes (Holovaty & Kaplan-Moss, 2009).

These tools and methodologies have been foundational in recent geotechnical research, underscoring Python's versatility and effectiveness in geotechnical modeling.

7.2 Uncertainty Quantification and Geostatistical Analysis

Uncertainty quantification is essential in subsurface modeling due to the inherent variability of geological materials. As Curran and Hammah (2006) noted, "uncertainty is king; make room for it." Geological materials, formed under diverse thermo-physio-chemical conditions, have histories

that are often incomplete or unknown. Consequently, their properties can vary significantly with location. A geotechnical parameter predicted from a model based on a single observation may be inaccurate, making it crucial to account for geospatial uncertainty.

7.2.1 Variogram Modeling

- **Definition:** The variogram measures the relationship between two locations based on the variance of differences between them, providing insight into spatial continuity. (Barnes, 2003).
- **Importance:** Variograms reveal spatial patterns that are not evident in standard statistical summaries, allowing for more accurate predictions.
- **Components:** Key components include the nugget effect, sill, and range, each representing different aspects of spatial variability.
- **Modeling:** Various models, including spherical, exponential, and Gaussian models, are used to fit experimental data, guiding geostatistical interpolation and improving model accuracy. These models help in understanding the degree of spatial correlation, which is essential for making reliable predictions in areas where direct data may be sparse (Cressie, 1993).

7.2.2 Geostatistical Interpolation

- **Kriging:** A statistical interpolation method that uses variograms to account for spatial correlation between sample points, providing the best linear unbiased prediction of unknown values. (Matheron, 1963)
- **IDW (Inverse Distance Weighted):** This deterministic method assigns weights inversely proportional to the distance between sample points, providing a simpler but less statistically robust approach than kriging. It is often used when a quick estimation is needed, though it may not account for spatial correlation as effectively (Shepard, 1968).

7.2.2.1 Types of Kriging:

- **Ordinary Kriging:** Assumes constant mean and variance across the spatial field, making it suitable for datasets without apparent trends.
- **Universal Kriging:** Accounts for trends in the data by incorporating a drift function, useful in areas where soil properties change systematically with location.
- **Block Kriging:** Estimates average values over a block of grids, reducing prediction error and making it ideal for regional-scale assessments.
- **Cokriging:** Enhances accuracy using correlated variables, beneficial when multiple related geotechnical properties are measured.

7.2.2.2 Limitations of Kriging

- The accuracy of kriging heavily depends on the quality of the variogram model, and poor variogram fitting can lead to unreliable predictions. Additionally, in cases where sample data is sparse or highly variable, kriging may result in "over-smoothing," masking important local variations (Webster & Oliver, 2007).

7.2.2.3 Geostatistical Simulation

Geostatistical simulation offers an alternative to interpolation by generating multiple realizations that reflect the full range of possible outcomes. This approach is valuable for risk assessment and decision-making in environments with significant uncertainty (Goovaerts, 1997).

- Sequential Indicator Simulation: Used for categorical data, such as soil classification, to produce multiple realizations that honor the data's spatial distribution.
- Sequential Gaussian Simulation: Applied to continuous data, allowing for the creation of multiple equiprobable realizations that capture the uncertainty in the predictions.

Simulations provide not only a single value for unknown locations but also the uncertainty associated with the estimated value through the conditional cumulative distribution function (CCDF), making them a powerful tool for uncertainty quantification.

7.3 Case Studies and Applications

Several case studies demonstrate the practical application of these methodologies in real-world scenarios:

7.3.1 Case Study 1: 3D Geotechnical Modeling in Turin, Italy (Rienzo et al., 2008)

- Created a geological map using data from 380 boreholes covering approximately 50 km².
- Developed a 3D GIS model to manage spatial data for a geological Turin model.
- Utilized deterministic triangular network (TIN) method for spatial interpolation.
- Conducted uniaxial compression and shear tests on soil samples to examine geotechnical properties.
- Correlated geotechnical parameters with cementation maps.

7.3.2 Case Study 2: Integration of BIM and GIS for Geotechnical Zoning in Peshawar, Pakistan (Khan et al., 2021)

- Created a geodatabase incorporating borehole information for geostatistical interpolation through kriging.
- Produced digital maps illustrating geotechnical parameters like SPT-N values, plasticity index, and unconfined compressive strength.
- Employed Autodesk Civil 3D for BIM to handle coordinate systems, visualize borehole data, and create geotechnical surfaces and solids.

- Integrated BIM-GIS approach facilitated storage and management of geotechnical information in IFC format.

7.3.3 Case Study 3: AI-Based Site Characterization by the University of Dayton (Wang et al., 2020)

- Developed a methodology combining AI-based machine learning algorithms with other disciplinary principles to interpret geotechnical site investigation data.
- Used a machine learning module to analyze original site characterization data and deduce subsurface stratification models.
- Incorporated stochastic simulation techniques to estimate subsurface conditions at unexplored locations with uncertainty quantification.
- Demonstrated benefits of the methodology through three case studies.

These studies highlight the effectiveness of combining geostatistical analysis with advanced visualization and modeling techniques to enhance the reliability of geotechnical mode

Appendix C includes a more detailed discussion of some of these technical concepts and analysis methods.

8. CONCLUSION AND PROJECT BENEFITS

The proposed 3D geotechnical model for Pennsylvania subsurface soils is a pioneering initiative that can integrate cutting-edge technology with comprehensive data management. This project can significantly enhance PennDOT's capabilities in infrastructure planning, design, and management by offering an advanced tool for visualizing and analyzing subsurface conditions. The benefits of this project are far-reaching and include the following:

1. **Informed Decision Making:** The interactive 3D model will provide PennDOT with a detailed understanding of subsurface conditions, leading to more informed and confident decisions in infrastructure planning and design. This will help mitigate risks and optimize project costs by reducing the uncertainties typically associated with subsurface investigations.
2. **Streamlined Data Management:** By integrating diverse data sources—ranging from historic geotechnical records to new data collections—into a unified model, the project will streamline data management processes, making it easier to analyze, store, and retrieve critical geotechnical information.
3. **Enhanced Collaboration:** The web-based interface, along with potential VR/AR integration, will facilitate better communication and collaboration among various stakeholders, including engineers, geologists, policymakers, and the public. This improved collaboration will support more cohesive and effective project execution.
4. **Scalability and Flexibility:** The use of open-source technologies and a modular design approach ensures that the model can easily be expanded and adapted to accommodate future data sources, technological advancements, and evolving project needs.
5. **Educational and Training Value:** The 3D model will serve as an educational resource for training new professionals in geotechnical engineering and public infrastructure management. It will also enhance public understanding of subsurface conditions and their impact on infrastructure.
6. **Cost Savings in Site Investigations:** By providing detailed preliminary insights into subsurface conditions, the model will help optimize site investigation strategies, potentially leading to significant cost savings on field investigations and testing.
7. **Advancement in Public Infrastructure Management:** This project positions PennDOT as a leader in the application of advanced technologies for infrastructure management, setting a benchmark for other state agencies to follow.
8. **Comprehensive Historical Data Integration:** By leveraging historic geotechnical data and integrating it with new data through tools like Geosetta, the model will offer a more complete and accurate representation of Pennsylvania's subsurface conditions, improving the reliability of geotechnical assessments.
9. **Standardized Data Practices:** The adoption of industry-standard data formats like DIGGS ensures consistent data collection, storage, and exchange, which will enhance collaboration with external partners and future-proof the database for technological advancements.
10. **Predictive Capabilities:** Incorporating machine learning models for predicting SPT values and other geotechnical parameters will allow for comprehensive preliminary assessments in areas with limited physical testing data, further reducing uncertainty.

11. **Wider Data Coverage:** By integrating Geosetta's API for accessing nearby historical data, the model will provide a more comprehensive view of subsurface conditions, enhancing PennDOT's ability to plan and execute projects effectively.
12. **Quality Assurance for New Data:** The model will serve as a platform for validating and ensuring the quality of new geotechnical data submissions, thereby improving the overall consistency and reliability of data used in future infrastructure projects.

Overall, this project represents a significant advancement in the field of geotechnical data management and infrastructure development, offering long-term value through improved decision-making, efficient data management, and innovative approaches to public infrastructure management.

9. LITERATURE REVIEW REFERENCES

1. Bachus, R., Machairas, N., & Cadden, A. (2020). Standardizing Geodata Transfer and Storage. *Deep Foundations*, 97-100.
2. Barnes, R. (2003). Variogram Tutorial. Colorado: Golden Software, Inc.
3. Bentley Systems. (2021). gINT Connect edition. Retrieved July 18, 2024, from <https://www.bentley.com/wp-content/uploads/PFB-gINT-LTR-EN-LR.pdf>
4. Columbia University. (2024). Kriging Interpolation. Retrieved July 16, 2024, from Columbia University Mailman School of Public Health: <https://www.publichealth.columbia.edu/research/population-health-methods/kriging-interpolation#>
5. Curran, J., & Hammah, R. (2006). Seven lessons of geomechanics software development. The 41st U.S. Symposium on Rock Mechanics (USRMS): "50 Years of Rock Mechanics - Landmarks and Future Challenges". Golden, CO: American Rock Mechanics Association.
6. Data Interchange for Geotechnical and Geoenvironmental Specialists (DIGGS). (n.d.). DIGGSML Official Website. Retrieved from <https://diggsml.org/>.
7. De Rienzo, F., Oreste, P., & Pelizza, S. (2008). Subsurface geological-geotechnical modelling to sustain underground civil planning. *Engineering geology*, 96(3-4), 187-204.
8. Deutsch, C. V., & Journel, A. G. (1998). *GSLIB: Geostatistical software library and user's guide* (2nd ed.). New York: Oxford University Press.
9. Esri Inc. (2024). ArcGIS Pro help. Retrieved July 17, 2024, from <https://pro.arcgis.com/en/pro-app/latest/help/main/welcome-to-the-arcgis-pro-app-help.htm>
10. Esri Inc. (2024). How Kriging works. Retrieved July 16, 2024, from <https://pro.arcgis.com/en/pro-app/latest/tool-reference/3d-analyst/how-kriging-works.htm>
11. Federal Highway Administration (FHWA). (2021). Looking to The Sky for Geotechnical Data. *Public Roads*, Spring 2021. Retrieved from <https://highways.dot.gov/public-roads/spring-2021/04>.
12. Fish, M., & Lane, R. (2012). GIS and the New Hampshire Rock Cut Management System. Report FHWA-NH-RD-12323V, Bureau of Materials and Research, New Hampshire Department of Transportation, Concord. Retrieved from http://www.nh.gov/dot/org/projectdevelopment/materials/research/projects/documents/12323v_report.pdf.
13. Goovaerts, P. (1997). *Geostatistics for Natural Resources Evaluation*. Oxford University Press.
14. Harris, C. R., Millman, K. J., van der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., ... & Oliphant, T. E. (2020). Array programming with NumPy. *Nature*, 585(7825), 357-362.
15. Holovaty, A., & Kaplan-Moss, J. (2009). *The Definitive Guide to Django: Web Development Done Right*. Apress.
16. Hunter, J. D. (2007). Matplotlib: A 2D Graphics Environment. *Computing in Science & Engineering*, 9(3), 90-95.
17. Jordahl, K., den Bossche, J. V., Fleischmann, M., Wasserman, J., Gerard, J., Tratner, J., & Cochran, M. (2020). *GeoPandas: Python tools for geographic data*. GitHub.

18. Khan, M. S., Park, J., & Seo, J. (2021). Geotechnical property modeling and construction safety zoning based on GIS and BIM integration. *Applied Sciences*, 11(9), 4004.
19. Kringe, K., & Chatterjee, S. (2020). Uncertainty quantification of structural and geotechnical parameter by geostatistical simulations applied to a stability analysis case study with limited exploration data. *International Journal of Rock Mechanics and Mining Sciences*, 125, 104157.
20. Li, Z. (2021). An enhanced dual IDW method for high-quality geospatial interpolation. *Scientific reports*, 9903.
21. Matheron, G. (1963). Principles of geostatistics. *Economic Geology*, 58(8), 1246-1266.
22. McKinney, W. (2010). Data Structures for Statistical Computing in Python. *Proceedings of the 9th Python in Science Conference*, 51-56.
23. Olsen, M., Raugust, J., & Roe, G. (2013). Use of Advanced Geospatial Data, Tools, Technologies, and Information in Department of Transportation Projects. NCHRP Synthesis 446. Transportation Research Board of the National Academies, Washington, D.C. Retrieved from <https://nap.nationalacademies.org/catalog/22539/use-of-advanced-geospatial-data-tools-technologies-and-information-in-department-of-transportation-projects>
24. Ohio Department of Transportation. (n.d.). Transportation Information Mapping System (TIMS). Retrieved from https://gis.dot.state.oh.us/tims_classic.
25. Pavlis, T. L., & Mason, R. (2017). New World of 3D Geologic Mapping. *GSA Today*, 27(8), GSATG313A. Retrieved from <https://rock.geosociety.org/net/gsatoday/science/G313A/GSATG313A.pdf>.
26. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... & Duchesnay, E. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, 12, 2825-2830.
27. Pyrcz, M. J., & Deutsch, C. V. (2014). *Geostatistical reservoir modeling*. New York: Oxford University Press.
28. Remy, N., Boucher, A., & Wu, J. (2009). *Applied geostatistics with SGeMS: A user's guide*. Stanford: Cambridge University Press.
29. Rose, G., Kirk, P., Gibbons, C., & Lander, A. (2018). Three Dimensional Geological Models in Ground Engineering: When to Use, How to Build and Review, Benefits and Potential Pitfalls. *Australian Geomechanics*, 53(3), 79–88. Retrieved from <https://www.nzgs.org/libraries/three-dimensional-geological-models-in-ground-engineering-when-to-use-how-to-build-and-review-benefits-and-potential-pitfalls/>
30. Shepard, D. (1968). A two-dimensional interpolation function for irregularly-spaced data. *Proceedings of the 1968 ACM National Conference*, 517-524.
31. Supergeo. (2017). Spatial Statistical Analyst Extension Help. Retrieved July 15, 2024, from Supergeo Technologies Inc.: https://www.supergeotek.com/Spatial_Statistical_ENG_HTML/isotropy_and_anisotropy.htm#:~:text=The%20Sill%20value%20of%20the,vary%20along%20with%20different%20angles.
32. Thibaut, R., & Vandekerckhove, G. (2021). pysgems–Use SGeMS (Stanford Geostatistical Modeling Software) within Python. Retrieved July 17, 2024, from <https://github.com/robinthibaut/pysgems>

33. Turner, L., Saito, T. Grimes, P. (2008) End-User Interest in Geotechnical Data Management Systems, Sacramento, California Department of Transportation, Division of Research and Innovation.
34. Wang, H., Wang, X., & Liang, R. (2020). Study of AI based methods for characterization of geotechnical site investigation data. Columbus: Ohio Department of Transportation. Office of Statewide Planning and Research.
35. Waskom, M. L. (2021). Seaborn: Statistical data visualization. *Journal of Open Source Software*, 6(60), 3021.
36. Webster, R., & Oliver, M. A. (2007). *Geostatistics for Environmental Scientists* (2nd ed.). Wiley.
37. Yu, Y. (2010). *Geostatistical Interpolation and Simulation of RQD Measurements*. Vancouver: The University of British Columbia.
38. Zhang, L., He, C., Li, J., Wang, Y., & Wang, Z. (2017). Comparison of IDW and physically based IDEW method in hydrological modelling for a large mountainous watershed, Northwest China. *River Research and Applications*, 912-924.

APPENDIX A -AGENCY INTERVIEW QUESTIONS ON 3D GEOTECHNICAL MODELING AND ADVANCED DATA INTEGRATION

Introduction:

The project aims to develop a three-dimensional (3D) geotechnical model for subsurface soils across Pennsylvania. The following questions are designed to gather insights from agencies regarding their data management, visualization practices, and the potential integration of advanced technologies.

• Current Practices

- How does your Agency currently manage and visualize subsurface geotechnical data?
 - How do you collect and manage data at the project or district level (e.g., bridge, subsurface, pavement, maintenance, cost)?
 - What systems do you use (e.g., GIS, SQL)?
 - What challenges do you face with your current geotechnical data management system?

• 3D Visualization

- Is the Agency implementing 3D visualization tools?
 - Has your agency implemented or considered implementing a 3D geotechnical modeling system? If yes, please describe your experience.
 - What potential benefits do you see in having a comprehensive 3D geotechnical model for your state's subsurface conditions?
- Web and VR Applications:
 - Does your agency use any web-based or VR applications for geotechnical data visualization or analysis?
 - How valuable do you think VR technology could be for exploring and analyzing geotechnical data?

• Data Integration

- Can you describe your data capture process and integration into your database systems?
- How have you incorporated historic information into your system?
- Do you use any standardized data formats (e.g., DIGGS) for geotechnical data management? If so, what has been your experience?

• Data Validation

- Are there tools within these platforms to understand and manage uncertainty in the data sets? If so, how are they implemented, what items are evaluated, and how were the acceptability targets established?
- Does your agency have any recommendations on assessing uncertainty and variability of field observations and history?

- cal subsurface records?
 - Are there recommended data governance procedures to handle subsurface data, including engineering measurements, geological data, maps, remote sensing data, or other sources?
 - Do you use any platforms or software to interpret missing data spatially and over time using statistical inference models?
 - Are there established procedures or research studies to incorporate data from different sources or time periods? This can include merging data from different devices or systems measuring the same or different characteristics.
 - Do you develop tools internally or use commercial tools?
- **Predictive Modeling and Machine Learning**
 - Has your agency explored using machine learning or AI for predicting geotechnical properties (e.g., SPT values) in areas with limited physical testing?
 - What potential benefits or concerns do you see in incorporating predictive models into geotechnical assessments?
- **Implementation Considerations**
 - What potential challenges do you foresee in implementing a comprehensive 3D geotechnical modeling system with advanced data integration?
 - What resources (e.g., training, software, hardware) would your agency need to effectively utilize such a system?
 - How does the agency maintain these systems? Is it a central support team, or is each department or project manager responsible for their data?
- **Collaboration and Data Sharing**
 - How important is the ability to easily share geotechnical data with other agencies, contractors, or academic institutions?
 - Would a standardized system for geotechnical data management and visualization facilitate better collaboration on projects? Why or why not?
 - How would such tools integrate with or change existing workflows?
- **Future Outlook**
 - How do you see the role of advanced technologies (e.g., 3D modeling, VR, AI, machine learning) in geotechnical engineering evolving over the next 5-10 years at your Agency?
 - What features or capabilities would you like to see in future geotechnical data management and visualization systems?
- **Value Assessment**
 - On a scale of 1-10, how valuable would a comprehensive 3D geotechnical model with integrated historic data and predictive capabilities be for your agency's infrastructure planning and design processes?

- What specific tasks or decisions do you think would benefit most from such a system?
- What benefits has the Department seen from new data management tools and visualization, and what impact have they had on the availability of actionable data to meet the needs of your operations?
- Have you conducted any cost-benefit analyses on implementing advanced geotechnical data management and modeling systems?
- What factors would be (or were) most important in justifying the investment in such a system for your agency?

- **Additional Comments**

- Please provide any additional thoughts, concerns, or suggestions regarding the implementation of a comprehensive 3D geotechnical modeling system with advanced data integration and predictive capabilities.

APPENDIX B – SGEMS AND ARCGIS PRO DETAILED COMPARISON

1. Table 1: Overview of SGeMS

Category	Description
File Format	GSLIB: an ASCII format. It has several lines, including the title, name of the properties and values of the mentioned properties.
	SeGMS binary file format. It stores objects in uncompressed binary file format for faster loading and less disk space.
Exploratory Data Analysis (EDA)	Common parameters
	Histogram
	Q-Q plots
	Scatter Plots
Variogram Models	Nugget effect
	Spherical
	Exponential
	Gaussian
Interpolation Algorithm	Simple kriging
	Indicator kriging
	Cokriging
	Block kriging
Simulation Algorithm	LU Simulation
	Sequential Gaussian Simulation
	Sequential Gaussian Co-simulation
	Direct Sequential Simulation
	Sequential Indicator Simulation
	Sequential Indicator Co-simulation
	Block Sequential Simulation
	Block Error Simulation
	Multipoint Simulation

Table 2: Overview of ArcGIS Pro Geostatistical Analyst Extension

Category	Description
File Format	ArcGIS supports 37 data formats. Using <i>ArcGIS Data Interoperability extension for Desktop</i> additional 70 data formats can also be used
Exploratory Data Analysis (EDA)	Common parameters
	Histogram
	Q-Q plots
	Scatter Plots
Variogram Models	Circular
	Spherical
	Tetraspherical
	Pentaspherical
	Exponential
	Gaussian
	Rational Quadratic
	Hole effect
	K and J Bessel
	Stable
Interpolation Algorithm	Simple kriging
	Ordinary Kriging
	Indicator kriging
	Universal Kriging
	Probability Kriging
	Disjunctive kriging
	Empirical Bayesian Kriging
Simulation Algorithm	Gaussian Geostatistical Simulation

APPENDIX C - DETAILED TECHNICAL CONCEPTS AND ANALYSIS

Variogram analysis is a core component of geostatistical methods for quantifying spatial variability. This section delves deeper into the formulation, components, and modeling of variograms.

Table 3: Basic Statistics of the Considered Datasets

Parameters	Dataset 1	Dataset 2
Count	5000	5000
Average	100	100
Standard Deviation	19	19
Median	100.47	100.83
10th Percentile	73.4	73.6
50th Percentile	125.33	124.95

The histograms of these two synthetic databases are shown in Figure 1. Comparing both Table 3 and Figure 1, the two datasets appear to be almost identical.

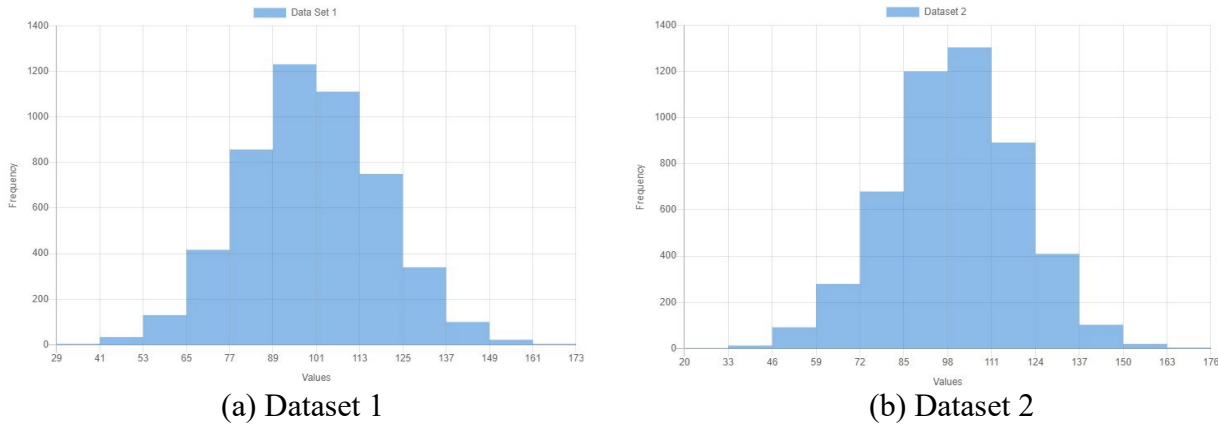


Figure 1: Histogram of the datasets

However, the apparent differences between the datasets can be observed from the contour plots shown in Figure 2. Dataset 1 is sharper or rougher compared to dataset 2. It is important to note that the variability in dataset 1 is not more compared to dataset 2, as both have the same standard deviation. The spatial textural changes in dataset 1 are more rapid, which is visible from the smaller and sharper high (red) or low (blue) zones in Figure 2(a). On the contrary, the spatial textural changes in dataset 2 are more gradual.

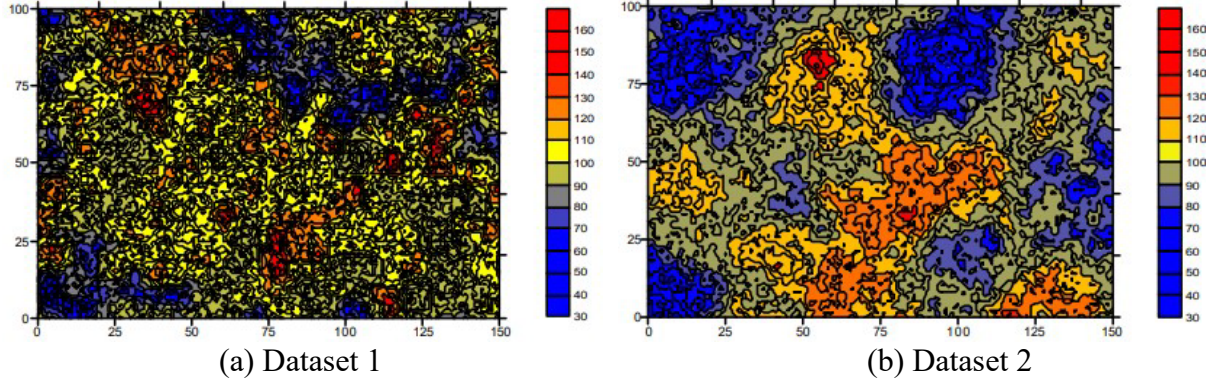


Figure 2: Contour plots of the datasets (Barnes, 2003)

It is not surprising that the common statistics or histograms cannot identify or quantify these textural variations, as their calculations do not include spatial information. On the other hand, variograms can quantitatively characterize the spatial continuity or roughness in a dataset. The variograms of the two datasets are shown in Figure 3. The slope for Dataset 1 is much steeper than Dataset 2.

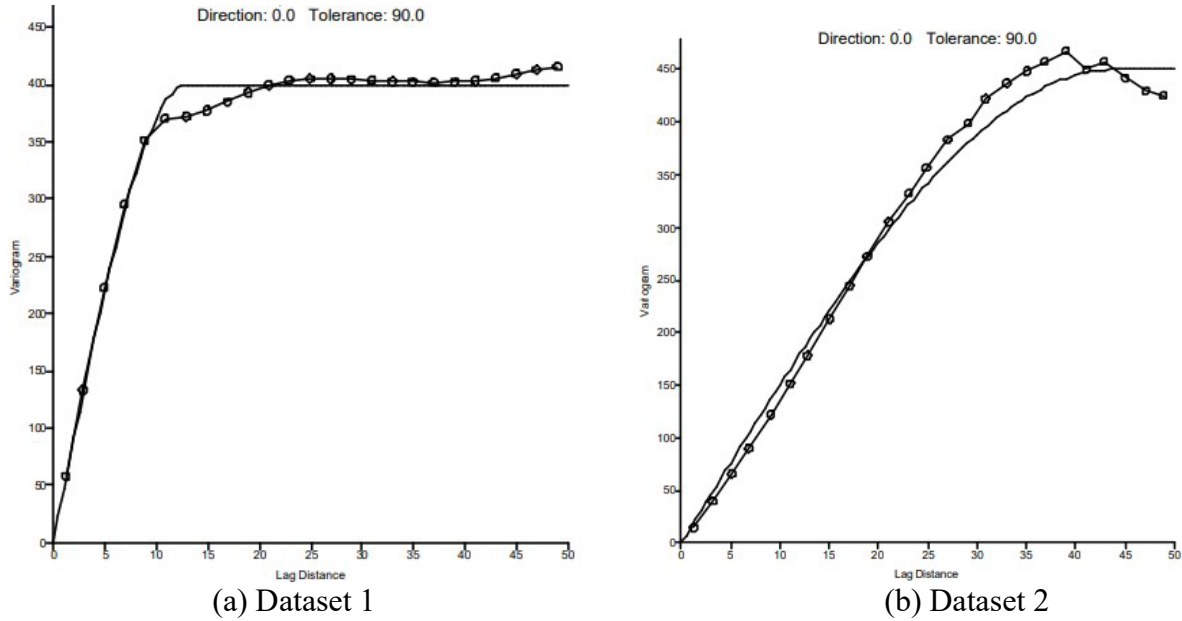


Figure 3: Variograms of the datasets (Barnes, 2003)

Formulation and Components

The mathematical formulation of semivariogram is shown in Eq. (1)

$$\gamma(h) = \frac{1}{2N(h)} \sum_{\alpha=1}^{N(h)} [z(u_{\alpha}) - z(u_{\alpha} + h)]^2 \quad (1)$$

where, $z(u_\alpha)$ and $z(u_\alpha + h)$ are the tail and head values of the lag vector at sample locations, respectively. The schematic of a lag vector with head and tail position is shown in Figure 4.

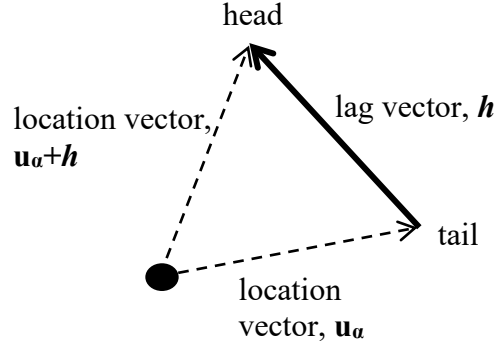


Figure 4: Schematic of lag vector (Pyrzcz & Deutsch, 2014)

A continuous function of Eq. (1) is obtained by considering different lag distances. It is worth mentioning that the $\frac{1}{2}$ term converts the variogram into a semivariogram, as semivariogram directly relates to covariance function $C_x(\mathbf{h})$ and univariate variance σ_x^2 , as shown in Eq. (2). The term semivariogram and variogram will be used interchangeably from here on in the rest of the document.

$$C_x(\mathbf{h}) = \sigma_x^2 - \gamma(\mathbf{h}) \quad (2)$$

The univariate variance σ_x^2 in Eq. (2) is also known as the sill of the variogram. Sill is the variogram value at range. The range is defined as the lag distance at which the variogram flattens or shows maximum variability in the data. Another component of the variogram is the nugget effect. It is the value of the variogram at an infinitesimal distance from the lag distance value of zero. The nugget effect can be attributed to the measurement error or spatial variations at a distance smaller than the sampling interval or both. Different components of the variogram are shown in Figure 5.

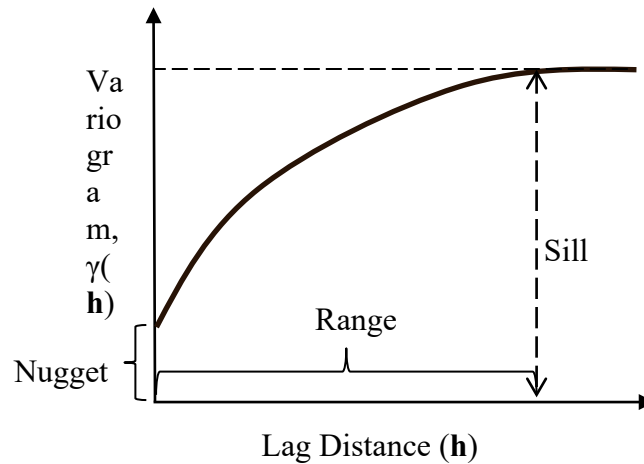


Figure 5: Different components of variogram

Modeling

Once the experimental variogram has been calculated based on the available data, it is used to develop a variogram model for prediction. The primary reasons for modeling are to know the value of the variogram for all possible lag distances ($\mathbf{h}$) and directions and to incorporate the available geotechnical or geological knowledge, such as information on directions of continuity (Pyrz & Deutsch, 2014). Nested and positive definite variogram structures can be used to model the spatial continuity using Eq. (3).

$$\Gamma_x(\mathbf{h}) = \sum_{i=1}^{nst} \gamma_i(\mathbf{h}) \quad (3)$$

where, $\Gamma_x(\mathbf{h})$ is the nested variogram structure resulting from the summation of a number of nested variograms $\gamma_i(\mathbf{h})$. Positive definiteness of the variogram models is an important characteristic, which ensures that the variances remain positive for any linear combinations. Thus, the following models are typically used for variogram modeling:

- Spherical
- Exponential
- Gaussian
- Nugget only

Variogram modeling is also used to identify the presence of isotropy or anisotropy in the data. As mentioned earlier, the variogram model is a function of distance. If the variogram of a spatial variance is merely a function of distance and the variability does not vary by direction, it is called isotropy. However, if the variability varies along spatial directions, it is called anisotropy. To identify the presence of isotropy or anisotropy, first, a variogram model must be fitted along all the directions. After that, different variogram models are fitted in different directions. If the range, sill, and nugget effect remain the same for all the variogram models, then the spatial data is isotropic, or else it is anisotropic. There are several sources of anisotropy in the spatial data. They are geometric anisotropy, zonal anisotropy, and mixed anisotropy.

The Sill value of the variogram with geometric anisotropy varies only with distances, not with directions. This means that when the distances between samples at different spatial angles are the same, their sill is also the same, but their ranges may differ along different angles. The direction with the largest range corresponds to the direction of the variogram and is known as the "Major Range". The smallest range is referred to as the "Minor Range" (Supergeo, 2017).

The Sill value of the variogram with zonal anisotropy changes not only with distance but also with direction. Essentially, when samples are at the same distance but in different directions, their Sill values and ranges differ. The Zonal Anisotropic variogram model comprises two or more anisotropic variograms. Typically, natural phenomena do not exhibit this type of anisotropy. Mixed anisotropy is the combination of both the geometric and zonal anisotropy. Figure 6 shows the variograms of different types of anisotropy in the spatial data.

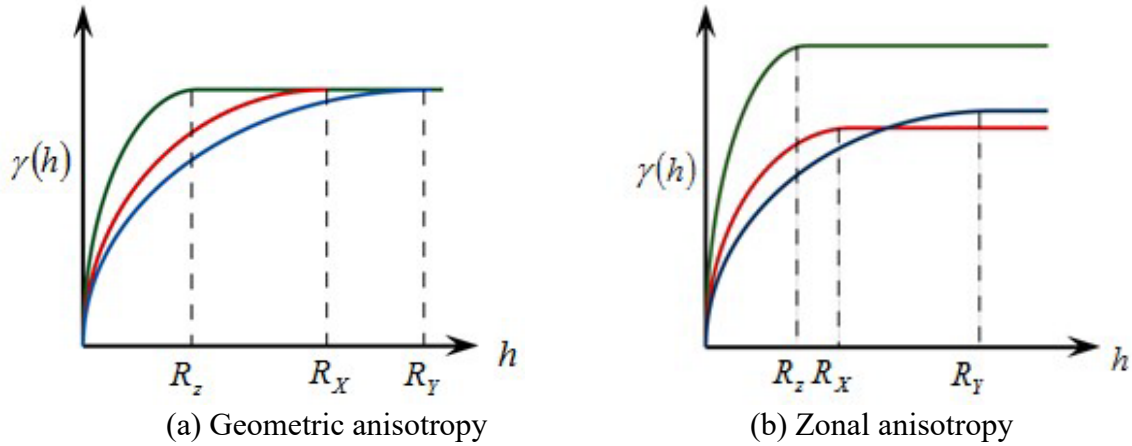


Figure 6: Variograms of different types of anisotropy in the spatial data

Geostatistical Interpolation

After a model that best describes the experimental data has been selected, it can be used to quantify the spatial uncertainty through various geostatistical interpolation schemes, such as kriging. Geostatistical interpolation is based on the assumption that the spatial objects that are closer to each other are more similar than the objects that are further apart. Interpolation is used to predict the unknown values of a new location based on the limited number of sample data points. Several geostatistical interpolation methods have been used to predict unknown values, such as inverse distance weighted (IDW), kriging, spline, etc. A brief discussion of these methods is presented below.

Inverse Distance Weighted (IDW)

In IDW, values of an unknown location are determined based on the combination of linear weights of a set of sample points. The weights are inversely related to the distance, as shown in Eq. (4).

$$\lambda_{ij} = \frac{1}{r_{ij}^p} \quad (4)$$

where, r is the distance between the point with known values and the point where the spatial value needs to be determined. p is the power parameter, which controls the significance and takes positive, real numbers like 1, 2, 3, etc.

The main assumption of IDW is that the influence of the variable decreases with distance. Thus, IDW can account for the closeness of the spatial sampling points. However, it cannot account for the data redundancy (Li, 2021). IDW suffers in the presence of clustered data.

Despite being of almost the same distances, the isolated locations with known values will have lower weights in the estimation of the predicted value according to IDW compared to the combined weights of the clustered points, as shown in

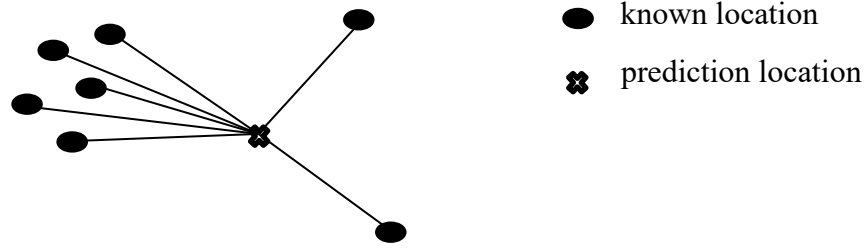


Figure 7: Limitation of IDW

Kriging

One of the popular spatial interpolation methods is kriging. IDW is a deterministic approach for spatial interpolation as it only considers the distance between the points. On the other hand, kriging is based on statistical models that include autocorrelation, i.e., the statistical relation among the measured points. Kriging is based on the assumption that the spatial correlation between sample points can be used to account for variations in the surface, which is reflected through their distance or direction from each other. In order to predict the value of an unknown location, the kriging process involves fitting a mathematical function to a designated number of points or to all points within a specified radius. This process is comprised of several steps, including exploratory statistical analysis of the data, variogram modeling, and surface creation. Kriging is particularly ideal when there is a spatial correlation between the distances or directional bias in the data (Esri Inc., 2024). The general formula for kriging is represented by Eq. (5).

$$z^*(u) = \sum_{\alpha=1}^n \lambda_{\alpha} z(u_{\alpha}) \quad (5)$$

where, $z^*(u)$ is the kriging estimated value at the unknown location, $z(u_{\alpha})$ is the spatial values at known locations, and λ_{α} is the weights, which are determined based on the covariances between the data points. An unbiasedness constraint is added to the kriging estimate by making the sum of the weights equal to one. For that, the remainder of the sum of the weights is multiplied with the global average estimate ($\bar{z}$), as shown in Eq. (6).

$$z^*(u) = \sum_{\alpha=1}^n \lambda_{\alpha} z(u_{\alpha}) + (1 - \sum_{\alpha=1}^n \lambda_{\alpha}) \bar{z} \quad (6)$$

It is worth noting that setting the weight constraint allows the prediction with the global average when no further information is available.

The consideration of the closeness and redundancy of the data sets kriging apart from IDW. Kriging weights consider not only the spatial continuity between the available data and the estimation location but also the spatial continuity between the available data themselves, as shown in Eq. (7).

$$\sum_{ij} \lambda_j = \sigma_{i0} \quad (7)$$

where, Σ_{ij} are covariances between the locations with known values, and σ_{i0} is the covariances between the known locations and the unknown location. As mentioned previously, a variogram model is required for the kriging estimate. The covariances in Eq. (7) can be obtained from the covariances of the variogram model shown in Eq. (2).

There are several types of kriging. They are:

Ordinary Kriging: It is simplest form of kriging which assumes that the mean and variances remain constant across the spatial field.

Universal Kriging: This method of kriging relaxes the stationary assumption of ordinary kriging, and it is particularly used when there is a trend in the spatial data.

Block Kriging: This method estimates the average values over a block of grids instead of a single point, which reduces the prediction error.

Cokriging: In this method of kriging, additional correlated variables are used to enhance the estimation accuracy at each location.

Despite its wide usage, kriging does have some limitations (Columbia University, 2024). The weights of the kriging are obtained from the modeled variogram. Thus, if the variogram model is not accurate, then the kriging prediction will be less accurate. Moreover, kriging provides the “best” prediction with the available sample data. If the available sample data is limited and the variability is high, then kriging will use the global average for prediction as shown in Eq. (6). It results in the “over-smoothing” of the data, disregarding the high and low spatial values (Yu, 2010). Thus, the “best” estimate of kriging may not be the “accurate” estimation.

Simulation

Interpolation techniques used in the geostatistical analysis generate a single value for each unknown point. In reality, each unsampled point can have numerous equally likely values, as the actual value at each unsampled location is uncertain. The geostatistical simulation generates multiple maps that replicate the actual phenomenon and present these potential values, offering a foundation for risk assessments, economic decision-making, and other estimations that involve uncertainty.

Two types of simulations are typically used for spatial data analysis. For categorical data, sequential indicator simulation is used, and sequential Gaussian simulation is used for continuous data. Simulations are used to generate multiple realizations to capture spatial uncertainty. Realizations are generated by keeping the inputs and parameters constant but changing only the random numbers. Simulations not only provide a single value for the unknown location but also provide the uncertainty associated with the estimated value through the conditional cumulative distribution function (CCDF). CCDFs can be used to measure uncertainty, probability threshold, and interval, general statistics, etc. The general steps that are followed in the geostatistical simulations are provided below (Kring & Chatterjee, 2020):

- Transform the data
- Calculate variogram
- Define a random path for all the grid node locations
 - Consider nearby available data and previously simulated values for simulation

- Solve kriging equations to construct a conditional distribution
- Perform Monte-Carlo Simulation to obtain random simulated value from the CCDF
- Assign the simulated value to the considered node locations
- Repeat the steps for all the nodes
- After generating a realization, it is checked whether the data, histogram, and variogram are honored or not
- Repeat the above steps to generate multiple realizations by considering a different random path in Step 3

Kring and Chatterjee (2020) used geostatistical simulations to identify the uncertainties in the structural and geotechnical parameters for stability analysis of rock slopes (Kring & Chatterjee, 2020). P number of realizations of faults were created using sequential indicator simulation. The presence of faults was represented using categorical variables. Sequential Gaussian simulation was used to generate a Q number of realizations for rock quality designation (RQD), which was a continuous variable. These $P \times Q$ number of spatial maps were used to quantify the joint uncertainty of the stability analysis arising from both the faults and RQD. The mean value of factors of safety (μ_F) of the rock slopes for the $P \times Q$ simulated maps can be obtained using Eq. (8)

$$\mu_F = \frac{1}{P \times Q} \sum_{i=1}^{P \times Q} F_i \quad (8)$$

where, F_i is the factor of safety for the i -th simulated spatial maps of fault structure and geotechnical parameter.

The central limit theorem can be used to obtain the probability of failure from the uncertainty-based model using Eq. (9)

$$p_F = \frac{n_F}{P \times Q} \quad (9)$$

where, $n_F = \frac{1}{P \times Q} \sum_{i=1}^{P \times Q} I(F_i)$, and the function $I(F_i)$ is defined as:

$$I(F_i) = \begin{cases} 1 & \text{if } F_i \leq 0 \\ 0 & \text{if } F_i > 0 \end{cases} \quad (10)$$

The proposed method was applied for the slope stability analysis of an open pit mine design with limited exploration data. From the geostatistical simulated maps, a safe slope angle was determined to be 44° .

Rienzo et al. (2008) utilized 3D subsurface geological-geotechnical modeling to optimize the planning and development of subsurface structures in urban environments, specifically illustrated through the case of the City of Turin in Northern Italy. They created a geological map using data from 380 boreholes covering approximately 50 km^2 . A 3D GIS (Geographical Information

System) model was developed to manage substantial spatial data and maintain it for a geological Turin model. They structured the database into various tables, including the well table, lithology table, strat_unit table, and cem_class table, to handle borehole identification, lithological descriptions, stratigraphic units, and cementation classes. Following stratigraphic analysis, the researchers georeferenced the top of the identified layers to an absolute system with a vertical axis starting from sea level. For spatial interpolation, they utilized the deterministic triangular network (TIN) method in MGE-Intergraph's Terrain Analyst module. Additionally, the team conducted uniaxial compression and shear tests on approximately seventy soil samples collected from six different sites along the metro route in the city center to examine geotechnical properties. The results revealed a broad range of uniaxial compressive strength (ranging from 1.9 MPa to 6.6 MPa) and elastic modulus (ranging from 500 MPa to 2300 MPa) values, with no discernible trends relating to depth or lithology. The researchers then correlated the geotechnical parameters with the cementation maps, finding a relationship between higher uniaxial compressive strength values and strongly cemented sediments and lower values with non-or-poorly cemented sediments. Overall, the study highlights the usefulness of 3D geological-geotechnical modeling in evaluating the effects of tunnel excavations on surface and subsurface structures through numerical simulations and in guiding design revisions to address potential subsidence issues.

Khan et al. (2021) studied the integration of Building Information Modeling (BIM) and Geographic Information System (GIS) to model geotechnical properties and zoning in the Peshawar region of Pakistan. The decision to focus on this area was influenced by the availability of borehole data obtained from soil mechanics laboratories and local geotechnical investigation consultants. The data consisted of in-situ tests such as Standard Penetration Test (SPT-N) values as well as various laboratory test parameters like unconfined compressive strength and water table information. Using ArcMap software, a geodatabase was created to incorporate borehole information for geostatistical interpolation through kriging, and digital maps were produced illustrating geotechnical parameters. The GIS mapping effectively represented SPT-N values, plasticity index, and variations in unconfined compressive strength across the study area, contributing to a user-friendly and comprehensive understanding of the geotechnical information. The accuracy of the generated maps was validated by comparing actual test results with values on the maps, demonstrating the effectiveness of the proposed method for geotechnical mapping. Additionally, the study compared its approach with a previous study in the Potohar region, which revealed reduced uncertainties and improved spatial coverage in the Peshawar region. Autodesk Civil 3D was employed for BIM to handle the coordinate system, visualize borehole data, and create geotechnical surfaces and solids, as shown in Figure 8. Each soil type was represented with colored surfaces and solids based on the borehole data linked with Coordinate Geometry (COGO) points. The volume of each soil type was determined using the surface-to-surface method. Additional properties, such as strength, classification, and compaction, were also linked to the appropriate solid layers within the BIM environment. The integrated BIM-GIS approach facilitated the storage and management of geotechnical information in the Industry Foundation Classes (IFC) format, enabling collaborative efforts and data exchange among stakeholders. The developed methodology was utilized to identify safe and suitable construction techniques for excavation design, as understanding soil types is crucial for worker safety. The paper provides an explanation of the OSHA soil classification system and details how the developed digital maps, which visualize the distribution and properties of soil

types, can assist practitioners in planning the appropriate protective systems for excavation works.

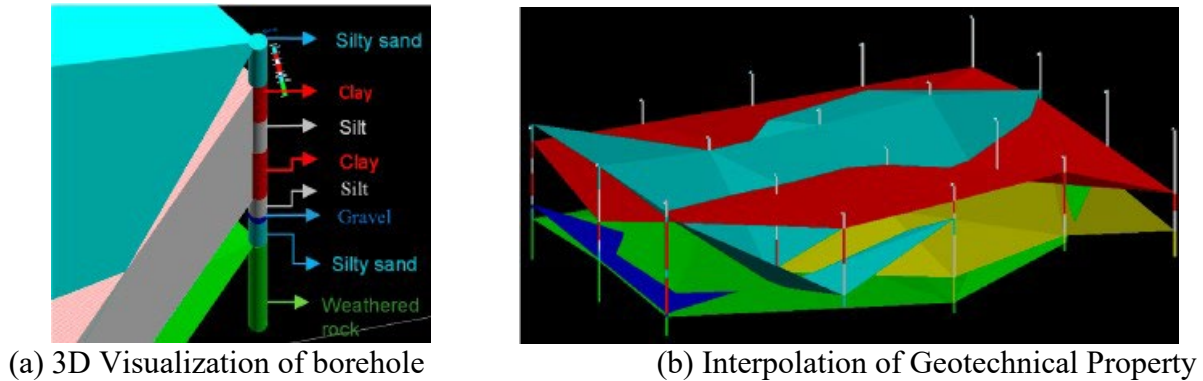


Figure 8: Visualization of geotechnical properties in BIM environment (Khan et al., 2021)

The research team at the University of Dayton has developed a new methodology combining Artificial Intelligence (AI) based machine learning algorithms and other disciplinary principles to interpret geotechnical site investigation data (Wang et al., 2020). The conceptual framework of the proposed methodology includes using a machine learning module to analyze the original site characterization data, like borehole logs, to deduce the subsurface stratification model and determine the uncertainty of layer boundaries. Geotechnical engineers can extract, visualize, and analyze the interpreted data for appropriate use in the planning, design, and construction. The mathematical foundation of the developed methodology is based on various disciplines, including a user-friendly Python interface for importing the required data from the DIGGS-compliant site investigation data, data fusion, stochastic simulation techniques to estimate subsurface conditions at unexplored locations with uncertainty quantification, and a visualization interface that presents the computational results along with the associated reliability and risk analysis. Three case studies were performed to demonstrate the benefits of the developed methodology in enhancing the site characterization capabilities.

APPENDIX B. REVISED INTERIM REPORT 4

Three-Dimensional Modeling and Uncertainty Quantification of Subsurface Soils

December 4, 2025

Revised March 2, 2026

RFQ 221216

Purchase Order (PO) 4300805489

Task 4: Develop Recommendations Interim Report

Prepared for:

Pennsylvania Department of Transportation

Prepared by:

Applied Research Associates, Inc. (ARA Inc.)

and

Schnabel Engineering, LLC (Schnabel)

TABLE OF CONTENTS

DEFINITIONS	II
1. EXECUTIVE SUMMARY	1
2. INTRODUCTION AND TASK GOALS	2
2.1 PROJECT OVERVIEW	2
2.2 PREVIOUS TASKS.....	2
2.3 CURRENT TASK (TASK 4).....	2
3. ESTABLISHING GDMS VALUE	4
3.1 STAKEHOLDER IDENTIFICATION.....	4
3.2 STAKEHOLDER ENGAGEMENT.....	6
3.3 STAKEHOLDER EXPANSION	8
3.4 RETURN ON INVESTMENT EVALUATIONS	9
4. PILOT RESPONSE AND EXPANSION.....	11
4.1 PILOT RESPONSE.....	11
4.2 STAGED EXPANSION	12
4.3 STAKEHOLDER ENGAGEMENT.....	14
4.4 DEVELOPMENT SCHEDULE	15
5. DATA COLLECTION AND INTEGRATION	18
5.1 NEW FIELD (BORING) RECORDS	18
5.2 HISTORICAL BORING RECORDS	20
5.3 HISTORICAL CONSTRUCTION RECORDS	22
5.4 GDMS MATURATION	24
6. GDMS TOOLS AND UTILIZATION.....	27
6.1 SIMPLE USE CONDITION: ARCHIVE AND RETRIEVAL.....	27
6.2 MODERATE USE CONDITION: STRUCTURED DATA AND VISUALIZATION	27
6.3 EXTENSIVE/COMPLEX USE CONDITION: LIFECYCLE DIGITAL TWIN	28
6.4 EXAMPLE USE CASES	28
6.5 EXPANSION OF GDMS TOOLS AND USES	30
7. CLOSING	33
REFERENCES	34

DEFINITIONS

Definitions of key terms are modified from publicly available definitions and software documentation.

API – Application Programming Interfaces (APIs) are a set of rules, protocols, and tools that allow different software systems to communicate and exchange data; serving as a bridge between applications. APIs enable applications to request and deliver information without knowledge of each systems’ internal workings and design.

CesiumJS – Cesium An open-source JavaScript library for building high-precision, hardware-accelerated 3D globes and 2D maps in the browser (WebGL), optimized for streaming large geospatial datasets and visualizing 3D Tiles, terrain, and imagery (<https://cesium.com/platform/cesiumjs/>).

DIGGS – Data Interchange for Geotechnical and Geoenvironmental Specialists (DIGGS) is a GML/XML-based schema and standard for exchanging geotechnical and geoenvironmental data between organizations, designed to standardize borehole, sampling, and test data for easier sharing and reuse (<https://geoinstitute.org/special-projects/diggs>).

GIS – Geographic Information System (GIS) is a computer-based technology used to capture, store, manage, analyze, and visualize spatial or geographic data (<https://www.esri.com/en-us/what-is-gis/overview>). PostGIS is a software that adds geospatial capabilities to PostgreSQL databases (<https://postgis.net/>).

JavaScript – JavaScript is a programming language that enables web pages and applications to perform dynamic actions (e.g., updating content, responding to user events, controlling multimedia). Along with HTML and CSS Javascript is a core web technology designed to manipulate the Document Object Model (DOM) in browsers to create interactive experiences (https://developer.mozilla.org/en-US/docs/Learn_web_development/Core/Scripting/What_is_JavaScript).

JSON – JavaScript Object Notation (JSON) is an open, language-independent data interchange format that uses human-readable text to represent objects, arrays, strings, numbers, booleans, and null values (<https://www.json.org/json-en.html>). GeoJSON is a open standard specifically developed for the representation of geographic features and their non-spatial attributes following JSON notation (<https://geojson.org/>).

Leaflet – Leaflet is a widely-used small and efficient JavaScript library designed specifically for creating interactive maps for both desktop and mobile browsers. Leaflet focuses on simple structure, high performance, and usability (<https://leafletjs.com/>).

Mortar 3D -

OCR – Optical Character Recognition (OCR) is a computational process that analyzes the pixel structure of text within an image and converts those patterns into encoded characters using feature extraction and classification algorithms. Modern OCR systems typically apply machine-learning based models to handle variable fonts, noise, and distortions, producing machine-readable text suitable for indexing, parsing, and downstream data workflows.

RESTful Software – RESTful software refers to systems or APIs that follow the principles of the Representational State Transfer (REST) architecture. REST is a set of guidelines for designing scalable, flexible, and efficient web services that enable communication between clients and servers over the internet (<https://aws.amazon.com/what-is/restful-api/>).

SQL – Structured Query Language (SQL) is a standard language for storing, manipulating and retrieving data in databases (<https://www.w3schools.com/sql/>). PostgreSQL is a open-source relational database management system (RDBMS) which include a focus on SQL compliance (<https://www.postgresql.org/>).

Stakeholder – An individual, group, or organization that has an interest in or is affected by a project, decision, or system. Stakeholders can be internal (e.g., employees, managers) or external (e.g., customers, regulators, communities).

1. EXECUTIVE SUMMARY

The Three-Dimensional Modeling and Uncertainty Quantification of Subsurface Soils (RFQ 221216) project, initiated in June 2024, is advancing PennDOT's development of a Geotechnical Data Management System (GDMS) to modernize subsurface data management and improve geotechnical decision-making. The effort integrates geologic mapping, PennDOT's geotechnical records, and external datasets into a unified, standards-based platform supported by robust governance and quality assurance protocols.

Initial tasks established the technical foundation with comprehensive literature review and stakeholder interviews identified national best practices, data standards, and operational lessons from peer DOTs. These findings informed the design of a web-based GDMS architecture built on Django with PostgreSQL/PostGIS, incorporating DIGGS-compliant schemas, API-first ingestion, and phased deployment. The pilot system successfully demonstrated integration of geotechnical datasets, interactive visualization, uncertainty quantification, and role-based access control. While not yet integrated into production systems, the pilot validated the technical approach and provided a platform for future testing and development.

Feedback from PennDOT engineers and IT staff has been positive, particularly regarding the system's visualization capabilities and database flexibility. Districts expressed interest in integrating historic datasets, CAD surfaces, and geotechnical asset management program data. But at the same time had concerns regarding data interpolation, liability, and workflow consistency underscoring the importance of formalized QA/QC and governance policies, accessibility compliance, and standardized procedures.

In Task 4 the project has moved to emphasize communication, outreach, and workflow development to increase the chance of organizational adoption and acceptance. Structured engagement with PennDOT leadership, geotechnical engineers, district staff, contractors, utilities, and federal agencies may all be used to help guide refinement and ensure the GDMS reflects PennDOT's needs. Documentation practices, stakeholder registries, and interactive workshops represent several development tools available to PennDOT to establish governance structures and support risk-informed planning made available through the GDMS.

Future expansion is recommended to proceed in staged phases. The first stage will establish standardized data governance and QA/QC protocols. District-level pilots are recommended to validate workflows and integrate historic datasets. Statewide expansion will embed GDMS into PennDOT's enterprise systems, broaden asset coverage, and introduce automated project tracking and centralized standards. Ultimately, GDMS will be positioned to evolve into a key element of a digital twin environment, supporting resilience planning, extreme weather considerations, and interoperability with BIM platforms.

In summary, the GDMS pilot has demonstrated technical feasibility and strategic value. Its success now depends on formalizing governance, expanding stakeholder engagement, and carefully managing the transition from pilot to enterprise system. With staged expansion and sustained investment in training and change management, PennDOT is positioned to institutionalize GDMS as a cornerstone of geotechnical asset management and digital delivery.

2. INTRODUCTION AND TASK GOALS

2.1 Project Overview

Research project *Three-Dimensional Modeling and Uncertainty Quantification of Subsurface Soils* (RFQ 221216) was initiated on June 3, 2024, following a comprehensive kickoff meeting held on May 15, 2024. This initiative aims to develop a comprehensive three-dimensional geologic model of Pennsylvania, integrating multiple data sources and advanced modeling techniques to enhance understanding of subsurface conditions throughout the state. The project scope, clarified during the kickoff meeting, encompasses four major components: enhancement of existing geologic mapping, integration of PennDOT's subsurface geotechnical data, incorporation of additional relevant data sources, and development of robust data management strategies. This multi-faceted approach ensures comprehensive coverage of both historical data and ongoing geotechnical investigations.

2.2 Previous Tasks

Task 1. Literature Review and Stakeholder Interviews

The team completed a thorough literature review and comparative analysis of national and international geologic and geotechnical data systems, cataloging best practices, software tools, and data standards. Structured interviews with Caltrans, PennDOT, LaDOTD, and Ohio DOT captured operational practices, common challenges, and lessons learned about inventorying geotechnical assets, adopting standardized schemas, and applying remote sensing and ML. Task 1 produced a written literature review, interview summaries, software comparisons, and appendices on geostatistical and uncertainty methods to inform system design.

Task 2. Implementation Plan and System Architecture

Task 2 translated Task 1 findings into a practical implementation plan and system architecture. The plan specifies a web-based GDMS built on a Django backend with PostgreSQL/PostGIS spatial storage, DIGGS-compliant data structures, API-first ingestion, and phased deployment steps from proof of concept to statewide rollout. Detailed components include data ingestion and legacy conversion strategies, quality assurance / quality control (QA/QC) procedures, access control, performance optimization, and integration points with PennDOT enterprise systems and authoritative external services.

Task 3. Pilot Development and 3D Model Demonstration

The pilot GDMS was developed and demonstrated as a proof of concept. The pilot implements the core stack and data model (borings, SPT records, lithology, lab tests, projects), provides interactive 2D/3D visualization (Leaflet, Mortar 3D), supports project grouping and queries, and includes an uncertainty visualization tool to show data coverage and confidence. The pilot also demonstrates vendor-neutral ingestion, basic legacy conversion, and role-based access control. The developed GDMS remains a demonstration system, not yet integrated into PennDOT production IT, and the dataset and interactive tools are not intended for engineering decisions.

2.3 Current Task (Task 4)

Collectively, Tasks 1 through 3 validated the technical approach, identified governance and QA priorities, and produced a usable pilot for user testing. The scope of Task 4 is to continue the

development of an effective tool that can be readily implemented across the organization, including workflows and value proposition development to PennDOT and associated geoprofessional organizations and communities. Workflow guidance was developed to better define data collection, PennDOT and state integration, and maintenance and development of the GDMS. This task also explores how to best establish GDMS value with stakeholders both inside and outside of PennDOT; how to stage the GDMS expansion for the greatest chance of success; and potential future tools and uses for the GDMS depending on the level of implementation.

3. ESTABLISHING GDMS VALUE

Successful development and implementation of the pilot GDMS and future iterations is dependent on demonstrating the value of the concept, technologies, features, and applicability to stakeholder interests. The following sections detail best practices to identify and engage stakeholders to establish and maintain GDMS value starting with stakeholder identification, mapping, and engagement strategies. Stakeholders presented include a brief description which tie into the GDMS. While listed as individual stakeholders, individual personnel, groups, or organizations within PennDOT may hold several concurrent stakeholder interests. The provided lists are not exhaustive and provide internal and external stakeholders who are anticipated to be the most engaged with the GDMS, a description, and their potential engagement. This list is developed to help established the potential impacts of workflows, processes, and long-term design decisions on stakeholder usability and engagement.

3.1 Stakeholder Identification

As practiced in Task 1 and onward, early engagement of potential stakeholders is crucial to develop and sustain GDMS interest and investment. Task 1 identified preliminary stakeholder interests in the GDMS through a series of agency interviews and discussions with PennDOT personnel indicating both desired use-cases (e.g., asset inventories and hazard assessments) and potential limitations (e.g., limited resources, structural efficiency). Identified use-cases based on Task 1 and subsequent discussions has resulted in a relatively robust range of potential stakeholders for future engagement internal and external to PennDOT, including:

Internal Stakeholders (PennDOT)

- **Executive leadership:** Program sponsors and senior managers who set strategic priorities, approve funding, and champion statewide adoption.
- **Geotechnical engineering division:** Subject matter experts who define technical requirements, validate geotechnical content, and lead quality assurance (QA) for subsurface data.
- **District personnel and project managers:** Local implementers who submit exploration requests, review local data, and provide pilot feedback for iterative improvements.
- **Design and review engineers:** Users who consume GDMS outputs for project design, foundation decisions, and risk assessments; reviewers who apply QA/QC checklists and approve data for production use.
- **Construction and field operations:** Field crews and construction managers who provide as-built and construction data and rely on GDMS for site planning and safety.
- **Maintenance and asset managers:** Teams and individual personnel that use the GDMS to inventory geotechnical assets, prioritize inspections, and plan lifecycle interventions.
- **Surveyors and mappers:** Providers of control, LiDAR, and topographic data who ensure accurate georeferencing and integration with the spatial database.
- **Spatial analysts:** Developers and analysts who manage spatial layers, mapping services, and visualization components.

- **IT and enterprise management:** Parties responsible for hosting, security, API integrations, system performance, and alignment with PennDOT IT policies.
- **Traffic and safety operators and oversight:** Coordinators of traffic control requirements and ensure safe field operations tied to exploration schedules.
- **Right of Way and permitting:** Property access managers and those who deal with utility clearances.
- **Data governors and record management:** Owners of data standards, metadata policies, retention rules, and DIGGS compliance enforcement.
- **Training and workforce development:** Responsible for developing role-based training, user guides, and adoption support materials.

External Stakeholders

- **Consulting geotechnical firms and materials testing laboratories:** Businesses which perform site investigations, laboratory testing, interpretations, and submit DIGGS-compliant deliverables or deliverable test results, QA data, and standardized reporting.
- **Construction contractors and subcontractors:** Those who execute field work, provide as-built and construction records, and implement mitigation measures.
- **Utility owners and operators:** Owners and operators of utilities and utility corridors who provide location data, grant clearances, and coordinate relocations or outages.
- **Local government units and municipalities:** Who may issue permits, provide local GIS and land-use data, and coordinate public outreach, and otherwise utilize public data for their own development and land-use.
- **Other state agencies:** Pennsylvania Department of Conservation and Natural Resources (DCNR), Pennsylvania Department of Environmental Protection (DEP), and other state level agencies who provide authoritative geologic maps, bedrock and surficial data, and technical guidance or supply environmental constraints, groundwater data, and permitting requirements.
- **Federal agencies:** USGS, FEMA, NOAA, NRCS, and other agencies who provide national geologic, flood, climate, and soils datasets and web services. Agencies who may use GDMS for hazard response planning, situational awareness, and post-event assessments.
- **Academic and research institutions:** Groups who conduct research, validate machine learning (ML) methods, and support training and peer review.
- **Industry associations and contractors' groups:** Advocates for industry needs, or assist with standards adoption, and provide training.
- **Data providers and historical data custodians:** Organizations and individuals who may be able to provide legacy borehole data such as scans, geophysical surveys, and proprietary datasets.
- **Professional societies and standards bodies:** Organizations such as ASCE, Geo-Institute, TRB, ASTM, and other organizations who offer standards, best practices, and training resources.
- **Funding partners and grant agencies:** Provider of project funding who may require performance reporting and milestones.

- **Public and community stakeholders:** Those who contribute local knowledge, receive project outreach, and influence permitting and acceptance.
- **Software and platform vendors:** Businesses and individuals who supply GDMS components, integration support, licensing, and technical maintenance.

This list reflects the potential iterative nature of a fully developed GDMS with data being both provided to and received from stakeholders (e.g., other state agencies, federal agencies, consulting geotechnical firms). As this list is based on ongoing discussions, inclusion or exclusion of individual stakeholders will depend on the intent, content, and development of the GDMS. The research team understands that not all these stakeholders may be at the forefront as part of the current development process. However, a fully developed GDMS may interact with each of these parties during or following development as the system changes to meet the requirements of end users. As each stakeholder is engaged it is also anticipated that secondary stakeholders will be identified based on the use of overlapping data and relationships.

3.2 Stakeholder Engagement

Engagement of identified stakeholders is recommended to occur in a structure manner to ensure uniform and goal-driven GDMS development. Selection and engagement should focus on prioritization of key stakeholders which are expected to include the geotechnical engineering division, district personnel, project managers, design and review engineers, construction and field operations, maintenance and asset managers, and IT and Enterprise Management.

These stakeholders should be engaged proactively to verify GDMS development focuses on technical and operational feasibility, and long-term suitability; ultimately leading to increased credibility. Engaged stakeholders should be mapped based on their interests, influence or GDMS impact and responsibilities, decision authority, anticipated interaction or communication channels, and anticipated interaction frequency with the GDMS. Establishing a register will provide the foundation for communication and decision-making throughout the GDMS development cycle while maximizing engagement across PennDOT. An example of a stakeholder registry for internal stakeholders is provided in Table 1.

Stakeholder workshops should be held regularly to address concerns or developments in interests, responsibilities, interaction processes, and specifically geotechnical risks and design assumptions such as the key interests identified in Task 1. This collaborative approach will help maintain a real-world oriented GDMS development and align expectations across stakeholders. In addition to specific internal and external engagement, broad public engagement should be pursued through accessible channels such as newsletter and a project website. Demonstrating PennDOT's interest in asset management and cost-effective governance. Engagement should be carefully managed to ensure identification, communication, consultation, and documentation of identified stakeholders, with activities tailored to the needs of different groups. The approach should be iterative, allowing feedback to be incorporated at multiple stages, ensuring resources are directed where uncertainty is greatest.

Table 1. Example internal stakeholder registry for GDMS development.

Stakeholder	Interest / Responsibilities	Decision Authority	Communication Channels	Interaction Frequency
Executive Leadership	Set strategy; approve funding; sponsor program.	Final approval on budget, policy, and statewide rollout.	Executive briefings; steering committee meetings; email.	Quarterly or as-needed
Geotechnical Engineering Division	Define technical requirements; validate geotechnical content; lead QA.	Approve technical standards and acceptance criteria.	Technical workshops; review meetings; GDMS notifications.	Weekly to monthly
District Personnel and Project Managers	Submit exploration requests; implement local workflows.	Local approval of project scopes, schedules, and priorities.	District meetings; project portals; GDMS tickets.	Weekly to monthly
Design and Review Engineers	Use GDMS outputs for design and foundation decisions and QA/QC; perform peer review.	Approve design use of GDMS data; request additional investigations.	Project meetings; design reviews; GDMS queries.	Project-driven
Construction and Field Operations	Provide as-built and construction data; perform field collection.	Approve field data submissions and schedules.	Field briefings; mobile app updates; daily logs.	Daily to weekly during active work
Maintenance and Asset Management	Inventory geotechnical assets; prioritize inspections; plan interventions.	Prioritize maintenance actions and allocate funds.	Asset dashboards; reports; coordination meetings.	Monthly
Surveyors and Mappers	Provide control, LiDAR, and topographic data; ensure georeferencing and spatial databases.	Approve spatial control and coordinate transforms.	GIS services and workshops; technical calls.	Project-driven
Spatial Analysts	Manage spatial layers, mapping services, and visualization components.	Approve data integration and representation methods.	Data transfers; API documentation; GIS services and workshops; technical calls.	Weekly to monthly
IT and Enterprise Management	Host systems; enforce security; manage API integrations; system performance; alignment with PennDOT IT policies.	Approve infrastructure, security posture, and integrations. Access control and feature implementation.	Architecture reviews; control and access; ticketing system.	Weekly to monthly
Traffic and Safety Operations	Coordinate traffic control for field operations.	Approve traffic control plans and schedules.	Traffic coordination meetings; permit systems.	Project-driven
Right of Way and Permitting	Manage property access and utility clearances.	Approve access permits and utility coordination.	Permit workflows; coordination meetings; GDMS links.	Project-driven
Data Governors and Records Management	Define data standards, metadata, retention, and DIGGS compliance.	Enforce data policies and retention rules.	Governance meetings; policy documents; compliance reports.	Monthly
Training and Workforce Development	Develop role-based training and adoption materials.	Approve training curriculum and delivery approach.	Training sessions; LMS; webinars.	Quarterly or as-needed

Developed documentation should be institutionalized as a consistent practice. Meeting minutes, action items, and decision logs should be maintained in a centralized repository, accessible to all stakeholders. Engagement should focus on benefits including early regulatory contact, technical workshops, and transparent documentation. Attention should be given to expanding outreach to newly identified stakeholder groups (e.g., utility owners, municipalities and municipal managers) to help strengthen relationships before and during GDMS expansion.

3.3 Stakeholder Expansion

As each set of stakeholders has been identified, stakeholder roles, responsibilities, and connections are anticipated to provide a natural pathway for expanding the stakeholder list and refining organizational involvement in GDMS development. Each stakeholder (e.g., state unit, consultant, contractor, or regulator) can be considered as a node in a broader network. By examining their functions and dependencies, additional stakeholders and organizations can be identified which may not have been apparent at the outset.

For example, a designer may highlight their reliance on survey data, which points to the inclusion of survey and mapping divisions. Similarly, construction contractors may identify inspection teams or materials testing labs as critical partners; expanding the stakeholder list to include quality assurance groups. If regulatory authorities are included, regulatory agencies may identify secondary stakeholders such as environmental review boards or municipal permitting offices whose data requirements must be integrated into the GDMS. This iterative expansion ensures that the GDMS is developed to reflect the full range of data producers, users, and decision-makers. It may also refine organizational involvement by clarifying which groups are central to GDMS governance, which are primarily data contributors, and which are largely end-users of data. Through the identified engagement strategies (e.g., workshops, interviews, and responsibility matrices) stakeholders also help identify gaps, overlaps, and emerging needs, reducing the burden on PennDOT for GDMS conceptual development and leading to a more comprehensive and resilient stakeholder network.

This process transforms stakeholder mapping from a static exercise into a dynamic tool for system design. By leveraging the knowledge and connections of identified stakeholders, agencies can ensure that the GDMS incorporates diverse perspectives, aligns with operational realities, and gains the legitimacy needed for long-term adoption. This deliberate process may also help identify specific individuals to champion GDMS expansion and development using one or several influence-commitment matrices. Where each individual stakeholder or stakeholder group is placed based on their ability to influence GDMS adoption and their commitment to seeing GDMS implementation. Influence-commitment matrices may also be developed for individual GDMS features which require substantial report (e.g., record digitalization, resource libraries, training). An example influence-commitment matrix is provided as Figure 1.

This example matrix outlines several potential parties including two theoretical districts; the Geotechnical Engineering Division [Div.] has a relatively high influence (y-axis) and high commitment (x-axis) to GDMS adoption, executive leadership hOCRas higher influence for GDMS adoption but less commitment; and theoretical Districts X and Y demonstrate how two districts may vary. In this example District Y is more committed to GDMS adoption as well as greater influence while District X has less commitment and less influence. Based on the

relatively high influence and commitment, the Geotechnical Engineering Division and District Y may both be suitable champions for GDMS adoption. While Executive Leadership may become a suitable champion, their commitment was able to be improved through demonstrating the benefits of the GDMS.

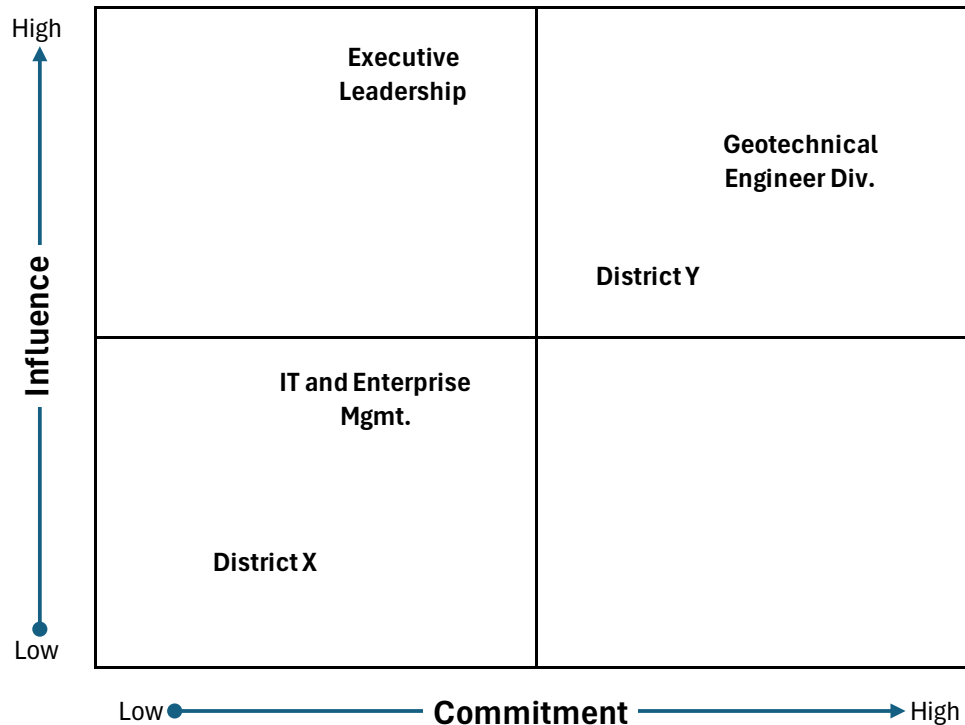


Figure 1: Example influence-commitment matrix for GDMS adoption.

3.4 Return on Investment Evaluations

Evaluating return on investment (ROI) for a GDMS requires a structured approach that accounts for both tangible and intangible benefits. Unlike more traditional capital projects, the GDMS delivers value through improved data governance, enhanced decision-making, and stakeholder collaboration. To capture this, ROI should be assessed across several dimensions.

1. **Direct financial savings can be quantified.** These include reduced duplication of field investigations, lower costs for data storage and retrieval, and efficiency gains from automated ingestion, retrieval, and standardized reporting. For example, district engineers and project managers benefit from faster access to borehole logs and lab results, which shortens design cycles and reduces consultant fees. IT and enterprise management teams realize savings by consolidating geotechnical data into a single, secure platform rather than maintaining fragmented systems.
2. **Productivity and time savings are critical measures.** Stakeholders including geotechnical engineers, surveyors, and construction staff spend less time searching for records or reformatting data, freeing resources for higher-value and critical tasks. Training and workforce development groups may benefit from standardized workflows and role-based training materials, reducing onboarding time and minimizing errors. These

efficiency gains translate into measurable wage savings and improved project delivery schedules.

3. **ROI must capture risk reduction and liability avoidance.** By providing uncertainty visualization and QA/QC workflows, GDMS reduces the likelihood of design errors and construction delays. Executive leadership and asset managers can quantify this benefit by comparing avoided costs of remediation, litigation, or emergency response. Federal and state agencies also benefit from improved hazard assessments and resilience planning, aligning with FHWA and AASHTO risk-based asset management guidance.
4. **Strategic and intangible benefits should be considered.** Consultants and contractors gain from streamlined data submission processes, while external stakeholders such as municipalities and utilities benefit from improved coordination and transparency. While PennDOT currently maintains a well-developed standard for borehole logs, academic institutions and professional societies may see further value in large-scale accessible datasets that support research and training. These benefits, while harder to monetize, can strengthen PennDOT's reputation as a partner in research projects, foster stakeholder trust, and accelerate adoption of standards.

Where possible, ROI evaluation should be stakeholder-driven. A stakeholder registry can be used to map costs and benefits by group, ensuring that each party's contributions and gains are documented. Executive leadership may choose to focus on statewide adoption and funding justification, while district personnel emphasize project-level efficiency. IT staff evaluate system performance and compliance, and contractors assess reduced administrative burden. By aligning ROI metrics with stakeholder priorities, PennDOT can build consensus and demonstrate value across the organization.

4. PILOT RESPONSE AND EXPANSION

4.1 Pilot Response

The current GDMS has been overall responded to with optimism. Across several months of discussions, district engineers and project staff have consistently indicated that the demo showcased strong visualization capabilities, particularly in 3D modeling of subsurface soils. The ability to generate subsurface statistics, visualize borehole variability, and display grout models in 3D was seen as a useful and powerful tool for exploration and communication of project progression and results. At the same time, participants emphasized that while 3D visualization adds significant value, traditional two-dimensional cross-sections remain essential for engineering practice including design. Several concerns were also raised about the risks of interpolating data for visualization, underscoring the need to balance accessibility and tool development with technical feasibility and liability considerations.

Database integration emerged as a recurring consideration with the project team highlighting the advantages of PostgreSQL with PostGIS extensions over ArcGIS Online, citing greater flexibility, fewer constraints, and stronger performance for complex queries. This approach was preferred by PennDOT staff, based on the importance of maintaining control over database structures and ensuring compatibility with existing enterprise systems. The GDMS's ability to connect to multiple databases was seen as a key strength, enabling broader integration across PennDOT's digital delivery ecosystem as detailed in the Task 2 report.

Feedback also focused on data incorporation with districts expressing interest in integrating historic datasets, geotechnical asset management (GAM) program data, and ongoing project files (e.g., borings and CAD surfaces) into the GDMS. JSON and geoJSON formats were identified as promising pathways for long-term storage and display, while automated methods for georeferencing logs and reports are being developed to streamline the inclusion of legacy data. These capabilities were viewed as critical for demonstrating GDMS's value in real-world projects.

Workflow and quality assurance considerations have also been extensively discussed across the project team and PennDOT. PennDOT emphasized the importance of retaining control over QA/QC processes, with draft-to-final workflows already in place for grouting data. Queries within the GDMS are SQL-based, offering transparency and flexibility, although caution must be practiced against open query builders due to security concerns. Accessibility and licensing issues will also have to be addressed, along with the need to address Survey123 seat policies and ensure compliance with Section 508 accessibility standards.

Integration with PennDOT's broader digital delivery plan remains a priority and meetings with IT staff confirmed that neither security nor ESRI tools are expected to pose major hurdles. Preliminary delivery methods are being evaluated for compatibility with PennDOT's current geospatial data systems including those such as PennDOT Multi-modal Project Management System (MPMS) presented as OneMap and District-level pilots, particularly in Districts 4 and 11, are being advanced to demonstrate GDMS's potential benefits. Discussions continue on expanding the system to include geotechnical standards, procedures, and training resources with the GDMS serving as a centralized hub.

Based on the pilot response, there is significant potential in expansion and development of the GDMS. This includes functionality to include automated project tracking, corridor risk flagging, and improved visibility of geotechnical data across districts and the state as a whole. The demo has established a strong foundation, and feedback reflects both satisfaction with its current capabilities and a clear roadmap for enhancements that will make GDMS a valuable part of PennDOT's geotechnical data management and digital delivery strategy. However, future expansion requires deliberate attention to several factors; the system must continue to balance visualization with engineering utility, maintain database flexibility, and establish robust workflows for new and historic and statewide data integration. QA/QC governance, licensing, and accessibility must be formalized to ensure sustainable adoption. Functional enhancements (e.g., project tracking, corridor risk flagging, and a centralized hub for geotechnical standards and training) may further increase the GDMS's value. Equally important, PennDOT must invest in training and change management to encourage district-level adoption and institutionalize GDMS within agency workflows.

4.2 Staged Expansion

As the GDMS pilot is a technical success, validating core functionality and earning positive feedback from engineers and IT staff, a staged expansion is an appropriate next step to expansion. This staged expansion is recommended to begin with standardization and governance, followed by wider district pilots, statewide integration, and finally digital twin optimization. This process is recommended to provide PennDOT a reasonable path to GDMS acceptance while ensuring sufficient control and a sustainable development path. This approach also aligns with FHWA guidance on Transportation Asset Management (TAM)/GAM development and supports PennDOT's digital delivery strategy. For each stage several potential key actions have been provided for consideration.

Stage 1: Standardization and Data Governance

The first stage focuses on establishing consistent data standards and governance protocols. FHWA's Geotechnical Technical Guidance Manual emphasizes the importance of standardized workflows for investigation, analysis, reporting, and archiving. As several of these processes are currently well developed at PennDOT this stage is anticipated to be brief focusing on confirmation of practices and their consistent use across districts and subcontractors. The FHWA Geotechnical Data Management Workshop highlights the adoption of DIGGS (Data Interchange for Geotechnical and Geoenvironmental Specialists) as a schema for digital data exchange which is the primary schema recommended by the research team.

Key actions:

- Implement DIGGS-compliant schemas for all GDMS data.
- Formalize QA/QC workflows to ensure draft-to-final transitions are consistent across districts.
- Establish accessibility and licensing policies (Survey123 seats, Digital Accessibility Policy applicability and compliance mechanisms).
- Create a governance framework for database management, prioritizing PostgreSQL/PostGIS for flexibility and performance.

Stage 2: District-Level Pilots

The second stage expands GDMS through wide-spread district-level pilots. This stage provides a thorough validation of new systems through phased implementation. By engaging district stakeholders this process also provides the opportunity to further receive stakeholder feedback and identify secondary stakeholders and tools which are desirable to GDMS users.

- Districts 4 and 11 should continue to provide project data, including boring logs and CAD surfaces, for integration for both individual projects and historical data sets. The inclusion of data for complete projects (i.e., planning to post construction) is recommended.
- Automated georeferencing workflows should be tested to incorporate historic datasets.
- GDMS must demonstrate its ability to generate 3D point clouds while maintaining 2D cross-sections for engineering use.
- Integration with PennDOT's GAM program processes should be prioritized, ensuring risk-based scoring and asset flagging are incorporated into pilot workflows. If the GDMS is anticipated as the primary tool for the GAM program, cross-stakeholder discussions should be performed as soon as possible to ensure that key stakeholders are satisfied with GDMS/GAM integration.

Stage 3: Statewide Expansion and Enterprise Integration

The third stage embeds GDMS into PennDOT's enterprise systems. FHWA's Transportation Asset Management Plan (TAMP) Guidance requires DOTs to integrate geotechnical data into asset management frameworks, considering lifecycle costs and resilience.

- GDMS should integrate with PennDOT's asset location data system to ensure geotechnical data flows into PennDOT's digital delivery ecosystem.
- Expansion should include statewide geotechnical (i.e. GAM) assets such as slopes, embankments, retaining walls, and hazardous sites not incorporated as part of Stage 1 or 2. This may include some assets which are identified but not monitored as part of other asset programs (e.g., return walls).
- Automated project tracking tools should be deployed to flag delays, boring changes, and corridor risks.
- A centralized GDMS hub should be developed to house geotechnical standards, procedures, and training resources.
- Training and change management programs should be implemented to encourage district-level adoption and institutionalize GDMS workflows.
- Development and implementation of consultant contract language incorporating digital data delivery and workflows should be fully implemented across PennDOT projects.

Stage 4: Long-Term Optimization and Digital Twin Integration

The final stage positions GDMS as part of PennDOT's digital twin environment, consistent with FHWA and AASHTO guidance on digital delivery.

- GDMS should evolve to support digital twin applications for maintenance, operations, and construction.
- Resilience and extreme weather considerations should be incorporated into geotechnical asset management, aligning with FHWA's risk-based asset management requirements.
- Continuous improvement processes should refine GDMS queries, visualization tools, and interoperability with BIM platforms.

4.3 Stakeholder Engagement

During the staged expansion of the GDMS stakeholder engagement should be maintained. Effective stakeholder engagement is critical to the staged expansion of the Geotechnical Data Management System (GDMS). FHWA's Transportation Asset Management Plan (TAMP) Guidance and AASHTO's DIGGS Implementation Framework emphasize that both internal and external stakeholders should be involved at each stage to ensure technical compliance, operational usability, and long-term adoption. Engagement should be structured to progressively broaden participation, moving from core technical teams to statewide and external partners. Recommended methods for stakeholder engagement are summarized below. However, external stakeholder involvement, particularly relating to federal agencies, may not be reliable or available depending on the GDMS deployment schedule.

Stage 1: Standardization and Governance

The next stage of engagement, like the pilot, focuses primarily on internal stakeholders to establish standards and governance. Both internal and external stakeholders may be engaged through technical workshops, schema validation sessions, and governance committee reviews.

- **Internal stakeholders:** PennDOT IT, geotechnical engineers, database administrators, and QA/QC staff. These groups define DIGGS-compliant schemas, formalize QA/QC workflows, and ensure accessibility and licensing compliance.
- **External stakeholders:** FHWA technical liaisons and AASHTO DIGGS working groups provide guidance on national standards and interoperability. Consultants and system vendors may be engaged to validate database architecture and compliance with Section 508 accessibility requirements.

Stage 2: District-Level Pilots

Stage two engagement relies primarily on district engineers and project teams to validate GDMS functionality in real-world contexts. Engagement for district-level pilots largely reflects the current pilot effort with pilot demonstrations, feedback sessions, and iterative refinement meetings. Additional district-level training may be required to ensure engineers can use GDMS effectively and to ensure documentation of recommended revisions and bug reports

- **Internal stakeholders:** District geotechnical engineers, project managers, and construction staff provide pilot data (borings, CAD surfaces) and feedback on

visualization tools. PennDOT IT supports additional integration with existing systems.

- **External stakeholders:** Consultants and contractors contribute project datasets and validate workflows. FHWA regional offices may observe pilots to ensure alignment with national asset management practices.

Stage 3: Statewide Expansion and Enterprise Integration

Following development and confirmation of GDMS performance at the district level additional executive stakeholder involvement is anticipated along with an increase in external stakeholders. As the GDMS is anticipated to be largely state-wide by Stage 3 additional workshops, integration sessions, and formal training programs are anticipated. A review by FHWA partners may also be suitable to gain feedback on MAP-21 compliance to the extent required by law.

- **Internal stakeholders:** PennDOT leadership, asset management teams, and geospatial database administrators oversee integration into enterprise workflows. District engineers continue to provide feedback on statewide adoption. Training and change management teams lead institutionalization efforts.
- **External stakeholders:** FHWA asset management specialists provide oversight to ensure GDMS aligns with TAMP requirements. Industry partners (consultants, contractors, software vendors) support integration of statewide geotechnical assets such as slopes, embankments, and retaining walls.

Stage 4: Long-Term Optimization and Digital Twin Integration

Stage four, is the final stage for the GDMS. Goals and engagement are targeted towards improved data integration into PennDOT workflows. Targeting the development of a digital twin environment which may allow significant improvements to project performance. As this field is still developing, engagement is anticipated to include strategic planning sessions, innovation roundtables, and continuous improvement reviews. Collaboration with FHWA and AASHTO is also recommended where possible to validate GDMS compared to other national digital delivery initiatives.

- **Internal stakeholders:** PennDOT's innovation office, IT architects, and project leadership drive digital twin integration. District engineers provide feedback on system resilience and extreme weather considerations.
- **External stakeholders:** FHWA innovation programs, AASHTO BIM for Infrastructure working groups, and academic partners contribute expertise on digital twin applications, resilience modeling, and advanced visualization. Contractors and consultants adapt workflows to integrate GDMS outputs into BIM platforms.

4.4 Development Schedule

The four stages of engagement are based on strategic outcomes and stakeholder participation which may not be directly tied to a schedule. However, based on the rate of expansion for similar geotechnical and asset management programs the project team has anticipated that the outlined stages can be approximated across a conservative timeline of seven years, reflecting the level of

effort and iterative nature of GDMS development. With this period subdivided into the near-term (1 to 2 years), mid-term (2 to 5 years), and long-term (greater than 5 years). These periods are expected to correspond to Stage 1 and Stage 2 for the near-term, Stage 3 for the mid-term, and Stage 4 for the long-term.

Near-Term (1 to 2 Years)

In the immediate future, the research team recommends that the GDMS serves as a vehicle for digitization and standardization. The priority is to establish consistent data structures and workflows that align with national standards such as DIGGS. By implementing DIGGS-compliant schemas, PennDOT ensures interoperability across consultants, contractors, and both internal and external systems.

GDMS development is anticipated to support QA/QC governance, formalizing draft-to-final transitions for boring logs, grouting data, and other field records. Pilot projects demonstrate the ingestion of JSON and geoJSON files, enabling visualization of borings, CAD surfaces, and point clouds are also a reasonable expectation. Accessibility and licensing issues (e.g., Section 508 compliance, Survey123 seat policies) should be resolved to ensure reasonable compliance and availability for GDMS users. These goals also align with FHWA's Geotechnical Technical Guidance Manual and early digital delivery initiatives. Revision to PennDOT publications (e.g., Publication 222 – Geotechnical Investigation Manual) and other organization documents which impact behaviors or workflows impacted by the GDMS should be revised to reflect GDMS inclusion.

Mid-Term Applications (2 to 5 Years)

As GDMS matures, its role is recommended to expand to statewide asset management. PennDOT can leverage the system to inventory geotechnical assets such as slopes, embankments, retaining walls, hazardous sites, and natural hazards otherwise excluded from current management practices. This expansion is similar to practices outlined in the Louisiana DOTD Geotechnical Asset Management Guide and FHWA's Transportation Asset Management Plan (TAMP) Guidance.

Integration with PennDOT's GAM program would enable risk-based scoring, consistent with NCHRP Report 903's framework for prioritizing geotechnical risks; allowing improved corridor-wide planning and emergency response. The GDMS is also expected to connect with enterprise systems such as the PennDOT OpenMap, ensuring geotechnical data flows seamlessly into PennDOT's broader digital delivery ecosystem.

Functional enhancements to the GDMS may include automated project tracking to flag delays, boring changes, or corridor risks, reducing burdens on district engineers. Training and change management programs institutionalize GDMS workflows across districts, supported by FHWA's Data Governance Primer.

Long-Term Applications (5+ Years)

In the long term, GDMS is targeted as a cornerstone of PennDOT's digital twin environment. By embedding geotechnical data into dynamic, continuously updated infrastructure models, GDMS

can be used to supported predictive analytics, resilience planning, and lifecycle optimization. This may include the ability to more easily simulate slope stability, settlement, and groundwater impacts under varying conditions, providing proactive insights for asset management. The GDMS may also incorporate resilience considerations for risk-based and climate-resilient asset management.

Interoperability with Building Information Modeling (BIM) platforms, as outlined in AASHTO's BIM for Infrastructure Roadmap, also provides a potential framework for GDMS output integration with construction sequencing, operations, and maintenance. Continuous improvement processes refine queries, visualization tools, and integration pathways, positioning GDMS as a fully embedded component of PennDOT's digital delivery strategy.

The research team understands that digital twin integration may be feasible earlier in the development and deployment cycle. If digital twin development is accelerated it is recommended to follow a similar development program starting with a small-group pilot, progressing through district-level expansion, and culminating at state-level expansion. This would allow for measured engagement including a feedback and correction cycle critical to successful program.

Commentary

The staged development of the GDMS can aid PennDOT in the institutionalization of geotechnical data management during a period of reduced staffing and uncertain long-term funding while maintaining alignment with FHWA and AASHTO guidance. Providing both immediate benefits and long-term value.

5. DATA COLLECTION AND INTEGRATION

Task 2 and Task 3 reports emphasize that field collection is most suitable when vendor-neutral, API-driven, and validated at the point of entry. This permits PennDOT and external field crews to use platforms such as BoreDM, RSLog, OpenGround, or Survey123, all of which support RESTful (Representational State Transfer) integration. Ensuring that the majority of tools available for field collection are acceptable for use with the GDMS. This is due to the GDMS enforcing schema compliance and quality checks before data is accepted. High-precision GPS/GIS devices complement these platforms by providing accurate coordinates for boring locations and geotechnical features, which is critical for spatial queries and 3D visualization. Task 2 and Task 3 reports also highlight the importance of embedding photographs, sketches, and metadata directly into field submissions, creating a richer dataset that can be validated and queried later.

Legacy data collection is equally important, given the vast archives of historical boring logs, reports, and geophysical surveys. Automated OCR and parsing tools can extract structured information from scanned PDFs, but expert review remains necessary to confirm stratigraphy, depths, and units. Metadata indexing—project identifiers, dates, and spatial footprints—ensures that historic data can be discovered and cross-referenced. PennDOT’s implementation plan stresses that legacy data should be normalized into DIGGS-compliant formats, while retaining provenance and original values for traceability as detailed in the Task 2 report.

National guidance reinforces these practices. FHWA’s Geotechnical Site Characterization manual stresses that field data should be collected with standardized forms, georeferenced coordinates, and immediate quality checks to reduce errors and duplication. TRB’s NCHRP Synthesis 446 on geospatial technologies in transportation projects highlights the growing use of mobile GIS and LiDAR for rapid field capture, but cautions that large datasets require structured workflows and validation to remain useful. NASEM reports on digital delivery emphasize that data collection must be designed for interoperability, with schema alignment (such as DIGGS) and metadata standards to ensure that information can flow across systems without loss of meaning.

5.1 New Field (Boring) Records

The collection of new geotechnical field records is essential to the integrity of modern infrastructure design and management. Subsurface conditions are inherently variable, and reliance on outdated or incomplete datasets introduces uncertainty that can compromise both engineering decisions and public safety. By gathering current field data—through borings, sampling, in-situ testing, and geophysical surveys; engineers obtain site-specific information that directly informs foundation design, slope stability analysis, and construction planning. These records reduce reliance on assumptions, validate geotechnical models, and provide the confidence necessary to mitigate risk.

Beyond immediate design applications, new field records play a critical role in asset management and lifecycle planning. They capture changes in subsurface conditions over time, such as groundwater fluctuations or soil degradation, enabling agencies to prioritize interventions and allocate resources efficiently. In this way, geotechnical data becomes a cornerstone of

risk-based asset management, supporting resilience against extreme weather events and aligning with national guidance on sustainable infrastructure stewardship. The collection of new records also ensures compliance with evolving data standards. When captured in digital, georeferenced formats such as DIGGS, these records integrate seamlessly with GIS platforms, BIM environments, and enterprise asset management systems. This interoperability strengthens institutional knowledge, facilitates collaboration across agencies and contractors, and ensures that geotechnical information remains accessible and authoritative for future generations of engineers.

New Record Integration Workflow

- 1. Pre-planning and project setup** Before crews mobilize, project managers define the exploration program in the GDMS. Existing subsurface data, borings, geologic maps, hazard inventories, and environmental overlays, are reviewed in the system to avoid duplication. Planned boring locations are georeferenced against PennDOT's OneMap, or similar geodatabase, and Survey/Mapping LiDAR surfaces. This ensures that crews have accurate coordinates and context before entering the field.
- 2. Digital form preparation** Standardized digital forms are generated through Survey123, BoreDM, or other supported platforms. These forms are DIGGS-compliant, meaning they enforce consistent fields for boring IDs, depths, sample intervals, test types, and metadata. The forms are pre-loaded onto tablets or mobile devices, with dropdowns and validation rules to reduce entry errors.
- 3. Field mobilization and real-time capture** Crews collect data directly into the digital forms while drilling or sampling. High-precision GPS devices record coordinates and elevations. Photos of samples, site sketches, and notes are attached to the record in real time. Because the GDMS supports RESTful integration, data is transmitted immediately when connectivity is available, or stored locally for later synchronization.
- 4. Inline validation and QA/QC** As data is entered, the system applies automated checks: ensuring depths are logical, units are consistent, coordinates fall within expected ranges, and required fields are complete. Records failing validation are flagged for correction before submission. This "validate at the source" approach reduces downstream cleanup and aligns with FHWA's recommendation for immediate quality control during site characterization.
- 5. Synchronization with GDMS** Once validated, the data is transmitted through RESTful APIs into the GDMS database. Each submission is tagged with project metadata, timestamps, user credentials, and data status (e.g., draft, reviewed, approved). The GDMS automatically indexes the data spatially, linking it to project boundaries and existing records. This ensures that new borings or tests are discoverable alongside historic data.
- 6. Supervisor review and approval** District geotechnical engineers or designated reviewers access the GDMS to confirm the accuracy of submitted records. They can approve, reject, or request corrections. Approved records lose their "draft" status and become part of the authoritative dataset. This mirrors PennDOT's QA/QC workflow described in Task 2 and Task 3, where grouting or pile data is initially marked "draft" until reviewed.
- 7. Integration with external layers** Once approved, the data is available for visualization against external sources such as DCNR bedrock geology, USGS topography, FEMA flood zones, or PennDOT roadway assets. This integration provides immediate context for planning and hazard assessment, without requiring crews to duplicate external datasets.

8. Archiving and metadata management All field records are stored with provenance information including: who collected the data, when, with what equipment, and under what project. This metadata ensures traceability and supports future uncertainty quantification. Legacy records can be linked to new field data, creating a continuous subsurface history for each location.

Workflow Importance

Geotechnical site characterization must be systematic, with clear documentation, georeferencing, and QA/QC at every stage. By following this workflow, PennDOT field crews or their consultants not only collect data efficiently but also ensure that every record is validated, traceable, and immediately useful within the GDMS. This reduces duplication, improves confidence in subsurface models, and supports statewide digital delivery initiatives.

5.2 Historical Boring Records

Historical boring records are often stored in diverse formats; scanned PDFs, gINT files, spreadsheets, and legacy reports. To make these records usable within a GDMS, the process begins with digitization and extraction. Optical character recognition (OCR) and form recognition tools can parse scanned logs, extracting key attributes such as boring ID, coordinates, depths, stratigraphy, and test results. However, automated extraction alone is insufficient; expert geotechnical review is recommended to confirm stratigraphic interpretations, unit consistency, and georeferencing accuracy. Once extracted, the data must be normalized into a standardized schema, such as DIGGS. This ensures that borehole locations, lithology intervals, and test results are represented consistently across projects and districts. This process may include the revision of coordination reference systems (CRS), terminology, units, and other data found to be variable across data providers. Metadata including project identifiers, collection dates, and source documents should be preserved to maintain provenance. As detailed in the Task 2 report legacy data is recommended to be indexed spatially within a PostGIS database, enabling queries by location, formation, or project lineage.

Quality assurance is critical for historical record integration. Automated validation routines are quickly becoming a well-established and reliable standard when trained on local (i.e. PennDOT) data. These routines can check for logical depth sequences, unit conversions, and coordinate ranges, while human reviewers confirm uncertain records. Records failing validation should be flagged for correction rather than discarded, preserving transparency. This process also improves the rate for data to be georeferenced and quality-checked to ensure reliability in downstream analysis. Metadata tagging and structured workflows are recommended to be well documented and consistently applied due to the anticipated size of the GDMS dataset with QA/QC directly embedded into the ingestion pipelines.

The final step is integration into visualization and analysis tools. Once normalized and validated, historical borings can be displayed in 2D maps or 3D subsurface models, linked to external datasets such as DCNR geology or USGS topography. This allows engineers to query historic subsurface conditions alongside new investigations, reducing duplication and improving planning.

Historical Boring Record Integration Workflow

This workflow is provided as a recommended process to convert historical and legacy subsurface information (e.g., PDFs, gINT, spreadsheets, CAD tables) into validated, spatially indexed, DIGGS-aligned records within the GDMS. It balances automation with expert review, preserves provenance, and enables immediate query and visualization.

Step 1: Source discovery and scoping

- Inventory sources: Identify district archives, ProjectWise repositories, consultant deliverables, and public portals holding boring logs, test data, and reports.
- Prioritize datasets: Select projects with clear georeferencing, complete logs/tests, and strategic coverage (e.g., corridors, districts, hazard areas).
- Define scope: Set batch size, timelines, and acceptance criteria (formats, completeness, coordinate references).

Step 2: Document intake and cataloging

- Ingest files: Pull PDFs, gINT databases, spreadsheets, CAD exports into a staging area.
- Catalog metadata: Record project ID, location, dates, authors, file type, and known coordinate systems.
- Assign IDs: Create persistent ingestion IDs linking source files to future normalized records.

Step 3: Automated extraction and initial normalization

- PDF/OCR parsing: Extract boring IDs, station/offset, coordinates, depths, stratigraphy, sample/test tables, and groundwater observations.
- gINT/CSV parsing: Map source fields to target schema; standardize units, data types, and coded values.
- Compute locations: Convert station/offset to coordinates using project baselines; validate or derive elevations from LiDAR/design surfaces.
- Normalize attributes: Align lithology codes, test names, and units to PennDOT/DIGGS conventions.

Step 4: Spatial validation and enrichment

- Coordinate checks: Verify CRS, datum, and precision; flag out-of-bounds locations and improbable elevations.
- Topology consistency: Ensure depth sequences, interval overlaps, and boring completion depths are logical.
- Contextual enrichment: Attach project extents, formation overlays (e.g., DCNR), flood zones (FEMA), and roadway assets for later queries.

Step 5: Quality assurance and expert review

- Automated rules: Enforce required fields, range checks, unit consistency, schema compliance (DIGGS), and lithology code validity.
- Geotechnical review: Confirm stratigraphy, corrections to depths/units, coordinate confidence, and special conditions (karst, fills, contamination).
- Provenance retention: Preserve original values alongside normalized values; log all transformations and reviewer decisions.

Step 6: DIGGS-aligned data modeling

- Entity mapping: Populate Projects, Borings, Samples, FieldTests, LabTests, LithologyIntervals, Groundwater, and Documents entities per DIGGS.
- Relationships: Link samples to borings, tests to samples/intervals, and documents to investigations; maintain external persistent identifiers.

- Versioning: Store record history, approvals, and change notes; mark uncertain records with confidence flags.

Step 7: Load to GDMS and index

- Database insertion: Write validated records to PostgreSQL/PostGIS tables; apply spatial indexes (GiST) for points/lines/polygons.
- Full-text indexing: Index document text for search by formation, material, or parameter names.
- Access controls: Assign object-level permissions based on project and role; mark status (draft, approved, archived).

Step 8: Verification and publication

- Spot checks: Run sample spatial queries, cross-sections, and tabular reports to verify integrity.
- UI exposure: Publish records to the map, tabular views, and query builder; link to source documents.
- User feedback loop: Capture DGE comments; route corrections through controlled updates and re-approval.

Step 9: Ongoing maintenance

- Incremental batches: Repeat ingestion in manageable waves; monitor validation metrics and error rates.
- Policy enforcement: Apply retention, archival, and correction policies; maintain schema conformance over time.
- Performance tuning: Optimize queries, caching, and indexes as volumes grow; adjust pipelines for new formats.

5.3 Historical Construction Records

Construction records, particularly as-builts, provide critical information about how infrastructure was actually built, including foundation details, material properties, and subsurface treatments. These records often exist as CAD drawings, PDFs, or project reports. To integrate them into the GDMS, the process begins with conversion into structured geospatial formats. CAD surfaces and boring data can be exported into GeoJSON or shapefiles, which are then ingested into the GDMS database. PennDOT's Task 3 pilot demonstrated this by converting project files into 3D GeoJSON for long-term storage and visualization.

As-builts should be linked to corresponding geotechnical records. For example, pile driving logs can be associated with boring data, grouting volumes can be tied to karst remediation records, and instrumentation can be linked to construction progress and boring data. This relational linkage ensures that construction data enriches the subsurface model rather than existing as isolated documents. Metadata such as project ID, contractor, date, and construction method must be captured to preserve provenance and enable lifecycle tracking.

As with other records, quality assurance plays a central role to record digitalization and use. Draft construction records should be marked as provisional until reviewed and approved by district engineers. PennDOT's QA/QC workflow, described in Task 2 and reinforced in meeting notes, uses draft stamps that are removed after validation. This ensures that only verified construction data enters the authoritative dataset.

Once integrated, as-builts may be used to support powerful analytical analyses. They may be used for engineers to compare design intent with actual construction (e.g., confirm redlines), assess performance of foundations and treatments, develop improvements to future designs, and support maintenance planning. The impact of these potential uses have been identified across several sources including FHWA's Geotechnical Engineering Circulars which emphasize the importance of documenting construction parameters (e.g., pile lengths, grout volumes) for future performance evaluation. NASEM's lifecycle data management guidance also details historical construction records indicating that construction records are essential for linking design, construction, and maintenance phases into a continuous digital record.

Historical Construction Records Integration Workflow

This workflow converts construction documentation (CAD/PDF as-builts, pile driving logs, grouting summaries, instrumentation records) into structured, spatially referenced, and relational GDMS entities linked to investigations and assets. It preserves the design-to-construction digital thread and supports lifecycle analytics.

Step 1: Source identification and scope definition

- Locate records: Identify as-built drawings, redlines, final reports, daily logs, machine data exports, and material tickets.
- Prioritize systems: Focus on foundation elements (piles, drilled shafts), ground improvement (grouting, DSM), walls, and instrumentation.
- Scope criteria: Select projects with adequate georeferencing and linkable IDs to investigations or assets.

Step 2: Intake, cataloging, and metadata capture

- Ingest files: Pull CAD (DWG/DXF), PDFs, CSV logs, and OEM exports into staging.
- Metadata capture: Record project ID, structure ID, element types, contractors, dates, coordinate systems, and sheet indices.
- Element identifiers: Establish persistent IDs for constructed elements to support cross-references.

Step 3: Geospatial conversion and normalization

- CAD to GIS: Convert CAD layers for foundation elements and surfaces to GeoJSON/shapefiles; assign CRS and validate geometry.
- Log parsing: Extract installation attributes (lengths, blows, set, grout pressures/volumes, tremie depths, concrete volumes).
- Unit alignment: Standardize units and nomenclature to PennDOT/ASTM/AASHTO conventions; normalize codes for methods and materials.

Step 4: Relational linkage to geotechnical entities

- Linking logic: Connect constructed elements to borings, samples, lithology intervals, and tests where relevant (e.g., pile locations to nearby borings).
- Asset association: Attach elements to bridges/roadway assets (BMS/RMS IDs) and project entities in GDMS.
- Document ties: Link as-built sheets and daily logs to specific elements and installation events.

Step 5: QA/QC and approval workflow

- Automated checks: Validate geometry within project bounds, compare installed vs. design parameters, check plausible ranges (e.g., grout pressure/volume).

- Engineer review: District reviewers confirm element attributes, reconcile discrepancies, and approve records; provisional status removed upon approval.
- Audit trail: Record reviewer notes, changes, and final status; preserve source files and parsed outputs.

Step 6: Data modeling and persistence

- Construction entities: Populate tables for FoundationElements, InstallationEvents, TreatmentOperations (e.g., grouting), Materials, and Instrumentation.
- Temporal data: Store time-series where applicable (e.g., pressure vs. volume, blow counts vs. depth), with timestamps and equipment IDs.
- Status and lifecycle: Track design, installed, and verified states; enable updates for post-construction modifications.

Step 7: Load to GDMS, index, and secure

- Database insertion: Write records to PostgreSQL/PostGIS with appropriate geometry types (points/lines/polygons) and temporal attributes.
- Indexing: Apply spatial and temporal indexes; enable full-text search on document metadata.
- Access control: Set permissions by project/role; protect sensitive contractor/OEM data as needed.

Step 8: Verification, visualization, and analytics

- Cross-checks: Compare design vs. as-built attributes; flag deviations for review.
- Visualization: Display elements and treatments in 2D/3D (e.g., piles with lengths, grout columns with volumes), overlay with subsurface and asset layers.
- Reporting: Generate standardized construction summaries, compliance reports, and export packages (DIGGS, CSV, GeoJSON).

Step 9: Integration with monitoring and maintenance

- Instrumentation links: Tie piezometers, inclinometers, settlement plates to constructed elements; ingest readings and maintain series.
- Performance loops: Track post-construction behaviors and correlate with installation parameters for lessons learned.
- Lifecycle updates: Record maintenance/rehabilitation events; update element status and associated documentation.

Step 10: Continuous improvement and governance

- Policy refinement: Update submission standards, required fields, and QA rules based on user feedback and audit findings.
- Training: Provide role-specific guidance for contractors, inspectors, and DGEs on digital capture and approvals.
- Sustainability: Schedule backups, retention, and schema evolution; keep alignment with DIGGS/ASTM/AASHTO and PennDOT enterprise systems.

5.4 GDMS Maturation

When integrating historical boring records and construction as-builts into the GDMS, the concept of implementation maturity should be used. Implementation maturity refers to how far along the organization (i.e., PennDOT) is in adopting structured, validated, and interoperable workflows. Early stages focus on demonstration and proof of concept, and mature stages emphasize scalability, governance, and lifecycle integration. A mature implementation is only

about technology; it is about institutionalized practices resulting in the system being reliable, repeatable, and trusted. Four fundamental indicators for maturity are recommended for consideration:

1. **Scope of integration:** Early implementations may only ingest a small set of legacy boring logs or a single pilot project's construction records. Mature implementations expand to statewide coverage, with automated pipelines that continuously ingest new data and reconcile it with existing archives.
2. **Quality assurance and governance:** At lower maturity, QA/QC may be ad hoc, relying on manual checks. At higher maturity, automated validation rules are embedded in ingestion pipelines, with role-based approvals and audit trails ensuring accountability.
3. **Interoperability and standards:** Initial systems may store data in proprietary formats. Mature systems enforce DIGGS compliance, ASTM/AASHTO standards, and metadata alignment, ensuring interoperability with consultants, contractors, and other DOTs.
4. **Lifecycle integration:** Early systems focus on investigation data only. Intermediate systems add the potential for correlations between construction test data and sampled subsurface material, lithology, and geologic strata. Mature systems integrate boring records, construction as-builts, and maintenance data into a continuous digital thread, supporting asset management and performance monitoring.

Example: Historical Borings

At an early maturity stage, PennDOT might digitize a handful of scanned boring logs from District 4, using OCR to extract depths and lithology. The records are stored in spreadsheets, with limited georeferencing and minimal QA/QC. Engineers can view them, but confidence is low, and duplication persists.

At a mature stage, PennDOT establishes an automated pipeline that ingests thousands of legacy boring logs from district archives, ProjectWise and consultant archives. Each record is parsed, normalized into DIGGS schema, and spatially indexed in PostGIS. Automated rules check logical depth sequences and unit consistency, while district geotechnical engineers review flagged records. Metadata ensures provenance, and the records are immediately queryable in both 2D and 3D visualization tools. This maturity level transforms static archives into a dynamic datasets.

Example: Construction As-Built

At an early maturity stage, as-built drawings for a bridge foundation are stored as PDFs in district files. Pile driving logs are scanned and attached to the project folder, but they are not linked to boring data or accessible in the GDMS. Engineers must manually cross-reference documents, which is time-consuming and error-prone.

At a mature stage, CAD as-builts are converted into GeoJSON and ingested into the GDMS. Each pile element is linked to nearby borings, with installation attributes (length, blows, set) stored as structured records. Grouting volumes for karst remediation are captured as time-series data, linked to treatment operations, and visualized in 3D. QA/QC workflows ensure that draft records are reviewed and approved before becoming authoritative. The GDMS now supports

lifecycle analysis: engineers can compare design intent with actual construction, assess performance, and plan maintenance based on verified as-built data.

6. GDMS TOOLS AND UTILIZATION

As outlined in previous sections the GDMS may be used to varying levels of complexity depending on the stage of implementation. This reflects that the GDMS is intended as a ‘living’ system which adapts to its current user-base and reflects the development and needs of PennDOT. Considering these stages of implementation through a framework of simple to extensive or detailed implementation provides an outline for how data may be used across stages of development. These uses are cumulative with simple implementation uses remaining present during later stages of development. These stages can be presented as three stages of complexity: simple, moderate, and extensive or complex. These stages or conditions are briefly summarized below followed by a more detailed explanation and examples.

- Simple: Archive and retrieval, with metadata enforcement.
- Moderate: Structured schema, QA/QC, and visualization, with version control and API integration.
- Extensive/Complex: Lifecycle digital twin, integrating investigations, construction, and maintenance, with uncertainty quantification and real-time monitoring (ASCE, ISSMGE).

6.1 Simple Use Condition: Archive and Retrieval

At the simple use condition, a GDMS provides a centralized repository for historical boring logs, geotechnical reports, and scanned PDFs. The primary benefit is reducing duplication of investigations and improving accessibility. Engineers can query by project name, location, or date, retrieving past records without searching district archives. Visualization is limited to 2D maps showing boring locations.

FHWA Geotechnical Technical Guidance Manual details that even basic repositories should enforce consistent metadata (e.g., project identifiers, coordinates, and dates) to ensure records are discoverable and traceable. The Geotechnical Data Management Workshop provided further recommendations including that DOTs begin with simple electronic archiving but adopt DIGGS or equivalent schema early to allow basic records to be migrated into structured formats on an as-needed basis.

6.2 Moderate Use Condition: Structured Data and Visualization

At a moderate use condition, the GDMS evolves into a schema-compliant, validated system. Historical records are normalized into DIGGS format, enabling consistent queries across projects. Field data collection is digitized through Survey123, BoreDM, or similar tools with borings entering the system in real time. Automated QA/QC checks enforce unit consistency, logical depth sequences, and coordinate accuracy, while district reviewers approve records before they become authoritative. Visualization expands to include interactive maps and simple 3D subsurface models, overlaying boring data with geology, flood zones, DOT utilities, and other assets.

Active work in this area is underway as part of NCHRP Synthesis 56-23 which focuses on integrating geotechnical data into digital delivery. This research develops on several existing standards from FHWA and AASHTO, state GDMS experience, and guidelines from NCHRP

with a focus on moderate-use systems. These moderate systems are recommended to include version control, traceability, and API-based integration to ensure data reliability and interoperability. The National Academies' Geotechnical Information Practices in Design-Build Projects also highlight the importance of providing structured geotechnical information in bid documents, reinforcing the need for schema compliance and QA/QC.

6.3 Extensive/Complex Use Condition: Lifecycle Digital Twin

At its most advanced level, the GDMS can become a fully integrated digital delivery platform, linking investigations, construction records, and maintenance data into a continuous lifecycle model. Historical borings should be digitized and spatially indexed; construction as-builts (e.g., pile driving logs, grouting volumes, CAD surfaces) are converted into structured geospatial records; and maintenance data (e.g., instrumentation readings, rehabilitation events) are tied directly to assets. The system supports complex geostatistical analysis and uncertainty quantification, enabling visualization of confidence ranges in subsurface models. Advanced 3D visualization tools allow interactive exploration of underground features, while relational linkages connect boring data to constructed elements and performance monitoring.

The development of 4D visualization and time-domain integration is a strong indicator of an extensive or complex use condition and should be a goal for a fully developed GDMS. Time-domain programs are a developing and important part of asset management in the future due to the increase in extreme events, land-use changes, and the need to better prioritize limited funds for system maintenance and improvement. This move towards 4D management is detailed by both NASEM and ASCE who emphasize that extensive systems should evolve into digital twins, integrating geotechnical data with BIM and asset management. The role of digital twins in linking investigations, construction, and maintenance, with real-time instrumentation feeding into 4D models plays a critical role in evaluating site performance and land-use changes. Digital geotechnical information management in a lifecycle perspective stresses that adding as-built geotechnical information to databases enables digital asset management and improves long-term suitability and system resiliency.

6.4 Example Use Cases

To help explain the potential use of the GDMS for individual cases four examples across the levels of complexity were prepared: one each for simple, moderate, and complex GDMS implementations.

Example 1: Locating Historical Borings for a Small Culvert Project (Simple Implementation)

A district engineer is tasked with replacing a small culvert. Instead of sending crews to drill new borings immediately, they log into the GDMS and search by location. The system returns scanned PDFs of boring logs from a nearby project completed several years prior and several borings for the existing culvert structure. The engineer can quickly review soil stratigraphy and groundwater notes, even though the records are not yet normalized into DIGGS.

This simple implementation provides value by consolidating archives into one searchable repository. Even basic electronic archiving reduces duplication and improves efficiency. At this

stage, the GDMS is essentially a digital repository or filing cabinet. Serving limited ‘smart’ features but acting to save time and improve project efficiency.

Example 2: Planning a Bridge Replacement with Integrated QA/QC (Moderate Implementation)

A design team is preparing for a bridge replacement. They query the GDMS for nearby borings and find normalized records stored in DIGGS format. Field crews collect new borings using boring log software (e.g., gINT, OpenGround, BoreDM, etc.), which transmits data directly into the GDMS. Automated QA/QC checks the newly entered data against data collection standards and nearby historical- and project-borings for material and geologic consistency. One record is flagged with inconsistent depth intervals and bedrock elevation, which the district geotechnical engineer reviews and corrects as-needed before approving. The team then uses the GDMS’s interactive map to overlay boring data with DCNR bedrock geology and FEMA floodplain layers. They visualize soil stratigraphy, compare it with flood risk zones, and identify potential preferred foundations based on soil data and foundation challenges. Based on the inconsistent bedrock elevations, additional borings may also be assigned. Bedrock variability is used to provide an improved estimate of pile lengths and potential overruns.

This moderate implementation presents the benefits of the GDMS when using structured workflows, metadata tagging, and API-based integration for managing large geospatial datasets. At this level, the GDMS serves as a system greater than a repository; serving as a validated interoperable system supporting planning and preliminary design.

Example 3: Karst Remediation with Lifecycle Integration (Complex Implementation)

PennDOT undertakes a karst remediation project involving sinkhole stabilization. The GDMS integrates multiple data streams:

- Historical sinkhole records and boring logs digitized into DIGGS schema.
- Maintenance records for the roadway including pavement overlays, patching, repairs, and surface elevation information are overlaid as geospatial records as time-series data.
- Construction as-builts, including grouting volumes and pressures, converted from CAD and CSV logs into structured geospatial records.
- Ongoing piezometer readings linked to the treated area, stored as DIGGS structured time-series data.
- Engineers visualize the magnitude and rate of change in site conditions based on surficial and subsurface indicators. The remediation is visualized in 3D using Motar3D, overlaying grouting columns with surface conditions, lithology, and groundwater data. They use uncertainty quantification tools to assess confidence in subsurface models, and performance monitoring data feeds back into the GDMS to track long-term effectiveness.

This complex implementation reflects digital delivery and lifecycle data management, which emphasize integrating investigations, construction, and maintenance into a continuous digital thread. At this stage, the GDMS functions as a digital twin, enabling predictive maintenance, risk-based asset management, and agency-wide adoption in addition to the benefits provided by the simple and moderate implementations.

Example 4: Highway Slope Validation and Risk Screening (Moderate-Complex Implementation)

PennDOT is evaluating a poorly performing slope along a rural highway. The GDMS record shows borehole data indicating clayey soils with moderate plasticity. When overlaid with rainfall intensity maps, vegetation changes, seasonal impacts, and historical landslide records, GDMS flags the slope as being in a high-risk zone. Instrumentation data stored in GDMS shows rising pore pressures during storm events and in the early spring, validating the hazard exposure and periods of increased exposure. At the same time, GDMS cross-checks construction compaction records and conditions during construction and finds they meet specifications, ruling out poor construction as the cause. This validation allows PennDOT to prioritize the slope for remediation in its GAM framework and to perform further root-cause analyses as part of remediation planning.

6.5 Expansion of GDMS Tools and Uses

There are several new and expanded GDMS use cases across the three stages of complexity. These uses may allow the system to serve as a site exploration planning and management tool, asset tracking system, hazard monitoring platform, and sustainability dashboard. By adopting these functions, PennDOT could align with emerging best practices in digital delivery and asset management, ensuring that geotechnical data supports not only design and construction but also long-term resilience and compliance.

- **Laboratory information management:** GDMS can integrate directly boring software or lab software workflows, storing results from Atterberg limits, triaxial tests, or concrete strength tests. This may include standardized reporting, linking photos of specimens, and ensuring traceability of personnel and equipment.
- **Field productivity and consultants management:** Comparison of field productivity rates (e.g., ft per hour drilled, yardage placed per day) across PennDOT and external crews can be examined to identify performance variables and indicators. Allowing the comparison of consultants, crews, equipment, and different field conditions to improve overall operations and resolve site-specific or condition specific performance inhibitors.
- **Equipment inventory and calibration tracking:** The system can maintain records of drilling rigs, sensors, and lab equipment, including calibration schedules, maintenance logs, and repair histories. This ensures compliance with ASTM/AASHTO standards and reduces risk of data collected with faulty instruments.
- **Real-time monitoring integration:** GDMS can ingest continuous data streams from geotechnical and structural sensors (piezometers, inclinometers, strain gauges), global navigation satellite system (GNSS), and TBM instrumentation. Real-time dashboards allow engineers to visualize trends, set alarms, and respond to anomalies immediately.
- **Hazard early-warning systems:** By linking monitoring data with thresholds, GDMS can generate automated alerts for landslides, sinkholes, or embankment failures. Several currently available commercial platforms allow alarms can be sent via email/SMS to authorized users, enabling rapid responses and improved safety.

- **Cross-disciplinary data correlation:** GDMS can correlate geotechnical data with environmental, hydrological, and structural datasets, enabling holistic analysis of site conditions. This is particularly useful for large infrastructure projects like tunnels, dams, or other critical utility facilities.
- **Sustainability and resilience metrics:** By integrating soil data, construction records, and monitoring results, GDMS can support lifecycle (i.e. carbon) accounting, resiliency and durability scoring, and sustainability reporting.
- **Site condition validation:** The GDMS could serve as the authoritative record of subsurface investigations, construction records, and performance monitoring. This could include:
 - **Site variability and consistency:** Comparison of site conditions including sample recovery, rock quality designation (RQD), jointing, soil classifications, and soil and site factors as part of an expanded uncertainty tool and subsurface consistency profile. This information may be used to evaluate acceptability of collected field data prior to the completion of field work or substantial design completion, reducing late-cycle remobilization to the field, re-design, and improving design parameter selection.
 - **Cross-checking investigations:** Borehole logs, lab test results, and in-situ testing data stored in GDMS can be compared against design assumptions and construction records. This ensures that site conditions match what was anticipated in design and identify outliers or atypical site conditions.
 - **Automated consistency checks:** Validation routines can flag discrepancies between field data (e.g., soil classification, groundwater depth) and design inputs. This reduces the risk of mischaracterization.
 - **Historical benchmarking:** GDMS allows engineers to compare new site data against historical projects in similar geologic settings, validating whether observed conditions are typical or anomalous. This may also be used as part of a larger dataset to revise district and state-wide data such as bedrock maps, groundwater flow, and other data used for civil and non-civil projects.
- **Natural hazard screening:** The GDMS could integrate geotechnical data with hazard layers (landslides, flooding, seismic zones) to validate risk exposure.
 - **Spatial overlays:** By linking asset locations to GIS hazard maps, GDMS can automatically flag slopes, embankments, or walls in high-risk zones.
 - **Event correlation:** Performance data (failures, repairs, monitoring results) can be correlated with hazard events (rainfall, seismic activity) to validate vulnerability models.
 - **Predictive triggers:** GDMS can store instrumentation data (piezometers, inclinometers) and set thresholds that validate hazard exposure in real time, supporting early warning systems.
- **Other areas of concern:** The GDMS may also serve as a tool for several other areas of concern, validating broader geotechnical and asset management concerns.
 - **Environmental constraints:** Integration with environmental datasets (wetlands, karst, contaminated soils) to flag sites where geotechnical conditions intersect with environmental risks.

- **Construction quality:** Validate whether as-built conditions (compaction, material properties) meet specifications by linking QA/QC records to design parameters and pre-identify potential trouble areas prior to visual distresses are present.
- **Lifecycle performance:** Validation of treatment effectiveness by tracking repairs, maintenance, and monitoring data; feeding back into PennDOT's risk-based asset management framework.

As demonstrated by these uses the GDMS has the potential to evolve into a multifaceted platform that supports a wide range of uses including laboratory information management, equipment tracking, real-time monitoring, hazard early-warning, and construction quality assurance. It can also serve as the authoritative record for site condition validation, offering automated consistency checks, historical benchmarking, and hazard screening through GIS overlays and predictive triggers. Beyond technical functions, the GDMS can strengthen compliance with industry standards, improve safety through rapid alerts, and feed lifecycle performance data back into PennDOT's risk-based asset management framework; aligning with national best practices and ensuring long-term resilience and accountability.

7. CLOSING

The development of PennDOT's GDMS represents a significant step toward modernizing subsurface data practices and advancing an understanding of risk relating to transportation systems and geotechnical data. The pilot has validated the technical architecture, demonstrated the value of integrated 2D and 3D visualization, and confirmed the feasibility of DIGGS-compliant workflows within PennDOT's enterprise environment. Moving forward, success will depend on disciplined governance, sustained stakeholder engagement, and careful management of the transition from pilot to production. By following the staged expansion strategy which includes standardization, district pilots, statewide integration, and eventual digital twin optimization, PennDOT can institutionalize GDMS as a durable platform that enhances engineering rigor, supports operational efficiency, and strengthens long-term resilience. This report underscores that GDMS is not simply a technical tool, but a foundational system for geotechnical stewardship, positioning PennDOT at the forefront of digital delivery and infrastructure management.

REFERENCES

- Burns, S. E., Torres, J., Jung, Y., Liu, C., & Tsai, Y. (2023). *Towards the Implementation of a Geotechnical Asset Management Program in the State of Georgia*. Atlanta, Georgia: Georgia Department of Transportation. Retrieved from <https://rosap.nhl.bts.gov/view/dot/67653>
- Federal Highway Administration. (2013, April). FHWA-HRT-13-031. *Implementation of Geotechnical Asset Management at State DOTs*.
- Federal Highway Administration. (2025, April 08). *Asset Management*. Retrieved from U.S. Department of Transportation, Federal Highway Administration: <https://www.fhwa.dot.gov/asset/guidance.cfm>
- Gransberg, D. D., & Loulakis, M. C. (2012). *NCHRP Synthesis 429: Geotechnical Information Practices in Design-Build Projects*. National Cooperative Highway Research Program, Transportation Research Board. Washington, D.C.: National Academies Press.
- Jebelli, H., Asadi, S., Mutis, I., Lui, R., & Cheng, J. (2024). *Digital Twins in Construction and the Built Environment*. Reston, Virginia: American Society of Civil Engineers. doi:<https://dx.doi.org/10.1061/9780784485606>
- Mayne, P. W., Christopher, B. R., & DeJong, J. (2002). *Subsurface Investigations-- Geotechnical Site Characterization: Reference Manual*. National Highway Institute. Washington, D.C.: Federal Highway Administration. Retrieved from <https://rosap.nhl.bts.gov/view/dot/41566>
- National Academies of Sciences, Engineering, and Medicine. (2024). *Consensus Study Report: Foundational Research Gaps and Future Directions for Digital Twins*. Washington, D.C.: National Academies Press. doi:10.17226/26894. Retrieved from <https://dx.doi.org/10.1061/9780784485606>.
- New York Department of Transportation. (2025). *Digital Delivery - Interim Guidance Document*. New York Department of Transportation. Retrieved from <https://www.dot.ny.gov/divisions/engineering/design/dqab/hdm/hdm-repository/Digital%20Delivery%20Guidance.pdf>
- Ohio Department of Transportation. (2005). *Geotechnical Data Management System (GDMS): Case Study 1*. Ohio Department of Transportation. Retrieved from https://www.tsaadvet.com/CaseStudies/1_CaseStudies_StateAgency.html
- Olsen, M. J., Raugust, J. D., Corvallis, O., & Roe, G. V. (2013). *NCHRP Synthesis 446: Use of Advanced Geospatial Data, Tools, Technologies, and Information in Department of Transportation Projects*. National Cooperative Highway Research Program, Transportation Research Board. Washington, D.C.: National Academies Press.
- Shishlov, A. S., Costin, A. M., & Davidson, M. T. (2023). Integration of Building Information Modeling Interoperability into Nonlinear Finite Element Analysis of Bridge Substructures. *Transportation Research Record: Journal of the Transportation Research Board*, (pp. 9-21). Washington, D.C. doi:10.1177/03611981231160172
- Svensson, M., & Friberg, O. (2024). Digital Geotechnical Information Management in a Life Cycle Perspective. *Proceedings of the 18th European Conference on Soil Mechanics and Geotechnical Engineering (ECSMGE2024)* (pp. 871-874). Lisbon, Portugal: Taylor Francis.
- Transportation Research Board. (2025, November 3). *NCHRP Synthesis 20-05/Topic 56-23*. Retrieved from <https://apps.trb.org/cmsfeed/TRBNetProjectDisplay.asp?ProjectID=5880>
- Unkefer, D., Saad, T., Mavandadi, S., Machairas, N., Collins, B., Hiel, L., . . . Jang, D. (2023). The Present and Future of 3D and 4D Geotechnical Data in BIM and Digital Twins. *2023 GI*

- Web Conference*. American Society of Civil Engineers. Available from <https://www.asce.org/education-and-events/explore-education/on-demand-webinars/the-present-and-future-of-3d-and-4d-geotechnical-data-in-bim-and-digital-twins/>.
- Vessely, M., Robert, W., Richrath, S., Schaefer, V. R., Smadi, O., Loehr, E., & Boeckmann, A. (2019). *NCHRP Research Report 903: Geotechnical Asset Management for Transportation Agencies*. National Cooperative Highway Research Program, Transportation Research Board. Washington, D.C.: National Academies Press.