Proceedings of the 21st ICSMGE, Vienna, Austria, 14 – 19 June 2026. Pistrol, Adam & Schweiger (eds.)
Published by: ÖGG, Austrian Society for Geomechanics, Salzburg, Austria, ISBN 978-3-9503898-4-5
https://doi.org/10.53243/ICSMGE2026-2004

3D Geospatial Databases – Unlocking the Full Potential of Ground Investigation Data

Joost Gevaert , Jules Blom
Bedrock.engineer, joost.gevaert@bedrock.engineer

Mats Kahlström
Awer, Bergen, Norway

ABSTRACT: Geospatial databases are designed to efficiently store, manage and query objects that are represented physically in a geographic space, such as in the real world. Ground investigation (GI) data is inherently tied to a physical real-world location, e.g. a borehole, and as such, using geospatial databases to manage GI data could be considered a perfect match. This is, however, far from the truth today. Most of the GI data today is either locked inside a PDF document or stored in structured text-files formatted based on different schemas such as AGS, DIGGS, GEF etc. Some GI data are even locked inside proprietary software. To overcome this challenge, an open-source Python library bedrock-ge is developed that enables converting GI data from structured text files to geospatial databases. This conversion makes GI data accessible to standard GIS software for analysis, visualization and further data management. Connecting the GI data into a GIS ecosystem creates pathways for integration with broader AEC workflows through platforms such as Speckle, allowing GI data to be visualized and federated together with CAD and BIM models. In addition, Speckle allows for the GI data to be loaded directly into computational design software such as Rhino/Grasshopper, enabling a more data-driven approach to geotechnical design.

KEYWORDS: Ground investigation data, Geospatial databases, Data interoperability, Open-source software, Computational design

1. INTRODUCTION

Ground Investigation (GI) data forms the foundation for ground modelling, risk assessment, and geotechnical design. Although engineering workflows across the construction industry are becoming increasingly digital and data-driven, the use of GI data still presents significant challenges in terms of data management, interoperability, and visualization. Getting access to this data in the right software or visualization tool takes considerable custom scripting or manual labor (El Sibaii et al., 2022; Gevaert, 2024). This is because GI data is either unstructured, such as in PDF documents, CAD drawings or Excel / CSV files with non-standard header names or stored as text in specialized GI data formats such as AGS (Association of Geotechnical and Geoenvironmental Specialists), DIGGS (Data Interchange for Geotechnical and Geoenvironmental Specialists), GEF (Geotechnical Exchange Format). These formats are designed primarily for data exchange and archiving, not for integration into broader digital engineering workflows.

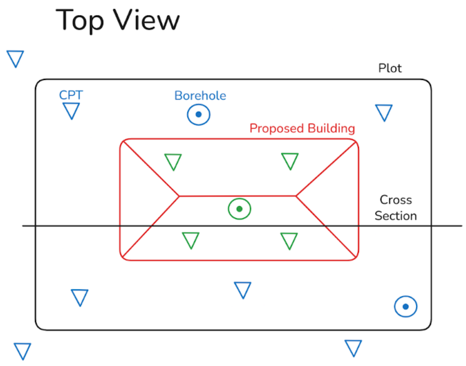
There have been efforts to align different GI data schemas, most notably during the "Geotech Interoperability Experiment" (Beaufils et al., 2024), which brought together people from AGS, DIGGS, the Open Geospatial Consortium (OGC) - responsible for open geospatial data standards, and buildingSMART International (bSI) - responsible for open BIM standards and the Industry Foundation Classes (IFC) format.

Although these data standardization efforts might be useful for the future, existing GI data exchange formats remain largely confined to specialized geotechnical software. They are rarely integrated directly with GIS, BIM, parametric design, geotechnical analysis, visualization tools, or the broader scientific computing ecosystem. This limits their usefulness in modern, multidisciplinary engineering workflows. As a result, GI data often remains siloed, making it difficult to integrate with the digital toolchains used by structural engineers, designers, planners, or data scientists. This fragmentation creates a “black box” effect: downstream engineers frequently lack access to the underlying GI data, design assumptions, and uncertainties, and must rely on static reports or simplified summaries. Consequently, collaboration, decision-making, and automated analysis across disciplines are constrained, and important details, such as subsurface variability, measurement uncertainty, and the rationale behind design choices, are often lost or misrepresented when GI analyses are handed off.

Gevaert & Blom (2025) propose to address these GI data management, interoperability, and visualization issues by structuring and storing GI data as relational 3D vector data in geospatial databases such as GeoPackage. To overcome the hurdle of converting GI data from text files to a geospatial database, the open-source Python library bedrock-ge is proposed. This paper presents the content from Gevaert & Blom (2025), showcasing the advantages of converting GI data to geospatial databases when it comes to GI data management, interoperability and visualization, and illustrates these benefits with selected real-world workflows to further clarify these benefits. Finally, the paper concludes with a discussion of future opportunities for integrating GI data into digital engineering ecosystems..

2. USING GEOSPATIAL DATABASES FOR GROUND INVESTIGATION DATA

Geospatial databases such as GeoPackage and PostGIS store geospatial vector data in tables that can be linked with relationships. Figure 1 shows a schematic plan view and cross section of GI data for a typical geotechnical engineering project, and helps to clarify that GI data is both relational as well as 3D geospatial data. This is further motivated in the next two subchapters.



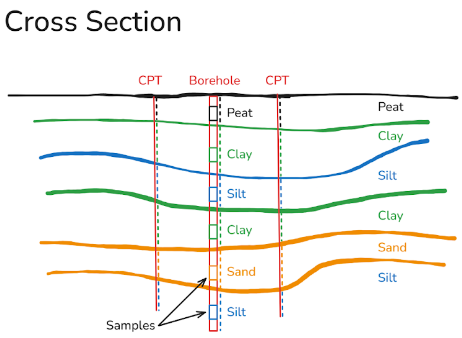
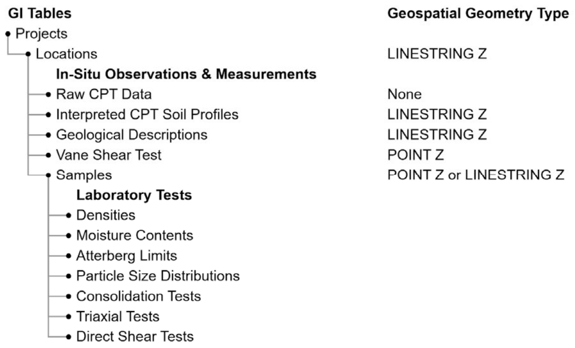


Figure 1. Top: Blue boreholes and CPTs represent historical GI data. The green borehole and CPTs represent new GI data obtained to inform the ground model and geotechnical design of the proposed building. Bottom: The schematic ground model shown in the cross section is the result of the interpretation of CPT data, in-situ observations & measurements and lab test data, and ground modelling.

2.1. Ground investigation data is relational

GI data inherently exists in a structured relational hierarchy. This hierarchy follows how ground investigations are performed. These relationships are typically one-to-many, meaning a GI project, for example, a new building or bridge, comprises multiple GI locations, e.g. borehole and CPT locations. At each GI location, multiple in-situ observations & measurements can be performed, e.g. raw CPT data, CPT interpretations, geological descriptions, Standard Penetration Tests (SPT). Additionally, each borehole can yield multiple samples from different depths. In turn, these samples may undergo multiple laboratory tests. These hierarchical relationships can be visualized in a tree diagram as shown in Figure 2.



GI Tables	**Geospatial Geometry Type**
Projects	
Locations	LINESTRING Z
In-Situ Observations & Measurements	
Raw CPT Data	None
Interpreted CPT Soil Profiles	LINESTRING Z
Geological Descriptions	LINESTRING Z
Vane Shear Test	POINT Z
Samples	POINT Z or LINESTRING Z
Laboratory Tests	
Densities	
Moisture Contents	
Atterberg Limits	
Particle Size Distributions	
Consolidation Tests	
Triaxial Tests	
Direct Shear Tests	

Figure 2. Hierarchical structure of GI data, together with the geospatial feature type corresponding to the different GI data types.

The hierarchical structure of GI data maps naturally onto a relational database, where projects, locations, samples, and laboratory tests can be stored in separate tables but remain consistently linked in the same way they relate in practice. Each entry in a table has a unique identifier, which makes it possible to connect locations to their projects, samples to their locations, and laboratory tests to their samples.

In contrast, common GI file formats differ in how they represent relationships. AGS files may contain incomplete or inconsistent links between projects, locations, and samples, while formats like GEF have no inherent relational structure. Storing GI data in a relational database enforces these hierarchies using a schema. This makes it easy to trace any sample, measurement, or test back to its corresponding location or project. Such databases are managed using Relational Database Management Systems (RDBMS), such as SQLite or PostgreSQL, and are accessed using Structured Query Language (SQL).

2.2. Ground investigation data is 3D geospatial data

Geospatial data is data that is associated with a location relative to Earth. There are two main categories of geospatial data: raster data and vector data. Raster data consists of a grid of pixels, representing continuous surfaces, of which satellite imagery and Digital Elevation Models (DEM) are good examples. Vector data represents discrete (3D) objects on Earth – according to the OGC Simple Feature Access (SFA) specification – with POINT, LINESTRING or POLYGON geometry. Properties can be attached to these geometries to describe them. Such an object represented by geospatial geometry + properties is often referred to as a feature ("Data Model (GIS)", 2025).

Unlike CAD or BIM geometry, geospatial features are georeferenced, meaning that it is aware of where on Earth its geometry is located ("Geographic Data and Information", 2025). This is achieved through a Spatial Reference System (SRS). A SRS is a standardized definition of the coordinate system + datum that links coordinates to locations on Earth. The most important types of SRSes are: geographic SRSes with coordinates in (Longitude [°], Latitude [°], Height [m]) and projected SRSes with coordinates in (Easting [m], Northing [m], Height [m]). Relational databases with a geospatial extension are SRS-aware and allow for adding a geospatial geometry column to its tables. In these tables, every row is a geospatial feature.

Because GI data is acquired at discrete locations somewhere on Earth, it can naturally be represented as geospatial vector data. GI locations are commonly represented as 2D POINT features, for example enabled by the open-source Python package python-AGS4 (Senanayake et al., 2022). However, GI data is inherently 3D with multiple observations, measurements and samples taken at several depths at every GI location. This allows 3D geospatial representations following the OGC Simple Feature Access spec. GI locations can be represented as 3D LINESTRING Z features. In-Situ Observations & Measurements can be represented as either 3D POINT Z features - e.g. vane shear tests and Standard Penetration Tests (SPT) - or 3D LINESTRING Z features - e.g. geological descriptions of borehole intervals or the layers of a soil profile resulting from CPT data interpretation. Samples can be represented as 3D POINT Z or LINESTRING Z features depending on the type of sample taken.

Figure 2 gives an overview of how these geospatial features are related to the hierarchical structure of GI data. Note that no geospatial geometry is assigned to raw CPT data. CPT data is recorded at a high vertical resolution. Assigning a POINT Z to every measurement would create an excessive number of features, which adds little practical value for visualization or analysis, while substantially increasing storage requirements and computational load.

3. ADVANTAGES OF CONVERTING GI DATA TO RELATIONAL 3D GEOSPATIAL VECTOR DATA

3.1. Workflow reliability and automation

By enforcing hierarchical relationships and consistent schemas, relational geospatial databases make it possible to automate data processing, analysis, and reporting. This reduces manual work, ensures reproducibility, and enables integration with downstream tools and models. It's the foundation that makes the other benefits (visualization, interoperability, dashboarding) possible.

3.2. Data sovereignty

Converting GI data to open geospatial database standards ensures complete data ownership and eliminates vendor lock-in that commonly occurs with proprietary GI data management software. Geospatial databases such as PostGIS can also be deployed on any hosting platform, allowing organizations to choose the cloud provider and data residency of their liking. This enables organizations to maintain full control over their valuable subsurface data without being dependent on specific software licensing agreements or facing risks of data inaccessibility due to software discontinuation.

3.3. *Scalability and performance*

Geospatial databases are built to store and manage large amounts of spatial data efficiently. They use indexing methods like R-tree and quad-tree to speed up spatial queries, even for datasets with millions of points. This is useful for large geotechnical projects or regional databases with data from many sources. Unlike file-based systems that load full datasets into memory, geospatial databases let users retrieve only the data they need—by location, depth, or test type. Related data, such as samples, tests, and locations, can be easily linked, making it straightforward to extract and analyze complex combinations of information. These systems also allow multiple users to access data simultaneously, support backups and recovery, and can connect with other enterprise systems.

3.4. *Geospatial databases*

Geospatial databases provide native support for spatial reference systems and coordinate transformations, automatically handling the complex geometric calculations required when working with georeferenced CAD and BIM models or (3D) web-based viewers. This spatial awareness eliminates common errors that occur when manually managing coordinate transformations.

Furthermore, all major GIS software platforms – including QGIS, ArcGIS, and web-based mapping applications – provide direct connectivity to geospatial databases. This allows the excellent geospatial and geostatistical analysis tools in these GIS software platforms to be applicable to GI data, simplifying ground modelling workflows, and enabling integration of GI data with broader GIS-based workflows such as urban planning, infrastructure planning or hazard mapping.

3.5. *Integration with scientific computing ecosystems*

The open-source scientific computing ecosystem, with Python as the dominant language and led by libraries such as NumPy, SciPy, GeoPandas, and scikit-learn, is experiencing unprecedented growth, and innovation in computational capabilities and machine learning applications. By structuring GI data as geospatial vector data, geotechnical engineers can leverage this rapidly advancing ecosystem for advanced statistical analysis, machine learning-based soil classification, and predictive modeling without requiring specialized GI software.

Additionally, code-based workflows ensure reproducibility by capturing every step rather than relying on point-and-click interfaces that leave no permanent record. Code inherently require explicit logic that can be independently verified and peer-reviewed, eliminating the ambiguity that can result from GUI-based software. Version control systems like git provide detailed histories of how modeling code has evolved.

3.6. *Easier visualization and spatial understanding*

Geospatial databases are compatible with desktop GIS applications and web-based mapping software libraries, like Deck.gl and CesiumJS, enabling 3D visualization of GI data in context with other geospatial information. These visualizations help stakeholders understand what types of GI data exist and where they are located, improving communication and decision-making, see for example Figure 3.

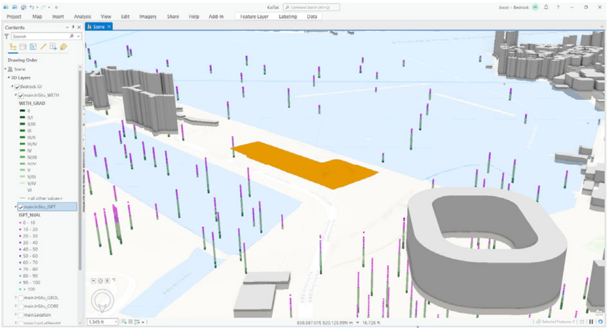


Figure 3. Boreholes in Hong Kong where granite weathering grades are displayed in different tones of green, and standard penetration tests performed on the residual soils are represented as pink dots in ArcGIS Pro. Using ArcGIS Online this visualization can also be made available as a web-app.

Apart from visualizing, GI data can be published to web platforms that integrate with BIM and other AEC models, such as Speckle. Speckle allows geospatial data to be shared, explored, and combined with models from CAD and BIM software, including Civil3D, Revit, and Rhinoceros 3D & Grasshopper. Once GI data is stored in a geospatial database and published to Speckle, users can interact with it in a web viewer, visualizing multiple data layers simultaneously and exploring subsurface information in context with architectural and engineering models (see Figure 4). Using this workflow, both GI data and interpolated ground models can be loaded directly into software like Rhinoceros 3D or Revit. This allows engineers to integrate subsurface information with architectural and structural models, supporting more informed design decisions and reducing manual data translation.

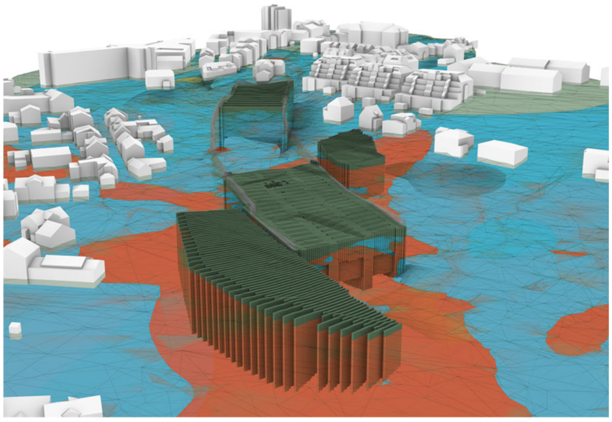


Figure 4. Illustration of ground model surfaces (red and blue) visualized jointly with building data, a Civil3D cut-and-cover tunnel and a proposed design for ground improvement (lime-cement columns) from Rhinoceros 3D.

3.6. *Interoperability with CAD and BIM software*

From Speckle, GI data and interpolated ground model surfaces can be loaded into CAD and BIM software (see Figure 5). This integration supports more flexible geotechnical workflows and improves interdisciplinary collaboration by connecting subsurface data with architectural and engineering models.

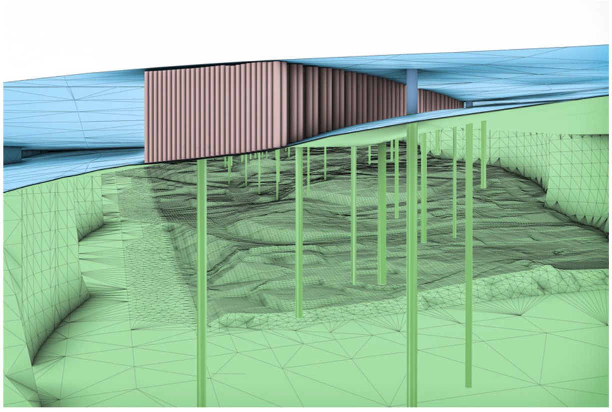


Figure 5. GI data that was published to Speckle from QGIS can then be loaded into CAD and BIM software: Rhinoceros 3D

3.7. *Interoperability with parametric design software*

With the use of bedrock-ge the parametric piled foundation design workflow presented by Gevaert (2024) was significantly simplified, because the geospatial GI data can be loaded from Speckle directly into Grasshopper for Rhinoceros 3D. In Grasshopper, the rockhead topography was interpolated using the interpolation algorithm of choice - Radial Basis Functions (RBF) in this case - after which the piles could be extended to the rockhead and checks were performed automatically.

In addition to simplifying geotechnical design workflows, integrating GI data into parametric design software makes it possible to parametrically generate the models required for geotechnical analysis.

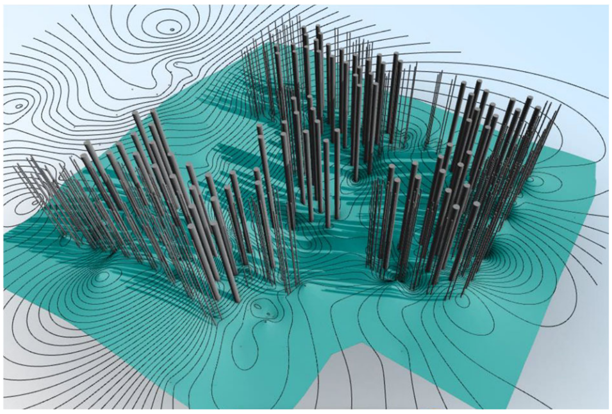
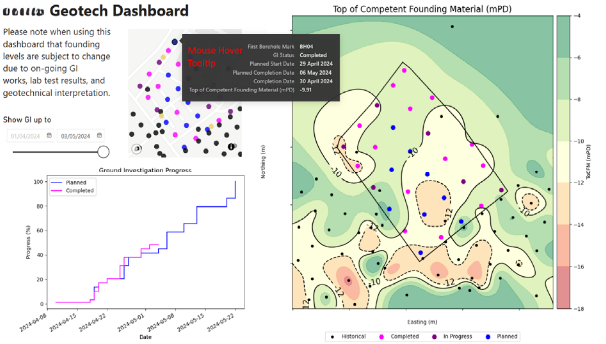


Figure 6. A piled foundation designed parametrically. Converting GI data to geospatial data significantly reduced the complexity of this parametric design workflow compared to the workflow presented by Gevaert, (2024).

3.8. *Interoperability with dashboarding software*

When GI data is available in standardized geospatial formats, it becomes possible to create dashboards with GI data. For example, to keep stakeholders up to date on ground investigation progress and the influence of new GI data on the feasibility of the envisaged foundation typology and design, see Figure 9.



Geotech Dashboard

Please note when using this dashboard that founding levels are subject to change due to on-going GI works, lab test results, and geotechnical interpretation.

Show GI up to 01/04/2024 03/05/2024

Mouse Hover Tooltip

First Borehole Mark: BH04
GI Status: Completed
Planned Start Date: 29 April 2024
Planned Completion Date: 06 May 2024
Completion Date: 30 April 2024
Top of Competent Founding Material (mPD): -9.91

Figure 7. A PowerBI dashboard that was used to keep stakeholders up to date on Ground Investigation (GI) progress and the influence of new GI data on the envisaged foundation typology and design (Gevaert, 2024, presentation slides).

6. CONCLUSION

The geotechnical engineering industry struggles with data management and interoperability, which hinders its progress toward more automated, data-driven workflows. Although ground investigation (GI) data is fundamentally geospatial, it's often trapped in proprietary software, PDFs, or various structured text files (e.g., AGS, DIGGS, GEF), limiting its accessibility.

To overcome this, we developed bedrock-ge, an open-source Python library that converts GI data into geospatial databases. This conversion makes the data compatible with standard GIS software and allows it to be federated with BIM and CAD models through platforms like Speckle. This integration enables enhanced visualization and direct use of GI data in computational design software, promoting a more data-driven approach to geotechnical design. Our approach aims to eliminate vendor lock-in, improve data interoperability, and connect GI data with the broader scientific computing ecosystem.

7. REFERENCES

Beaufils, M., Schleidt, K., Van der Schaaf, H., Ponti, D., Chadwick, N., & Dasenbrock, D. (2024, July 5). OGC Geotech Interoperability Experiment Engineering Report. https://docs.ogc.org/per/24-008.html Data model (GIS). (2025). In Wikipedia. https://en.wikipedia.org/w/index.php?title=Data_model_(GIS)&oldid=1287820692

El Sibaii, M., Granja, J., Bidarra, L., & Azenha, M. (2022). Towards efficient BIM use of geotechnical data from geotechnical investigations. Journal of Information Technology in Construction, 27, 393–415. https://doi.org/10.36680/j.itcon.2022.019

Geographic data and information. (2025). In Wikipedia. Available at: https://en.wikipedia.org/w/index.php?title=Geographic_data_and_information&oldid=1291003601

Gevaert, J. (2024). Bridging the Gap Between Geosciences and Civil Engineering. 2024(1), 1–5. https://doi.org/10.3997/2214-4609.2024101650

Gevaert, J. & Blom, J. (2025) Unearthing the hidden potential in ground investigation data by converting to geospatial databases. Geotechniekdag 2025

Senanayake, A. I., Chandler, R. J., Daly, T., & Lewis, E. (2022). python-ags4: A Python library to read, write, and validate AGS4 geodata files. Journal of Open Source Software, 7(79), 4569. https://doi.org/10.21105/joss.04569