Soils and Rocks
www.soilsandrocks.com
An International Journal of Geotechnical and Geoenvironmental Engineering
ISSN 1980-9743
ISSN-e 2675-5475

Development of a logical model for geotechnical databases

Jorge Brock Júnior[1#], Elisangela do Prado Oliveira[1], Rogerio Rodrigues de Vargas[1]

Article

Keywords
Geotechnical engineering
Data standardization
Logical model
Database

Abstract

The advent of construction informatization technologies and the success of the BIM (Building Information Modelling) methodology in the infrastructure market have increased the need for collaboration between different areas of knowledge, professionals, and agents involved in an infrastructure or civil project. For this collaboration to be effective, the correct standardization of procedures is essential to ensure efficient communication between the parties involved in these complex processes. The field of geotechnics and geotechnical investigations has the AGS (Association of Geotechnical and Geoenvironmental Specialists) standard for electronic transmission of files produced during drilling and laboratory tests. However, there is a gap regarding the standardization of a digital structure that can be used for the implementation of geotechnical databases. This paper proposes a logical model for the demands of the geotechnical market, which can be used by the entire community. The proposed model allows for easy access to large-scale data and the possibility of performing mathematical operations more easily.

1. Introduction

1.1 History of geotechnical databases

In the field of geotechnical engineering, data management is crucial for the success of various projects involving the construction and protection of urban areas, rural works, and natural resource exploration. Geotechnical engineering heavily relies on geological and geotechnical knowledge, necessitating efficient data storage, processing, and organization. This work aims to incorporate these elements into a visual model aggregating important geotechnical information into a single database.

The history of geotechnical databases traces back to recognizing their value for urban development and engineering projects. Pioneering efforts, such as those of Legget (1973) in the book "Cities and Geology," highlighted the importance of creating and maintaining public geotechnical databases. Legget (1973) argued that the availability of geotechnical data could lead to more efficient and less wasteful urban planning and construction. His work emphasized the benefits of such databases for geotechnical risk assessment, urban planning, geological mapping, and public education. These early concepts laid the foundation for subsequent developments in geotechnical data management and accessibility, demonstrating the significant role of comprehensive geotechnical records in advancing the field. Figure 1 shows a timeline of the biggest events in the geotechnical database history.

In the 1950s and 1960s, geotechnical data was primarily collected manually. Engineers and researchers would record information on physical spreadsheets and paper, which lacked standardization and posed challenges in both handling and organizing the data. The storage was rudimentary, and managing this information was labor-intensive. However, the rise of early computers marked the beginning of the digital storage of data, though the process was still very basic compared to modern systems.

The 1970s saw a significant leap with the adoption of relational databases, a model proposed by Edgar Codd. This model allowed geotechnical data to be organized into tables with defined relationships, paving the way for efficient data management. Additionally, the first geotechnical analysis systems, like GEOSYS, emerged, combining database technology with computational tools for soil and rock analysis. The use of structured storage, facilitated by relational databases, enabled more effective organization and retrieval of field and laboratory data through Structured Query Language (SQL) (Codd, 1970).

[#]Corresponding author. E-mail address: jorgebrockjunior@gmail.com
[1]Universidade Federal do Paraná, Centro de Estudos do Mar, Pontal do Paraná, PR, Brasil.

Submitted on April 12, 2024; Final Acceptance on October 28, 2024; Discussion open until May 31, 2025.
Editor: Renato P. Cunha
https://doi.org/10.28927/SR.2025.003024

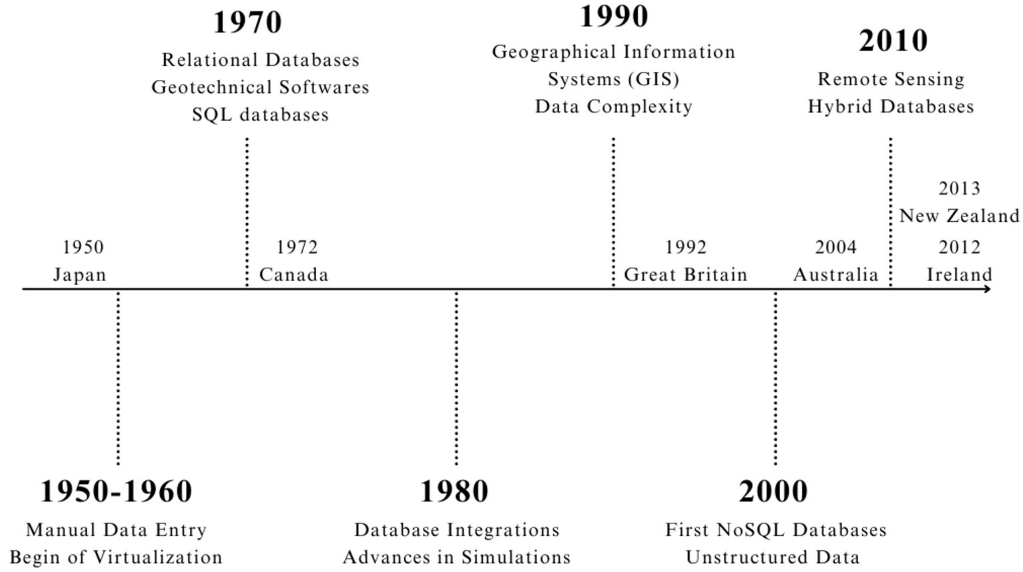

Figure 1. Timeline of major geotechnical database events.

In the 1970s, the Geological Survey of Canada (GSC) initiated a national urban geology program, resulting in the collection and collation of geotechnical data from 27 cities. This initiative led to the creation of the Urban Geotechnical Automated Information System (UGAIS), aimed at compiling, evaluating, and presenting geological information to meet the needs of planners, administrators, and engineers. Despite challenges such as the need for mainframe computers and the eventual obsolescence of some databases, the initiative underscored the enduring value of geotechnical data. In regions like Ontario, the borehole database remains a vital resource, accessed thousands of times annually, illustrating the long-term benefits of maintaining and updating geotechnical databases for urban development and engineering practices.

In the 1980s, computational tools for geotechnical analysis became more advanced, and their integration with databases improved. This decade saw an increase in the use of software to analyze geotechnical data, enhancing civil engineering projects. These tools were applied in the design and construction of major infrastructure projects like highways and dams, where the ability to analyze and manage large datasets became crucial for project success.

By the 1990s, geotechnical databases began to integrate with Geographic Information Systems (GIS), a development that allowed geotechnical data to be combined with geospatial data. This practice became particularly useful in environmental studies and mapping. As the decade progressed, the complexity of geotechnical data grew. With the advent of three-dimensional and temporal data, the volume of data being collected expanded rapidly, challenging the traditional database systems in use at the time.

The early 2000s marked the beginning of a shift towards more flexible data systems. The increasing prevalence of unstructured data, such as that collected from sensors or satellites, highlighted the limitations of traditional relational databases. To address this, NoSQL databases like Bigtable and MongoDB were introduced, offering solutions to handle large volumes of unstructured data and enabling horizontal scalability (Chang et al., 2008). This shift was crucial for the geotechnical field as data collection became more complex and diversified.

From the 2010s onwards, hybrid databases, combining both relational and non-relational systems, became more popular in geotechnical engineering. These databases offered the integrity and consistency required for structured data while providing the flexibility needed for handling large volumes of unstructured data. This period also saw the application of advanced geotechnical monitoring techniques in large-scale projects. Modern systems, including the Internet of Things (IoT) and InSAR, were used to map and predict landslides and ground displacements, enhancing environmental monitoring and large-scale infrastructure projects (Guilhot et al., 2021).

As of today, the current biggest geotechnical databases are related to Japan, Great Britain, Australia, New Zealand, and Ireland, the most recent being from New Zealand, with its creation in 2016, this database is based on a interface with the Web including data from boreholes, Cone Penetration Tests (CPTs) and trial pits. This data is stored via a private provider with undisclosed data management techniques, in an analysis of a random data point the system took 15 seconds to display the information of the point. The database is compatible with the AGS format and has also the inclusion of non-structured data in its points.

Today, some of the largest and most comprehensive geotechnical databases are maintained by countries such as Japan, Great Britain, Australia, New Zealand, and Ireland. Among these, New Zealand's database, established in 2016, stands out as one of the most recent and innovative. This database is accessible via a web-based interface and includes a wealth of geotechnical information such as data from boreholes, Cone Penetration Tests (CPTs), and trial pits. The data is securely managed by a private provider, though the specific data management techniques employed remain undisclosed. Despite these unknowns, the database's performance remains notable; for example, retrieving information from a randomly selected data point took approximately 15 seconds, demonstrating reasonable efficiency. Furthermore, the database is compatible with the widely used AGS (Association of Geotechnical and Geoenvironmental Specialists) data format, ensuring its utility across various platforms and systems. It also supports the inclusion of non-structured data within its datasets, expanding its flexibility and adaptability for diverse geotechnical applications.

In addition to New Zealand's system, numerous other significant geotechnical databases exist around the world, playing vital roles in regional planning and development. These databases serve as indispensable tools for urban planners, engineers, and geologists, helping them navigate the complex challenges of construction, risk management, and resource exploration in various terrains and environments. While the specifics of each database vary—ranging in size, scope, and technological sophistication—their collective importance cannot be overstated. These systems offer a structured way to store, access, and analyze vast amounts of geotechnical data, providing essential insights that inform safer and more efficient engineering decisions. Each database has its own unique history and evolution, shaped by the specific needs and technological advancements of the region it serves, yet all share the common goal of advancing geotechnical engineering practices through enhanced data accessibility and management. Table 1 provides an overview of some of the most notable geotechnical databases around the globe.

Kokkala & Marinos (2022) highlight the urgent need for organized databases to enhance information for decision-making processes in important geotechnical projects. The primary focus is on managing geological and geotechnical knowledge within a well-structured database capable of promoting correlations and offering valuable insights during the design and construction phases of critical projects. This perspective is exemplified in the city of Thessaloniki, northern Greece, where a vast repository of geological and geotechnical data from over 600 drillings is meticulously stored in a georeferenced database. The goal is to facilitate the management and processing of geoengineering data, a substantial asset for various geotechnical endeavors.

Table 1. Geotechnical databases and their respective technology.

Region	Year	Number of boreholes	Technology
Ontario, Canada	1972	90,000+	CSV
Great Britain	1992	100,000+	AGS files (SQL)
Perth, Australia	2004	650+	Mapbox (SQL)
Japan	1950	200,000	Excel Sheets (SQL)
New Zealand	2013	20,000+	ArcGIS Map (SQL)
Ireland	2012	27,800	ArcGIS Map (SQL)

The geotechnical database presented by Kokkala & Marinos (2022) heralds a shift from conventional approaches by incorporating multi-source datasets. Its primary function is to provide stakeholders with a tool for data analysis, correlation, and visual representation on maps, tables, and diagrams. This article suggests the application of geographic information system techniques in conjunction with geostatistical methods. These techniques help establish geotechnical zoning maps that highlight areas susceptible to liquefaction, settlement and assess the suitability of the site for different infrastructure projects. The article aims not only to provide quantitative ranges for influencing parameters but also advocates for the creation of accurate and reliable geotechnical models that comprehensively elucidate surface and subsurface conditions in the research area.

Li et al. (2022b) contribute to the discussion, highlighting the importance of soil characteristics, a critical factor for foundation engineering. According to this author, the availability of free, modern, and easily accessible data on soil characteristics is very promising for streamlining preliminary assessments for foundation engineering purposes. Existing databases produced by institutions such as the Czech Geological Survey offer valuable resources to construction companies. Further improvements in these databases can simplify the assessment of foundation soil characteristics. The potential to interconnect this information with data on well drilling, slope instabilities, and area geology opens up the possibility of generating comprehensive preliminary assessment reports, enabling rapid and cost-effective guidance in specific locations.

Phoon et al. (2022), studying the field of geotechnics with a focus on the role of data for the industry, highlights the current trend of needing more insights and actions to translate the state of found data, using descriptive statistics, machine learning algorithms as studied by Kim & Ji (2022) and Li et al. (2022a), and data-centric decisions. Thus, Phoon et al. (2022) adds to the discussion by emphasizing the importance of geotechnical data in engineering works, revealing a trend in the increasing amount of information required for a well-conducted study.

In the context of geotechnical engineering, where data plays a fundamental role in decision-making, this article investigates the urgent need for such databases, examines existing databases, and introduces a novel approach to geotechnical database modeling. It provides a roadmap for leveraging geological data, mitigating challenges, and finding opportunities presented in the geotechnical field, ultimately contributing to safer and more efficient urban development and infrastructure projects.

2. Materials and methods

The type of database considered in the design of the schemas was non-relational (NoSQL), providing greater adaptability to new unstructured information in the system. An environment that stores such data may serve as a future basis for extracting structured geotechnical information by extracting information from percussion drilling through field bulletin photos, for example. Furthermore, the structure can be easily modified to include other attributes not initially modeled by the authors, allowing for continuous evolution of the database.

To visualize and understand the relationships between database attributes and objects, the U-type schematic metamodel proposed by Candel et al. (2022) was used. This metamodel, specifically developed for non-relational databases (NoSQL), provides an accessible visual representation of the complex relationships present in geotechnical data.

The structured data present in reports and surveys include properties represented by numbers or categorical texts, such as soil color and type. These data allow for mathematical or logical operations, such as calculating averages, including in mathematical equations, and creating graphical visualizations through the work of engineers and geologists. The flexibility of these operations allows database users to select and filter values according to their specific needs.

In addition to structured data, this research observed and made use of unstructured data from surveys and geotechnical studies. These data include image and document format reports, photos of any nature, or extensive texts containing descriptions and observations. This diversity of unstructured data serves as the basis for the extraction of structured information, enabling an integrated and comprehensive approach to managing and analyzing geotechnical data. This is as advanced artificial intelligence algorithms and parameter extraction are developed.

2.1 AGS transfer file format

In this study, the AGS-4 data schema and digital transfer standards, developed by the Association of Geotechnical and Geoenvironmental Specialists (AGS), were adopted. The selection of attributes focused on the most commonly observed elements in surveys and geotechnical studies, aiming for a comprehensive and accurate representation of this information.

The .ags archive intends to be a text file format used to transfer data reliably between organizations in the site investigation industry, independent of software, hardware or operating system. The format was set in 1991 and provides a standard way to transfer ground investigation, laboratory testing and monitoring data between the contributing parties of a project that involves geotechnical or geoenvironmental elements. This data is generated by contractors, professionals and companies that can save it for later use.

The AGS standard is one of the most used in geotechnical projects, as it has the advantage of delineating procedures and standards for digital file transfer. Used in projects and research requiring a vast amount of data, the procedure only solves the problem of standardizing files. Associations such as AGS Brasil add value by providing procedures for conducting surveys and geotechnical studies that follow the AGS file standard, complementing the scenario.

The AGS data schema stands as a data structure aimed at enriching digital files with additional information relevant to a geotechnical project. Figure 2 shows the data schema followed by the digital transfer standard.

According to AGS-BR (2022), the data format file must follow 20 rules to ensure compatibility between .ags files.

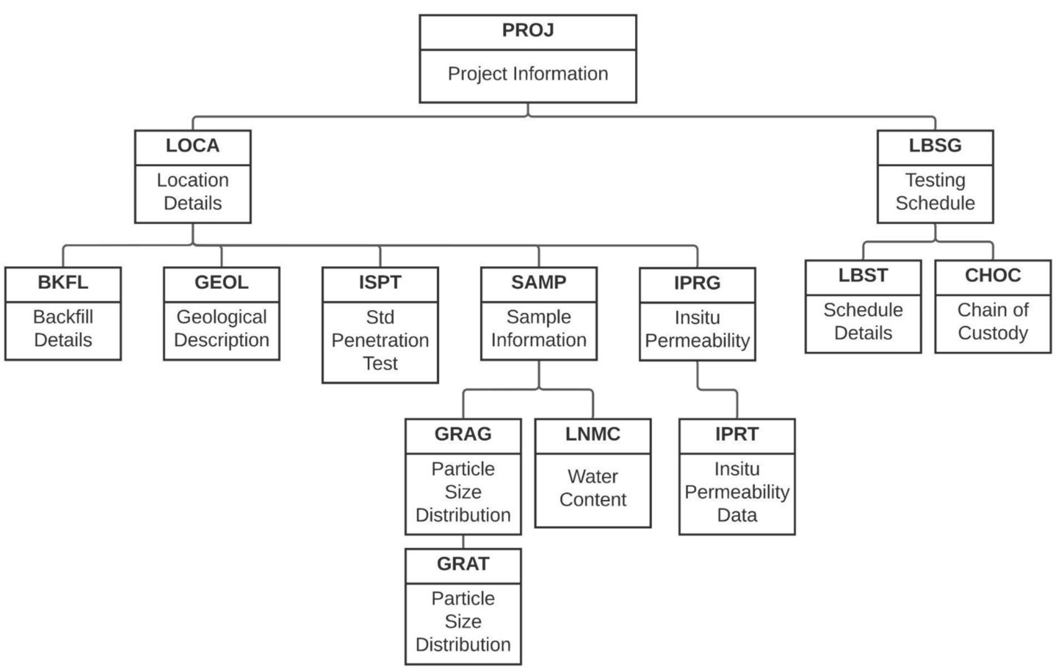

Figure 2. Data group schema followed by AGS-4. Adapted from Faversham, Kent (AGS-BR, 2022).

These rules help ensure a format readable by any personal computer provided with an operational system.

The first rule states the encoding system used in the file. Rules 2 to 4 impose some structures on the fields and variables to be saved. Rules 5 to 9 declare naming restrictions for better use readability both by users and electronic systems. The rules 10 to 16 give minimum requirements for variables and groups. For example, there must be at least one PROJ group in each data point. This ensures that there are no blank data in the database. The rules 17 to 20 detail requirements for group names, naming existence and requiring structure to naming.

Each object in the figure represents a relevant geotechnical element, such as project and location data. These pieces of information, in turn, have their own data schema containing various more detailed information about the geological procedures performed. Thus, Figure 3 shows an extract of one of these objects represented in the schema of Figure 1. Therefore, AGS 4 standard aids in aggregating information for digital files transferred between parts of a project.

To date, no academic work has explored the possibility of incorporating this data into non-relational databases.

Virtual systems that store complex information such as social networks and management systems rely on NoSQL databases, more flexible systems, which have horizontal scaling availability, that is, adding more schemas and objects to the data schema, and higher performance for mass data read and write operations. The work of Atzeni et al. (2020) addresses the use of this database category.

3. Analysis and results

In recent years, the choice between relational and non-relational databases has become a recurring topic in system architecture, mainly due to the exponential growth in data volume and the diversity of its structures. Relational databases, based on the model proposed by Codd (1970), offer a robust approach to ensuring data integrity and consistency, making them suitable for transactional environments. However, the growing need for scalability and flexibility in modern applications has led to the rise of non-relational databases (NoSQL), which enable distributed and unstructured data storage, as highlighted by Strauch (2011). Table 2 presents a comparison between these two types of databases, highlighting their main characteristics:

Many works already conducted on the subject have used structured databases, which are tools based on table relationships. Relational databases use tables with interrelations, allowing the use of SQL queries to manipulate and retrieve data in a structured manner. This model is highly efficient for transactional operations and when there is a need to maintain referential integrity and ensure consistency. However, these databases can struggle with scalability and performance under high-demand operations and large data volumes. In contrast, non-relational databases are designed to store data more flexibly, often in formats like JSON documents or wide columns. This flexibility allows for superior performance in large-scale read and write operations and is recommended for unstructured data or data that requires horizontal scalability. However, this flexibility may come at the cost of reduced consistency and integrity, and they may lack the robust transactional support found in relational databases.

Consider a case study involving the construction of a residential building in an urban area. To ensure the safety and stability of the structure, a comprehensive geotechnical investigation is crucial. The process begins with conducting Standard Penetration Test (SPT) assays to assess soil resistance. In parallel, laboratory procedures are carried out to determine Atterberg limits, providing information about the plasticity and liquidity characteristics of the soil. Additionally, granulometric analysis is performed to understand the distribution of particle

Group Name: PROJ - Project Information					
Status	**Heading**	**Suggested Unit / Type**		**Description**	**Example**
*R	**PROJ_ID**		ID	Project identifier	121415
	PROJ_NAME		X	Project title	ACME Gas Works Redevelopment
	PROJ_LOC		X	Location of site	High Street, Anytown
	PROJ_CLNT		X	Client name	ACME Enterprises
	PROJ_CONT		X	Contractors name	ACME Drilling Ltd
	PROJ_ENG		X	Project Engineer	ACME Consulting
	PROJ_MEMO		X	General project comments	
	FILE_FSET		X	Associated file reference (e.g. project specification, site location drawings)	FS1

Figure 3. Data object schema followed by AGS-4 (Faversham, Kent (AGS-BR, 2022).

Table 2. Comparative between SQL and NoSQL.

Characteristic	Relational Databases (SQL)	Non-Relational Databases (NoSQL)
Data Structure	Data organized into tables with rows and columns, based on the relational model proposed by Codd (1970).	Store data in flexible formats like documents (JSON), key-value, wide-columns, or graphs (Strauch, 2011).
Data Model	Fixed and strictly structured schema, requiring prior definition.	Flexible schema or schema-less, allowing modifications in data format.
Query Language	Structured Query Language (SQL), the standard language for data manipulation.	Use various APIs and specific query languages, such as JSON-like queries in MongoDB (Strauch, 2011).
Scalability	Vertical scalability (increasing resources on a single server).	Horizontal scalability (adding more servers), effective for large distributed datasets (Stonebraker & Hellerstein, 2005).
Data Consistency	Strong consistency and integrity through ACID transactions (Silberschatz et al., 2010).	Eventual consistency in distributed systems, focusing on availability and scalability (Strauch, 2011).
Integrity and Transactions	Robust support for ACID transactions, ensuring integrity in complex operations (Silberschatz et al., 2010).	Limited or no support for ACID transactions, depending on the NoSQL database type.
Ideal Data Type	Structured and organized data, suitable for transactional environments.	Ideal for semi-structured or unstructured data (logs, social media, big data).
Performance	Efficient performance in complex queries over structured data.	Optimized for fast read and write operations on large datasets distributed across nodes.
Examples	MySQL, PostgreSQL, Oracle, SQL Server (Silberschatz et al., 2010).	MongoDB, Cassandra, Redis, DynamoDB (Strauch, 2011).
Use Cases	Traditional enterprise applications (ERP, CRM).	Modern, distributed applications (social media, recommendation systems).

Source: The authors.

sizes in the soil. The granulometric curve is generated from all this data, and proper sizing of foundations and support structures is performed. Photographic records, hand-filled spreadsheets, and files resulting from the analysis of this information such as reports and geological maps are also produced during the geotechnical geological studies of the project in question. All this data is then stored in the database following the logical model proposed in this article.

To exemplify, some of the mentioned data formats will be presented. The individual drilling profile is usually presented in geotechnical study reports as a graphical component that assists in understanding the properties of the analyzed soil layers. The document is presented as an image or within PDF files and is characterized by being unstructured, making it difficult to extract values and information that can be processed by computers.

In Table 3, the presence of a table describing data extracted from a percussion test can be observed, this is a common document fulfilled by the probers. This data has the possibility of being stored in a database to allow mathematical operations. Because they are information with lower storage demand for computers, they can be stored in large quantities without incurring high costs. However, if transferred to the database through images or within text files, they become unstructured data.

Finally, the AGS data schema itself can be stored for direct use in geotechnical processing and modeling software, or it can be updated with information extracted from other files in the database, through Optical Character Recognition (OCR) algorithms, manual manipulation of information, or derivation of properties through mathematical or machine learning models. For example, finding physical index values through equations or algorithms using the data present in the system. Figure 4 shows an excerpt from the code that generates the file, with entities being delimited by "**" and subsequently attributes being delimited by "*", where the value assigned to the attributes is one line below. However, the authors suggest using a JSON (JavaScript Object Notation) file instead of a proprietary format with delimitation fields, as it offers various advantages in terms of readability, interoperability, and simplicity in data manipulation. Furthermore, its structure is language-independent, allowing easy integration into different development environments. Unlike delimited formats, which depend on specific characters to separate fields, JSON uses a hierarchical notation, making it easier to represent complex data. An excerpt of a translated code from AGS to JSON is presented in Figure 5.

After a rigorous examination of the AGS methodology, the recommendations present in the AGS Brasil guideline, and the structure of variables present in the reports of the organization, these formats and procedures refer to the distribution and hierarchical structure of data within the archive, this hierarchy is maintained to keep compatibility with the format.

Table 3. Percussion test results tables.

Standard Penetration Test									
Depth	Weight	G1	P1	G2	P2	G3	P3	NSPT	Fraction
1.00	11	1	25	1	16	1	16	1	2/32
2.00	5	1	35					1	1/35
3.00	3	1	20	1	20			1	1/20
4.00	0	3	15	4	15	3	15	7	
5.00	0	4	15	5	15	7	15	12	
6.00	0	8	15	9	15	11	15	20	
7.00	0	8	15	10	15	13	15	23	
8.00	0	12	15	13	15	15	15	28	
9.00	0	15	15	15	15	18	15	33	
10.00	0	18	15	20	15	10	5	45	30/20
11.00	0	22	15	30	10			50	30/10
12.00	0	30	5					50	30/5

Source: AGS-BR (2018).

```
"GROUP","PROJ"
"HEADING","PROJ_ID","PROJ_NAME","PROJ_LOC","PROJ_CLNT","PROJ_CONT","PROJ_ENG","PROJ_MEMO","FILE_FSET"
"UNIT","","","","","","","",""
"TYPE","X","X","X","X","X","X","X","X"
"DATA","121415","ACME Gas Works Redevelopment","Anytown","ACME Enterprises","ACME Drilling Ltd","","",""

```

Figure 4. Raw AGS-4 file excerpt (AGS-BR, 2022).

```
{
  "GROUP": "PROJ",
  "HEADING": {
    "PROJ_ID": "121415",
    "PROJ_NAME": "ACME Gas Works Redevelopment",
    "PROJ_LOC": "Anytown",
    "PROJ_CLNT": "ACME Enterprises",
    "PROJ_CONT": "ACME Drilling Ltd",
    "PROJ_ENG": "",
    "PROJ_MEMO": "",
    "FILE_FSET": ""
  },
  "UNIT": {},
  "TYPE": {
    "PROJ_ID": "X",
    "PROJ_NAME": "X",
    "PROJ_LOC": "X",
    "PROJ_CLNT": "X",
    "PROJ_CONT": "X",
    "PROJ_ENG": "X",
    "PROJ_MEMO": "X",
    "FILE_FSET": "X"
  },
```

Figure 5. Proposed file structure.

From the AGS guidelines regarding file structure, we have fully respected and incorporated the rules from 1 to 20, except the sixth rule that restricts the use of two characters, that are permitted in our format, and rules 19, 19a and 19b, rules regarding size limits to titles and data variables, which to keep compatibility to the AGS format must have an identification field, or id, to keep the same meanings across platforms.

The parts of rules 6, 11a, 11b and 16a, which state specific characters to the structural division of the file have no application in our JSON approach and are all substituted by the standard double brackets used in this format. For retro compatibility between these two formats, an algorithm that correctly converts the information between these two files should be developed. However, the development of such a tool is not in the corpus of this study.

The book Geological-Geotechnical Investigations: Good Practices Guide of the Brazilian Association of Environmental and Engineering Geology by Monticelli (2021) was also used to build the schema. Tables of field tests, borehole logs and standard reports used in the industry were used to elaborate the variables and pieces of information that are used by the parts of geotechnical projects, then the proposed model presented in Figure 6 was built:

The model represents the system's relationships and data through entities and attributes. Thus, the model above represents the AGS standard and its structure. As it is a file composed of several entities, which can be thought of as spreadsheets containing columns represented in the model from the attributes, next to the attribute nomenclature, the type of information expected in the field is declared, which can be texts, numbers, or other types of data understandable

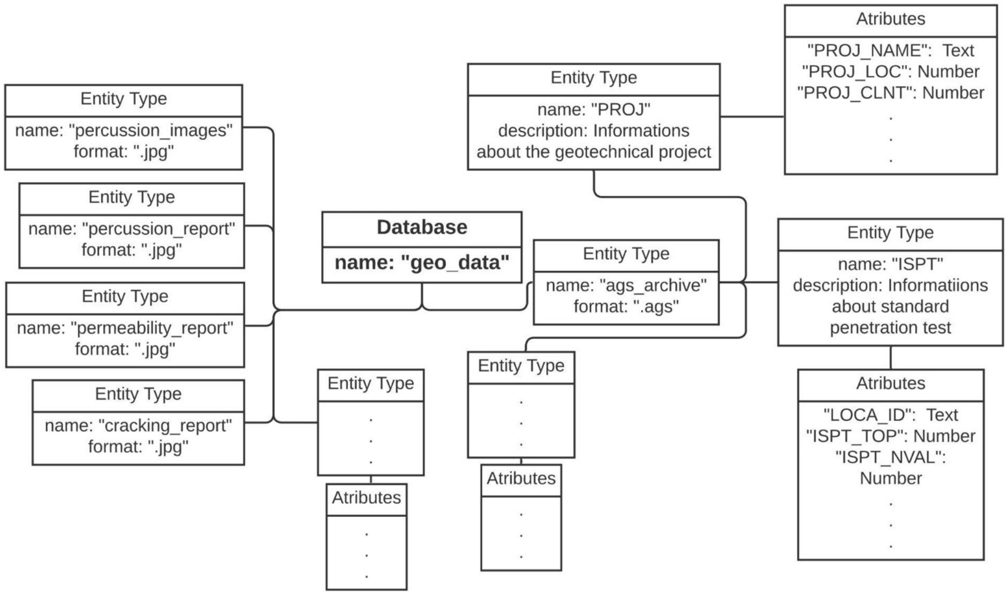

Figure 6. U-type scheme of the proposed model.

by digital systems. The three dots in parts of the schema represent continuity, that is, where there are more entities than those chosen for representation in the schema.

The nomenclature followed by the attributes and entities follows a pattern based on succinct characters and without the presence of accents and blank spaces, a mandatory format in database management systems. For example, the attribute "PROJ_NAME" represents an identifying text of the name of the geotechnical project being executed. The descriptions of AGS standard attribute names are available in the document defining it or in the Brazilian translation made by AGS Brasil.

Several data entities commonly found in the geotechnical field were arranged alongside the AGS standard, which does not have a well-defined digital structure and can be inserted into the database without generating sluggishness to the AGS standard scheme, enriching the amount of information available for download by users and enabling the processing of this information on a scale by algorithms and specific tools. In the "percussion_bulletins" entity, images of hand-filled bulletins by drillers can be inserted together with the AGS file and photos of the drilling execution. Other entities that allow the insertion of pertinent geotechnical data were inserted. There is also the possibility of developing new entities after the implementation of the database.

4. Conclusion

The geotechnical market is increasingly focused on data, modeling through artificial intelligence algorithms, and collaboration between areas of knowledge present in mining, infrastructure works, and civil construction, all activities with a high demand for quality, updated, and easily accessible data. Therefore, the construction of robust digital systems that integrate data within the various stages of a geotechnical engineering project is crucial.

The evaluation of the modeled scheme and its effectiveness through the construction of a real system and field tests is not within the scope of this research. However, the correct assessment of the functioning of the modeled scheme is crucial for its actual use and determines whether the hypothesis of a more effective system in the field is validated. For this, we consider the studies of Kokkala & Marinos (2022) and Li et al. (2022b) to primarily evaluate if the most important soil characteristics for the analyzed studies were incorporated into the proposed model.

If the model is in line with what is presented in the analyzed works, it is necessary to evaluate the methodology of constructing the database model, which follows the data modeling principles present by Atzeni et al. (2020) and the methodology of Candel et al. (2022). Thus, it can be inferred that the modeling performed has technical grounding and could be effective in the field.

Due to the complexity and plurality of important variables in various geological scenarios, this work focused only on establishing a contemporary methodology for the development of a geotechnical database. Thus, several geological investigations commonly carried out in the area, such as permeability tests and vane tests, were left out of this primary analysis.

In a scenario of application and development of the database, market solutions that are capable of receiving the proposed model should be considered. For this, some commercial products such as MySQL, PostgreSQL, and Microsoft SQL Server are not SQL type. MongoDB, Redis, DynamoDB, and Firebase are some commercial products that fit the proposed model, as they are NoSQL models.

Acknowledgements

The authors are very grateful to CNPq for the research grants given to the research.

Declaration of interest

The authors have no conflicts of interest to declare. All co-authors have observed and affirmed the paper's contents and there is no financial interest to report.

Authors' contributions

Elisangela do Prado Oliveira: conceptualization, visualization, methodology, supervision. Rogerio Rodrigues de Vargas: data curation, methodology, supervision, validation, reviewing. Jorge Brock Júnior: formal analysis, investigation, data curation, writing – review & editing.

Data availability

Data generated and analyzed during the current study are available in the pyagsapi repository, https://github.com/BritishGeologicalSurvey/pyagsapi/tree/main/test/files/real.

Declaration of use of generative artificial intelligence

This work was prepared with the assistance of generative artificial intelligence (GenAI), specifically ChatGPT GenAI was used to drafting text, brainstorming ideas and coding assistance in certain sections/elements of the manuscript. The entire process was actively supervised, thoroughly reviewed, and edited by the authors as needed to ensure accuracy and alignment with publication standards. The authors assume full responsibility for the content and conclusions presented in this publication, including any portions that involved GenAI assistance.

References

Association of Geotechnical and Geoenvironmental Specialists – AGS-BR. (2018). *Sondagem a percurssão (SP) - programação dos serviços.* AGS-BR.

Association of Geotechnical and Geoenvironmental Specialists – AGS-BR. (2022). *Electronic transfer of geotechnical and geoenvironmental data: AGS4 edition 4.1.1.* AGS-BR.

Atzeni, P., Bugiotti, F., Cabibbo, L., & Torlone, R. (2020). Data modeling in the NoSQL world. *Computer Standards & Interfaces*, 67, 103149. http://doi.org/10.1016/j.csi.2016.10.003.

Candel, C.J.F., Ruiz, D.S., & García-Molina, J.J. (2022). A unified metamodel for NoSQL and relational databases. *Information Systems*, 104, 101898. http://doi.org/10.1016/j.is.2021.101898.

Chang, F., Dean, J., Ghemawat, S., Hsieh, W. C., Wallach, D. A., Burrows, M., Chandra, T., Fikes, A., & Gruber, R. E. (2008). Bigtable: a distributed storage system for structured data. *ACM Transactions on Computer Systems*, 26(2), 1-16. http://doi.org/10.1145/1365815.1365816.

Codd, E.F. (1970). A relational model of data for large shared data banks. *Communications of the ACM*, 13(6), 377-387. http://doi.org/10.1145/362384.362685.

Guilhot, D., Martinez del Hoyo, T., Bartoli, A., Ramakrishnan, P., Leemans, G., Houtepen, M., Salzer, J., Metzger, J.S., & Maknavicius, G. (2021). Internet-of-Things-based geotechnical monitoring boosted by satellite InSAR data. *Remote Sensing*, 13(14), 2757. http://doi.org/10.3390/rs13142757.

Kim, H.-S., & Ji, Y. (2022). Three-dimensional geotechnical-layer mapping in Seoul using borehole database and deep neural network-based model. *Engineering Geology*, 297, 106489. http://doi.org/10.1016/j.enggeo.2021.106489.

Kokkala, A., & Marinos, V. (2022). An engineering geological database for managing, planning and protecting intelligent cities: the case of Thessaloniki city in Northern Greece. *Engineering Geology*, 301, 106617. http://doi.org/10.1016/j.enggeo.2022.106617.

Legget, R.F. (1973). *Cities and geology*. McGraw-Hill.

Li, K.-Q., Kang, Q., Nie, J.-Y., & Huang, X.-W. (2022a). Artificial neural network for predicting the thermal conductivity of soils based on a systematic database. *Geothermics*, 103, 102416. http://doi.org/10.1016/j.geothermics.2022.102416.

Li, Y., Rahardjo, H., Satyanaga, A., Rangarajan, S., & Lee, D.T.-T. (2022b). Soil database development with the application of machine learning methods in soil properties prediction. *Engineering Geology*, 306, 106769. http://doi.org/10.1016/j.enggeo.2022.106769.

Monticelli, J.J. (2021). *Investigações geológico-geotécnicas: guia de boas práticas*. ABGE (in Portuguese).

Phoon, K.-K., Ching, J., & Cao, Z. (2022). Unpacking data-centric geotechnics. *Underground Space*, 7(6), 967-989. http://doi.org/10.1016/j.undsp.2022.04.001.

Silberschatz, A., Korth, H.F., & Sudarshan, S. (2010). *Database system concepts* (6th ed.). McGraw-Hill.

Stonebraker, M., & Hellerstein, J.M. (2005). What goes around comes around. In M. Stonebraker & J.M. Hellerstein (Eds.), *Readings in database systems* (4th ed.). MIT Press.

Strauch, C. (2011). *NoSQL databases. Lecture selected topics on software-technology ultra-large scale sites, manuscript* (149 p.). Stuttgart Media University. Retrieved in April 12, 2024, from http://www.christof-strauch.de/nosqldbs.pdf