MODELING OF SPATIAL DATABASE FOR GEOTECHNICAL DATA

Bruno Rodrigues de OLIVEIRA, Newton Moreira de SOUZA, Rafael Cerqueira SILVA,
Eleudo Esteves de Araujo SILVA JUNIOR

University of Brasília. Faculdade de Tecnologia. Bloco C – ENC. Campus Universitário Darcy Ribeiro, Brasília-DF.
E-mails: brunorodriguesoli@gmail.com; nmsouza@unb.br; rafael.silva@unb.br; eleudo@unb.br

Introduction
Geographical data
Conceptual modeling
Understanding the problem and requirements
Requirements gathering
Field investigations
Laboratory tests
Samples
Trenches
Administrative regions
Boundary of the Distrito Federal
Spatial and spatial-time relationships
Database conceptual modeling
Conclusions
References

RESUMO - Os dados geotécnicos frequentemente são produzidos e gerenciados por diversos produtores de forma isolada, em formatos e padrões próprios, os quais visam atender única e exclusivamente às necessidades individuais de usuários específicos. No Distrito Federal existe uma grande quantidade de dados geotécnicos oriundos de pesquisas acadêmicas, órgãos executores de obras e do setor privado, mas o Governo do Distrito Federal não dispõe de uma gestão de informações geotécnicas adequada apesar de utilizar constantemente essas informações. Este trabalho apresenta uma proposta modelo conceitual de banco de dados geotécnicos, mais especificamente para investigações de campo e ensaios de laboratório. Para a modelagem foi utilizado a Técnica de Modelagem de Objetos para Aplicações Geográficas, adotado pela Infraestrutura de Dados Espaciais do Distrito Federal (IDE/DF), e compatível com as especificações homologadas pela Comissão Nacional de Cartografia, que define padrões de estrutura de dados para levantamentos topográficos de grandes escalas facilitando o compartilhamento de dados espaciais, a interoperabilidade e a racionalização de recursos entre os produtores e usuários de dados e informação cartográfica entre diferentes instituições. Essa abordagem permitirá a implantação do modelo de dados geotécnicos proposto na IDE/DF e a sua integração futura com a Infraestrutura Nacional de Dados Espaciais (INDE).


ABSTRACT - Geotechnical data is often produced and managed by several producers in isolation, in their own formats and standards, which aim to meet the unique needs of specific users. In the Federal District in Brazil there is a large amount of geotechnical data from academic research, public and private sector, but the Government of the Federal District does not have adequate geotechnical information management despite using that information constantly. This work presents a conceptual model of a geotechnical database, more specifically for field investigations and laboratory tests. For modeling the object modeling technique for geographic application, adopted by the Spatial Data Infrastructure of the Federal District (IDE / DF), was used and compatible with the specifications approved by the National Cartography Commission, which defines data structure standards for large-scale topographic surveys facilitating the sharing of spatial data, interoperability and the rationalization of resources between producers and users of data and information map between different institutions. This approach will allow the implementation of the geotechnical data model proposed at SDB/DF and its future integration with National Spatial Data Bank (NSDB).

Keywords: Conceptual data base model. Class diagram. Object modeling technique.

INTRODUCTION

After the natural disasters that occurred in Brazil, especially in the years 2010 and 2011, there was a growing awareness about the importance of improving disaster risk management in Brazil, which resulted in the institution of the National Civil Defense and Protection Policy (Política Nacional de Proteção e Defesa Civil - PNPDEC) by Law N°. 12,608 of April 10, 2012 (BRASIL, 2012) which, in addition to other measures, introduced changes to the City Statute (Estatuto da Cidade, BRASIL, 2001) and the Soil Installment Law (Lei de Parcelamento do Solo, BRASIL, 1979) related to geotechnical maps and the mapping of areas susceptible to geodynamic phenomena.

The Federal Government then began to encourage the elaboration of geotechnical maps, which consequently results in the production of data that subsidize their elaboration, which added to the data produced by the public sectors (e.g.: universities, agencies responsible for the drainage system and highways), and private (e.g.: survey
companies, laboratories and universities) results in a large volume of data.

Data production is not a problem since the production of geospatial data and information is facilitated by the accessibility resulting from technological advances.

However, geotechnical data is often produced, managed, used and stored by several producers in isolation, in their own formats and standards, which aim to meet the unique needs of individual users.

In 2017, the National Commission for Brazilian Cartography (Comissão Nacional de Cartografia Brasileira - CONCAR) presented version 3.0 of the Technical Specifications for Structuring Digital Vector Geospatial Data (Especificações Técnicas para Estruturação de Dados Geoespaciais Digitais Vetoriais - EDGV) which aims to standardize the vector reference geospatial data structures produced to compose cartographic bases related to scales 1:1,000 and lower. However, geotechnics is not one of the themes addressed by EDGV and it is the responsibility of other institutions to structure this data, which in the case of geotechnics, does not have an institution responsible for the theme.

The Geological Survey of Brazil (Serviço Geológico Brasileiro – SGB), despite being the agency responsible for preparing the cartographic products provided for in Law No. 12,608 of April 10, 2012 (BRASIL, 2012) and having a well-stabilized database, it does not have a standardization for the storage of geotechnical data and does not provide its model geological data.

Therefore, this work aims to present a proposal for a conceptual model of geographic database using the Object Modeling Technique for Geographic Applications (OMT-G) for the theme Geotechnics, more specifically for field investigations and laboratory tests, compatible with the Spatial Data Infrastructure of the Federal District (SDI/DF) and the Geographic and Vector Data Structure (EDGV) and its implementation in a database management system (SGBD) free, PostgreSQL / PostGIS.

**GEOGRAPHICAL DATA**

Geoinformation is the information that is distinguished by the spatial component, where each information record of a phenomenon has a location on Earth, at a given time or period.

To create or import data in a Geographic Information System (GIS), it is necessary to choose a geometric primitive, which is defined by rules in the conceptual representation and by the relationship with other features. That is, the primitive influences the data visualization, topology, and the possibility of manipulating and data options selections. Two-dimensional space objects are a representation of the real world, modeled by lines, points or polygons (Figure 1), associated with a location on Earth through pairs of coordinates, linked to a geodetic reference system.

The success of any computer implementation of an information system is dependent on the quality of the transposition of real-world entities and their interactions to a computerized database. Abstraction works as a tool that helps us to understand the system, dividing it into separate components (Borges et al., 2005).

The process of space discretization, as part of the abstraction process, aiming at obtaining adequate representations to geographic phenomena, depends on the way people perceive space, on the diversified nature of geographic data (geometry, location in space, associated information and temporal characteristics), the existence of spatial relationships and the form of transcription of geographic information into logical data units. In this way, it is necessary to...
construct an abstraction of objects and phenomena in the real world to obtain a convenient, though simplified, form of representation that is suitable for the purposes of the database applications (Câmara et al. 2005).

For geographic applications, four levels of abstraction are considered (Figure 2): the level of the real world that contains the real geographical phenomena to represent, such as rivers, streets and vegetation cover; the level of conceptual representation that offers a set of formal concepts with which geographic entities can be modeled as they are perceived by the user, at a high level of abstraction; the level of presentation that provides tools with which you can specify the different visual aspects that geographic entities have to assume during their use in applications; and the level of implementation that establishes standards, forms of storage and data structures to implement each type of representation, the relationships between them and the necessary functions and methods (Borges et al., 2005).

![Abstraction levels of geographic applications](image)

**Figure 2** - Abstraction levels of geographic applications (Adapted from Borges et al., 2005).

**CONCEPTUAL MODELING**

A data model is a set of concepts that can be used to describe the structure and operations in a database (Elsmari & Navathe, 2006). The first data models for geographic applications were aimed at the internal structures of the GIS and the user was forced to adapt the spatial phenomena to the structures available in the GIS to be used (Borges et al., 2005), generating models farther from the user's mental model. In view of the need to represent the complexity of phenomena and objects in the real world, models emerged capable of presenting a better abstraction of concepts, types of entities and their interrelationships.

The choice of the model should be made keeping in mind the modeling needs regarding the abstraction of geographical concepts, meeting the usual requirements for data models, such as clarity and ease of use (Borges et al., 2001), and the possibility of mapping the schemes produced for implementation in spatial DBMS, which includes the necessary identification of spatial integrity constraints (Borges et al., 2002; Davis Jr. et al., 2005).

Among the existing models, the OMT-G model starts from the primitives defined for the class diagram of the Unified Modeling Language (UML), introducing geographical primitives in order to increase the semantic representation capacity of that model and therefore reducing the distance between the mental model of the space to be modeled and the usual model of representation (Borges et al., 2005) being able to model the geometry and topology of the data and specify alphanumeric attributes and associated
methods for each class.

The OMT-G model is based on three main concepts: classes, relationships and spatial integrity constraints used to create static application schemes. The classes defined by the OMT-G model represent the three large groups of data (continuous, discrete, and non-spatial) that can be found in geographic applications, thus providing an integrated view of the modeled space.

The classes can be of the georeferenced type, which describes a set of objects represented spatially and associated with regions of the Earth's surface, or conventional (non-spatial), which do not have geometric properties, however, show behaviors, relationships, and some relationship with spatial objects (Figure 3) (Borges et al., 2005).

Georeferenced classes are specialized in geo-field classes, which represent objects or phenomena spatially in a continuous way such as soil or geology, and of the geo-object type that represent individualizable geographic objects, associated with elements of the real world, such as buildings, rivers and trees (Figure 4).

**Understanding the problem and requirements**

The process begins by choosing the objects of interest, that is, the phenomena of interest observed in the “real world” that will be represented by conventional or georeferenced classes during abstraction in the conceptual model, which will later compose the database. For these objects of interest, the requirements of the information itself and the spatial data that will represent them are raised.

**Requirements gathering**

Despite the various types of investigations and geotechnical tests, this work chose to work with the main types of investigations and tests carried out in academic research and executing agencies of the Federal District Government (GDF) but guaranteeing the possibility of expanding the proposed conceptual model.

In this research the main objects of interest of the model are field investigations, laboratory tests and samples.

**Field investigations**

Method of obtaining information in the field, on the surface or on the subsurface, in which the researcher may or may not have contact with the sampled material to obtain its physical properties (Adapted from Marrano et al., 2018).
Investigations are represented by point-type geo-objects which can overlap if they are not carried out in the same period. This class must be within the administrative boundary. Table 1 presents the subclasses of the investigations and their respective spatial restrictions and descriptions, which were based on Marrano et al. (2018) and NBR 8044:2018.

**Table 1 - Subclasses of the investigation’s superclass and their respective descriptions and spatial restrictions.**

<table>
<thead>
<tr>
<th>Subclasses of the investigations</th>
<th>Description</th>
<th>Spatial restrictions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Field Points</td>
<td>Any point on the Earth's surface or underground that contains information relevant to an engineering project.</td>
<td>Inherits superclass restrictions and the unique identifier.</td>
</tr>
<tr>
<td>Concentric Rings</td>
<td>Field test to determine the rate of water infiltration into the soil.</td>
<td></td>
</tr>
<tr>
<td>Palette</td>
<td>Field test to determine undrained resistance, dented undrained resistance and sensitivity of so-called soft soils.</td>
<td></td>
</tr>
<tr>
<td>Piezometer</td>
<td>Field test that allows to check the water pressures and the position of the phreatic level in the rock mass and soil respectively.</td>
<td>Inherits superclass restrictions and the unique identifier. Relationship with conventional tables with measurement values.</td>
</tr>
<tr>
<td>Guelph permeometer</td>
<td>Field testing allows the assessment of the <em>in situ</em> permeability coefficient of the soil.</td>
<td>Inherits superclass restrictions and the unique identifier. Relationship with conventional tables depending on the type of borehole.</td>
</tr>
<tr>
<td>Boreholes</td>
<td>Field test to collect deformed samples of soil or rock depending on the type of borehole. The drilling is commonly used to perform soil infiltration tests or loss of water by pressure on rocks. The auger, PANDA penetrometer, rotary, cone or piezocone and standard penetration test are types of drilling, whose information is stored in conventional classes related to this subclass.</td>
<td></td>
</tr>
</tbody>
</table>

**Laboratory tests**

Tests performed inside a laboratory, on soil or rock samples, to obtain the physical, mineralogical, mechanical, and hydraulic properties of the materials and/or to categorize the tested materials according to their geotechnical properties (Adapted from Head, 2006). The tests are represented by point type geo-objects which cannot overlap, regardless of the date of execution. This object must be within the administrative boundary. Table 2 presents the subclasses of the tests and their respective spatial restrictions and descriptions which were based on Head (2006), Head & Epps (2011; 2014) and NBR 8044:2018.

**Samples**

Rocky or earthy material collected through field investigations that can be used to perform laboratory tests. The deformed sample is one that does not maintain all the characteristics that occur *in situ* and the undisturbed sample is obtained to preserve the characteristics that occur *in situ* (NBR 6502/1995). The samples are represented by polygon geo-objects which can only overlap if they were carried out at different depths, regardless of the date of execution. The sample must always be related to the investigation in which the collection was performed, but it does not need to be related to laboratory tests as it may not have been subjected to tests. This object must be within the administrative boundary.

**Trenches**

Stores information regarding wells and trenches. The manhole on the ground is a vertical excavation (circular or square section) that aims to allow the access of an investigator to make a visual inspection of the walls and bottom and the removal of representative samples - undeformed and / or deformed. The trench is a vertical excavation (of rectangular section) made to obtain a continuous exposure of the soil in a certain part of the terrain (NBR 9604: 2016).

Although the trench is also a field investigation, we chose to represent them by polygon-type geo-objects which may not overlap and must be within the administrative boundary. This choice is made because the generation of centroid inserted inside the polygon that originates it is a simpler procedure than generating the polygon from a point, besides allowing to store different excavation formats.
Table 2 - Subclasses of the tests superclass and their respective descriptions and spatial restrictions.

<table>
<thead>
<tr>
<th>Subclass of the tests</th>
<th>Description</th>
<th>Spatial restrictions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Humidity</td>
<td>Laboratory test performed to determine the amount of water present in the soil structure.</td>
<td>Inherits superclass restrictions and the unique identifier.</td>
</tr>
<tr>
<td>Atterberg</td>
<td>Laboratory test performed to measure and describe the range of soil plasticity in numerical terms.</td>
<td></td>
</tr>
<tr>
<td>Apparently Density</td>
<td>Laboratory test carried out to determine the apparent density considering the total volume of the sample, including the empty space between the grains that compose it.</td>
<td></td>
</tr>
<tr>
<td>Actual grain density</td>
<td>Laboratory test performed to determine the real or relative density considering only the volume of the set of grains that make up the sample, without considering the empty space between the grains.</td>
<td></td>
</tr>
<tr>
<td>Granulometry</td>
<td>Laboratory test for obtaining the distribution of soil particles.</td>
<td></td>
</tr>
<tr>
<td>Compaction</td>
<td>Laboratory test to determine the adequate state of soil compaction for the construction of engineering works.</td>
<td></td>
</tr>
<tr>
<td>Constant Load Permeability</td>
<td>Laboratory test to measure the ability of a fluid to flow through its structure. Used on non-cohesive soils.</td>
<td></td>
</tr>
<tr>
<td>Variable Load Permeability</td>
<td>Laboratory test to measure the ability of a fluid to flow through its structure. Used in cohesive soils.</td>
<td></td>
</tr>
<tr>
<td>Simple Compression</td>
<td>Laboratory test to measure the unconfined compressive strength of cohesive soils.</td>
<td></td>
</tr>
<tr>
<td>California Support Index (ISC)</td>
<td>Laboratory test to measure the carrying capacity of the sub-base and sub-grade.</td>
<td>Inherits superclass restrictions and the unique identifier.</td>
</tr>
<tr>
<td>Direct Shear</td>
<td>Laboratory test to determine soil shear strength parameters (cohesion and friction angle)</td>
<td>Relationship with conventional tables with measurement values.</td>
</tr>
<tr>
<td>Densification</td>
<td>Laboratory test determines the compressibility characteristics of soils under the condition of lateral confinement. Laboratory tests composed of Mini-MCV, loss of mass by immersion and specific classification.</td>
<td></td>
</tr>
<tr>
<td>MCT</td>
<td>Laboratory test to determine soil resistance and deformability parameters.</td>
<td></td>
</tr>
<tr>
<td>Triaxial</td>
<td>Laboratory test for determining the unconfined compressive strength of cohesive soils.</td>
<td></td>
</tr>
</tbody>
</table>

Administrative Regions
The administrative regions of the Federal District are the administrative divisions of the Distrito Federal defined by law. The boundary of the administrative regions of the Distrito Federal is represented by a geo-object of the multi-polygon type in which there can be no overlap or gaps and must be contained in the boundary of the Distrito Federal.

Boundary of the Distrito Federal
Administrative boundary of the Federal District, which is one of the 27 federative units in Brazil. The administrative boundary of the Distrito Federal is represented by a polygon-type geo-object formed by the spatial aggregation relation that confers the constraint that overlapping or gaps within it cannot occur.

Spatial and Spatial-Time relationships
All geo-objects, whether polygons or points, have a spatial relationship with each other or with conventional classes. These relationships are specified in the conceptual model. The date of execution or registration of an investigation is recorded in each of the classes of the model.

DATABASE CONCEPTUAL MODELING
The geotechnical data were abstracted in the investigations and tests superclasses and in the sample class. The investigations superclass is responsible for generating the unique identifier code of the geotechnical data, here called geociu, relating field investigations to samples and tests. The investigations and tests store the information common to their respective subclasses while the class of samples is responsible for relating the investigations and tests through the unique identifier code.

Investigations and tests are represented as geo-objects of the point type while samples as a geo-object of the polygon type. All three classes must be within the administrative boundary, which in the case of the proposed model is the DF limit generated from the spatial aggregation of the class of administrative regions.

From the investigation’s superclass, any number of subclasses can be derived if it meets its definition. In the proposed model, the investigations superclass, through specialization
using the attribute type of investigation, gives rise to the subclasses Field Points, Concentric Rings, Guelph Permeameter, PANDA Penetrometer, Cone or Piezocone, Palette, Piezometer and Standard Penetration Test (SPT). Each of these objects is responsible for storing information related to geotechnical investigation of the subclass type. The Field Points and Concentric Rings classes inherit the geometry of the superclass and can be treated as conventional classes during database implementation and, more specifically in the field class, it was decided to segment it into two tables during the implantation, one to store the soil profiles and another to store the rock outcrops.

In the relationship between the investigation’s superclass and their respective subclasses, the special overlapping partial relationship was adopted, overlapping because it defines that two investigations can be carried out in the same location and partial because the subclasses presented do not constitute all possibilities for field investigations.

This relationship validates situations such as borehole followed by the installation of a piezometer but does not prevent the case of a PANDA penetrometer borehole followed by the execution of a penetration test, whose problem would be that the soil would no longer be in the conditions in situ. Time could be a variable that makes the second example feasible if the tests were not performed at the same time.

The rotary, percussion and auger borehole are generated by consulting the investigations superclass, the borehole subclass and the conventional rotating, percussion, and auger tables respectively. In the DBMS, this selection was made by creating a materialized view.

As for the trench, it was decided to keep it as a class not derived from the investigations so that the geometry would be more adequate to represent the geometry of the investigation. The points corresponding to the trench in the investigation table are obtained by determining the centroid inserted within the polygon that originates it (Figure 5).

In the case of the test’s superclass, any number of subclasses can be derived if it meets the definition of the superclass. In the proposed model, the tests superclass, through specialization using the test type attribute, gives rise to the subclasses Humidity, Atterberg, Density, Granulometry, Constant and Variable Load Permeability, California Support Index, Direct Shear, Simple Compression, Densification, MCT and Triaxial.

The test subclasses inherit the geometry of the test superclass and can be treated as conventional classes during implantation. The dimensions of the tested samples are represented in the sample class.

In the relationship between the test’s superclass and their respective subclasses, the partial

Figure 5 - Transformations in the database to obtain new classes.

In the case of the test’s superclass, any number of subclasses can be derived if it meets the definition of the superclass. In the proposed model, the tests superclass, through specialization using the test type attribute, gives rise to the subclasses Humidity, Atterberg, Density, Granulometry, Constant and Variable Load Permeability, California Support Index, Direct Shear, Simple Compression, Densification, MCT and Triaxial.

The test subclasses inherit the geometry of the test superclass and can be treated as conventional classes during implantation. The dimensions of the tested samples are represented in the sample class.

In the relationship between the test’s superclass and their respective subclasses, the partial
A disjoint, disjointed specialization relationship was adopted, since two tests cannot be performed on the same undeformed and partial sample because the subclasses presented do not constitute all possibilities for geotechnical tests. The other conventional tables are intended to store the results of test measurements. Based on the analyzes performed regarding the topological and geometric characteristics of the classes, a conceptual model is proposed that represents the investigations and geotechnical tests in an objective and coherent way. The proposed conceptual model (Figure 6) was generated using the OMT-G Designer module for visualization and modeling of OMT-G diagrams available for free at the following website: http://aqui.io/omtg-designer/.

Figure 6 - Conceptual model OMT-G proposed for the geotechnical database.

Considering the peculiarities of the classes Concentric Rings, Field Points, the subclasses of the tests superclass and the need for some investigations and tests requiring additional tables to store the results, a logical scheme was elaborated, without the information of the attributes to facilitate the visualization, showing the database implementation (Figure 7).

Figure 7 shows the tables mct_mini_mcv, triaxial_medicao, isc_medicao, isc_expansao, guelph_medicao, cisalhamento_direto_medicao, adensamento_medicao, which are complementary.
tables to store the results of test measurements. Other tables, such as domain tables, are optional and therefore are not provided for in the logical scheme.

Figure 7 - OMT-G logic scheme proposed for the geotechnical database.
CONCLUSIONS

Considering the reality of the Distrito Federal, in which more than four thousand geotechnical investigations that were compiled to compose this database were restricted in their respective sources, there is a real need to build a geospatial database that is compatible with the SDI / DF and National Spatial Data Bank (NSDB) to disseminate this information.

The great advantage of implementing the proposed model is the possibility of systematic and periodic organization of data produced by different agencies or companies in the Distrito Federal. The adoption of the model and the implementation of the database would avoid investments by different agencies or companies in the same region, optimizing the execution of future construction projects and assisting in urban environmental planning.

In the case of the OMT-G model, this proved to be appropriate to obtain adequate representations of the tests and investigations. Relationships as specialization can define more specific classes from generic classes, adding new properties in the form of attributes, as is the case with field investigations and laboratory tests and their respective subclasses. This type of relationship also allows specifying that two field tests can be done in the same place as a percussion survey followed by the installation of a piezometer (Specialized Overlapping), but two tests cannot be done on the same sample as a particle size and a density test (Disjoint Specialization).

As for the temporal variation, although the OMT-G model is not intended to model data with this property, due to the characteristics of the geotechnical data queries related to the date of execution or registration of an investigation are easily constructed and enough to represent the temporal variation of geotechnical data.

Regarding the DBMS, PostgreSQL is extremely robust and stable to serve as a base for the database and allows to store a large amount of data, multiuser access, change control and access to information in addition to being the DBMS used by the database. IDE / DF, which facilitates the adoption of this standard. The PostGIS spatial extension (2020) has a wide range of spatial functions, integrates with GIS such as Quantum GIS and moves towards 3D storage and analysis.

Finally, this model will serve as the basis for the development of an application for managing the geotechnical database in Quantum GIS and will later be expanded to include more objects of interest in the Geotechnical theme, the storage of geotechnical data in three dimensions and greater interoperability with other databases like the Federal District Multifinalitary Technical Cadastre (Cadastro Técnico Multifinalitário do Distrito Federal CTM/DF). The project is ongoing, and it is available at https://github.com/bro-geo/geotechnical_database as it is being implemented and full tested.

REFERENCES


