DOI: 10.5965223811712342024732

Revista de Ciências Agroveterinárias 23 (4): 2024
Universidade do Estado de Santa Catarina

Conceptual modeling of database object-oriented applied to soil studies – DATASOLOS_SC

Modelo conceitual de banco de dados orientado a objeto aplicado a estudos pedológicos – DATASOLOS_SC

Pablo Grahl dos Santos *(ORCID 0000-0002-3509-9616), Jaime Antonio de Almeida (ORCID 0000-0001-5808-9421)

Santa Catarina State University, Lages, SC, Brazil. *Corresponding author: pablograhlsantos@gmail.com

Submission: 10/01/2024 | Accepted: 29/07/2024

RESUMO

Sistemas de informação são capazes de armazenar dados importantes de modo que estes possam ser consultados e analisados para os mais diversos usos e aplicações. A maneira como os dados são armazenados em um banco de dados facilita a organização, a consulta e a atualização das informações. A modelagem conceitual tem sido aplicada com sucesso na construção de bancos de dados permitindo representar, de maneira abstrata, formal e não ambígua, a realidade da aplicação. O objetivo deste trabalho foi projetar e implementar uma estrutura de um banco de dados por meio de modelagem conceitual para aplicações na área da Ciência do Solo através do modelo OMT, visando à organização, sistematização e gerenciamento da informação disponível sobre este recurso natural, por meio de um sistema gerenciador de banco de dados objeto relacional - SGBD. O DATASOLOS_SC reúne dados gerais, físicos, químicos e morfológicos de perfis modais de solos de Santa Catarina, além de uma biblioteca de amostras de solo (banco físico). A versão inicial do banco de dados estará disponível para acesso no endereço eletrônico https://datasolos-sc.sistemas.udesc.br/.

PALAVRAS-CHAVE: banco de dados; orientação a objetos; solos.

ABSTRACT

Information systems are able to store important data so as that these can be consulted and analyzed for the most varied uses and applications. The way how the data is stored in a database makes it easy the organization, query and update of information. The conceptual modeling has been successfully applied in the building of databases enabling represent in an abstract way, formal and unambiguous, the reality of application. The objective of this study was to design a structure of a database by means of conceptual modeling for applications in the field of soil science through of OMT model, aimed the organizing, systematizing, and managing of the information available about this natural resource, through a relational object database management system. The DATASOLOS_SC gathers general, physical, chemical and morphological data of soil profiles from Santa Catarina State – Brazil, besides a soil samples library (physical bank). The initial version of the soil database will be available for access at the electronic address https://datasolos-sc.sistemas.udesc.br/.

KEYWORDS: database; object orientation; soils.

Detailed knowledge of natural resources and physical environmental characteristics is essential for assessing land use potential for various agricultural activities. To analyze factors related to soils, it is necessary to create a consistent, georeferenced and quantitative database (COOPER et al. 2005). The way data is stored in a database makes it easier to organize, consult and update information (ASSAD & SANO 1998).

Due to the need for information to support multidisciplinary studies involving soils and their relationships with other natural resources, international institutions have developed various expert information systems. 2004). Among existing initiatives, the Digital Soil Map of the World stands out as the sole global-scale soil database (FAO 1996). SOTER - The World Soils and Terrain Database emerged from efforts to compile, update, and expand the FAO database, based on mapping landscape units with distinct patterns of lithology, geomorphology, and soils (VAN ENGELEN & WEN 1995). The Canadian Soil Information System (CANSIS) is a geographic database that integrates climate, soil, and land use data to assess agricultural crop yields across diverse Canadian landscapes (MACDONALD & KLOOSTERMAN 1984). The National Soil Information System (NASIS) emerged from efforts to digitize the vast soil survey data collected by the United States Department of Agriculture (SOIL SURVEY STAFF 1991). And SIGSOLOS - Brazilian Soil Georeferenced Information System (EMBRAPA 1998).

Brazil is at the forefront of tropical soil research; however, traditional methods of disseminating this information have proven inefficient due to the large volume and lack of standardization of data, limiting its utilization. However, with the advent of modern computing technologies, conditions are created so that all this existing collection can be managed by information systems, composing a georeferenced database (BHERING et al. 1998). Soil information systems can include data with spatial dimensions or point-specific data that describe and quantify specific properties of a soil profile (BAUMGARDNER 1999). This study aimed to develop a prototype of an object-relational database structure (DATASOLOS_SC) through an object-oriented conceptual model applied to Soil Genesis and Classification studies to enable data management and systematize public access to soil information from Santa Catarina State, derived from Master's and Doctoral research projects linked to UDESC's Graduate Program in Soil Science.

In designing an information system, database design represents a critical phase. The traditional database design approach encompasses three distinct phases: conceptual design, logical design, and physical design (ELMASRI & NAVATHE 2017). According to LISBOA FILHO (2000) and LISBOA FILHO et al. (2000) data modeling is the process of real-world abstraction that emphasizes only the essential elements of observed reality, providing a notational and semantic foundation, encompassing the description and definition of data contents, structures, and rules. Generally, conceptual schemes are methodologically used in the development of databases (NETO 2000, NETO et al. 2006). According to CEN (1996), modeling is invariably based on a conceptual formalism (relational, hierarchical, object-oriented), regardless of the level of abstraction employed.

Several data conceptual models currently exist, which differ primarily in their conceptual and relational formalism. In this study, we employed the conventional Object Modeling Technique (OMT) developed by RUMBAUGH et al. (1991), due to its ability to represent semantic aspects of an application through its object-oriented approach and widespread use in modeling. Currently, there are several languages for specifying class diagrams according to object-oriented formalism. In this study, we opted to use the UML (Unified Modeling Language) class diagram notation (BOOCH et al. 1998, FURLAN 1998) using StarUML™ The Open Source UML/MDA Platform Version 5.0.2.1570 software. The conceptual model of the database (Figure 1) was designed based on basic technical concepts in the application area, with its logical design and physical structure constructed using the Microsoft Access 2007® application (Figure 2). Access was selected due to its relational database management capabilities, enabling data linking for visualization, querying, editing, and reporting, while offering versatile data import/export functionality across multiple formats and user-friendly interface that requires no prior programming or advanced database language expertise.

The data and information were obtained through technical and scientific projects conducted by the Soil Genesis and Mineralogy Laboratory at the Center for Agricultural and Veterinary Sciences (CAV) of UDESC, with or without support from partner institutions, thus holding copyright and/or intellectual property rights over this database.

The project aims to systematize public access to information for professional, personal, or academic purposes, requiring mandatory citation of the source in publications that directly or indirectly use data and/or information from the soil database.

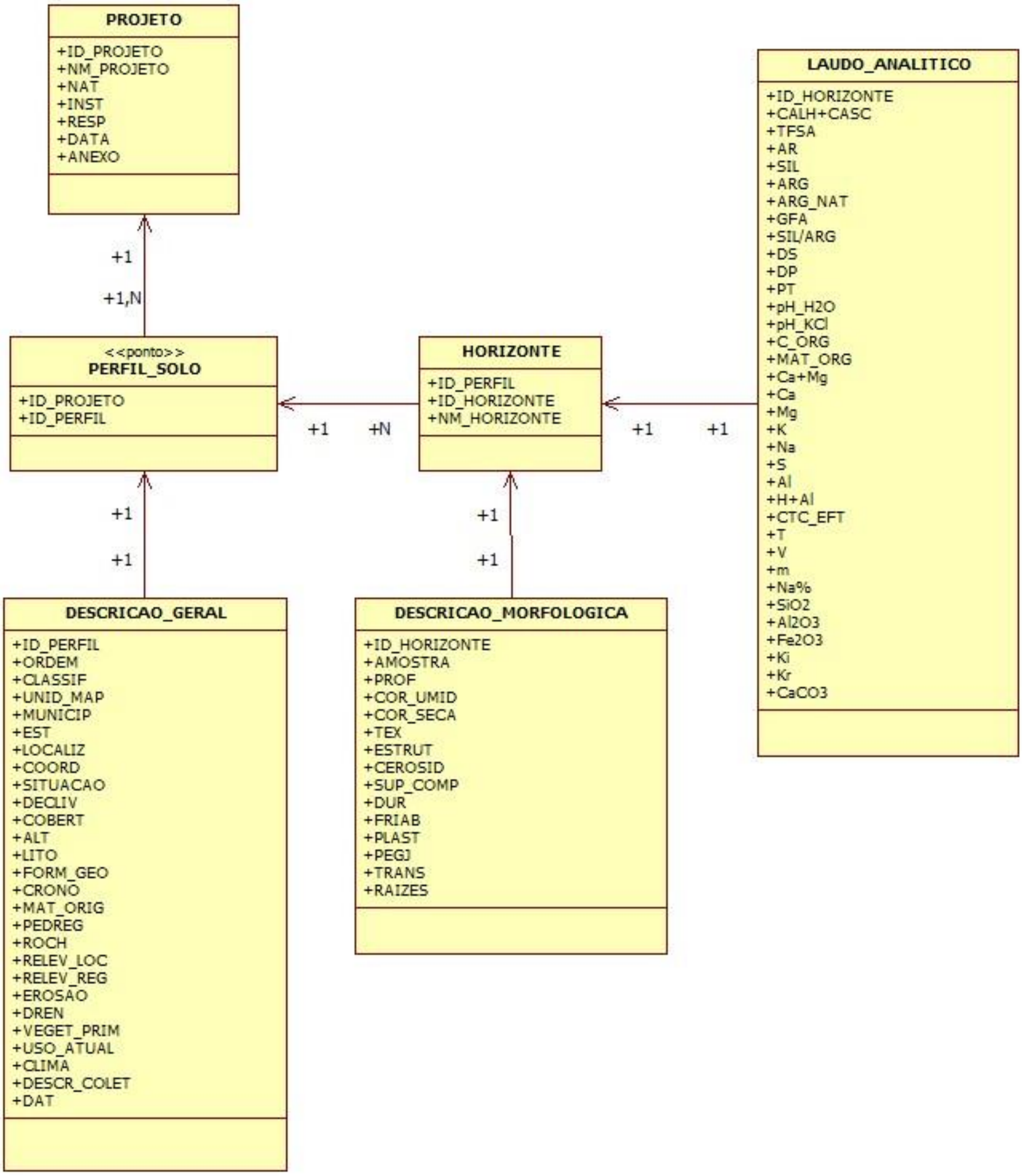

Figure 1. Conceptual model of the DATASOLOS_SC with the classes, attributes and relationships.

The input data were selected from 17 academic works (theses, dissertations, scientific articles, and technical reports), initially comprising 123 complete soil profiles, totaling 790 samples (horizons), including general, physical, chemical, and morphological data from soils in Santa Catarina (Tables 1 to 6).

DATASOLOS_SC also maintains a physical soil library where samples are coded and stored, available upon request for complementary analyses in other research projects.

Version 1.0 of DATASOLOS_SC will be available in the future for access on the internet at the following electronic address: https://datasolos-sc.sistemas.udesc.br/.

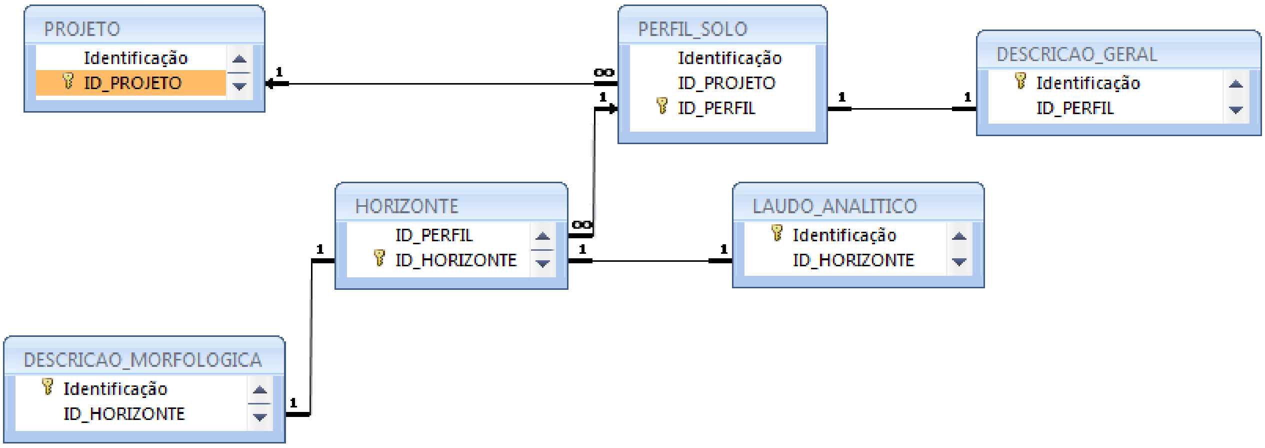

Figure 2. Logical project of the DATASOLOS_SC.

Table 1. Attributes list of the PROJECT class with the fields description.

Name Field	Type	Size	Description
PROJECT_ID	Text	2-3	Project identification code in the database (capital letters of the student/researcher's first/last name)
NM_PROJECT	Text	255	Project name (title of project or academic work)
NAT	Text	255	Nature of work (thesis/doctorate, dissertation/master's, scientific article, report, others)
INST	Text	255	Executing institution (acronym)
RESP	Text	255	Responsible (Coordinator/advisor; technical team; student)
DATE	Text	6	Abbreviated date of publication or defense of the work (mmm-yy)
ANNEX	Annex		PDF extension file of the work

Table 2. Attributes list of the SOIL_PROFILE class with the fields description.

Name Field	Type	Size	Description
PROJECT_ID	Text	2-3	Project identification code in the database (capital letters of the student/researcher's first/last name)
ID_PERFIL	Text	4-10	Soil profile identification code in the database (profile name in the work plus the project identification code in the database)

Table 3. Attributes list of the HORIZON class with the fields description.

Name Field	Type	Size	Description
ID_PERFIL	Text	4-10	Soil profile identification code in the database (profile name in the work plus the project identification code in the database)
ID_HORIZON	Text	3-5	Identification code of the soil horizon or layer in the physical and digital soil bank (sequential number of the sample plus the project code)
NM_HORIZON	Text	1-10	Horizon name (symbol - nomenclature of the horizon or soil layer)

Table 4. Attributes list of the GENERAL_DESCRIPTION class with the fields description.

Name Field	Type	Size	Description
ID_PERFIL	Text	4-10	Soil profile identification code in the database (profile name in the work plus the project identification code in the database)
ORDER	Text	11	Soil class at the 1st categorical level of the Brazilian Soil Classification System (Embrapa) (capital letters)
CLASSIF	Text	255	Taxonomic classification at subgroup level according to the Brazilian Soil Classification System (Embrapa)
UNID_MAP	Text	55	Mapping Unit (Soil Survey)
MUNICIP	Text	25	Name of the Municipality
EST	Text	2	Name of the State of the Federation (Acronym)
LOCALIZ	Text	255	Location (description)
COORD	Text	55	Geographic Coordinates LAT; LONG (dd mm ss,ss) - Geodetic Reference System (Horizontal*Datum* WGS84)
SITUATION	Text	255	Soil profile situation in the landscape
DECLIV	Text	1-5	Slope (%) absolute value or range
COBERT	Text	55	Vegetation cover over the soil profile
ALT	Text	10	Altitude (m) Orthometric or Geometric
LITO	Text	55	Lithology
FORM_GEO	Text	100	Geological Formation
CHRONO	Text	55	Chronology
MAT_ORIG	Text	255	Soil parent material
PEDREG	Text	55	Stony (class)
ROCH	Text	55	Rockiness (class)
RELEV_LOC	Text	55	Local relief (class)
RELEV_REG	Text	55	Regional relief (class)
EROSION	Text	25	Type and degree of soil erosion
DREN	Text	55	Drainage (class)
VEGET_PRIM	Text	100	Primary vegetation
CURRENT_USE	Text	55	Current land use
CLIMA	Text	3	Climate subtype according to Koppen classification (symbol)
DESCR_COLET	Text	255	Described and collected by (names)
DAT	Text	10	Date of profile description and sample collection (dd/mm/yyyy)

Table 5. Attributes list of the MORPHOLOGICAL_DESCRIPTION class with the fields description.

Name Field	Type	Size	Description
ID_HORIZON	Text	3-5	Identification code of the soil horizon or layer in the physical and digital soil bank (sequential number of the sample plus the project code)
SAMPLE	Text	1	Do you have a soil sample available in the physical bank (Y – yes; N – no)
PROF	Text	25	Depth/thickness of soil horizon or layer (cm)
COR_UMID	Text	25	Moist soil color (Munsell code)
COR_SECA	Text	25	Dry soil color (Munsell code)
TEX	Text	55	Soil texture (textural class or grouping)
STRUCTURE	Text	255	Soil structure (grade, size and type)
CEROSID	Text	100	Waxiness (degree of development and quantity)
SUP_COMP	Text	25	Compression or friction surfaces
DUR	Text	55	Hardness or tenacity (dry consistency of the soil)
FRIAB	Text	55	Friability (moist consistency of the soil)
PLAST	Text	55	Plasticity (wet consistency of the soil)
PEGJ	Text	55	Stickiness (wet consistency of the soil)
TRANS	Text	55	Transition between horizons or soil layers (topography and contrast)
RAIZES	Text	100	Presence of roots (quantity, type, size)

Table 6. Attributes list of the ANALYTICAL_REPORT class with the fields description.

Name Field	Type	Size	Description
ID_HORIZON	Text	3-5	Identification code of the soil horizon or layer in the physical and digital soil bank (sequential number of the sample plus the project code)
CALH+CASC	Text	5	Pebbles (>20mm) + Gravel (20-2mm) ($g.kg^{-1}$ of the fraction in the total sample)
TFSA	Text	5	Air-dry fine soil (<2mm) ($g.kg^{-1}$ of the fraction in the total sample)
AR	Text	6	Total sand fraction (2-0.05mm) ($g.kg^{-1}$)
SIL	Text	6	Total silt fraction (0.05-0.002mm) ($g.kg^{-1}$)
ARG	Text	6	Total clay fraction (<0.002mm) ($g.kg^{-1}$)
ARG_NAT	Text	5	Natural clay ($g.kg^{-1}$)
GFA	Text	5	Degree of clay flocculation (%)
SIL/ARG	Text	4	Silt/clay ratio (dimensionless)
DS	Text	4	Soil density (apparent) ($g.dm^{-3}$)
DP	Text	4	Particle density ($g.dm^{3}$)
PT	Text	4	Total soil porosity ($m^3.m^{-3}$)
pH_H2O	Text	4	pH in water (1:2.5)
pH_KCl	Text	4	pH in salt (potassium chloride) (1:2.5)
C_ORG	Text	5	Organic carbon content ($g.kg^{-1}$) or TOC ($g.kg^{-1}$)
MAT_ORG	Text	5	Organic matter content (%)
Ca+Mg	Text	5	Exchangeable calcium+magnesium content ($cmol_c.kg^{-1}$)
Ca	Text	5	Exchangeable calcium ($cmol_c.kg^{-1}$)
Mg	Text	5	Exchangeable magnesium ($cmol_c.kg^{-1}$)
K	Text	5	Exchangeable potassium ($cmol_c.kg^{-1}$)
In the	Text	5	Exchangeable sodium ($cmol_c.kg^{-1}$)
S	Text	5	S value (sum of bases) ($cmol_c.kg^{-1}$)
Al	Text	5	Exchangeable aluminum ($cmol_c.kg^{-1}$)
H+Al	Text	5	Potential acidity ($cmol_c.kg^{-1}$)
CTC_EFT	Text	5	Effective cation exchange capacity ($cmol_c.kg^{-1}$)
T	Text	5	T value (CTC at pH7) ($cmol_c.kg^{-1}$)
V	Text	4	V value (base saturation) (%)
m	Text	4	Aluminum saturation (%)
In%	Text	4	Sodium saturation (%)
SiO2	Text	5	Silica in sulfuric extract ($g.kg^{-1}$)
Al2O3	Text	5	Aluminum in sulfuric extract ($g.kg^{-1}$)
Fe2O3	Text	5	Iron in sulfuric extract ($g.kg^{-1}$)
Ki	Text	4	Ki index (dimensionless)
Kr	Text	4	Kr index (dimensionless)
CaCO3	Text	5	Calcium carbonate equivalent ($g.kg^{-1}$)

The analysis of the proposed model demonstrates that this simplified approach for a specific application can provide a conceptual framework for more complex and comprehensive systems, thereby reducing implementation challenges.

This study presented a conceptual data model in which classes and relationships, through logical design, enabled the physical structuring of a soil database, supporting information management in Pedology.

The proposed model enables better organization and systematization of input data, representing a collection of readily accessible and useful information about soils in Santa Catarina state.

REFERENCES

ASSAD ED & SANO EE. 1998. Sistema de informações geográficas: aplicações na agricultura. Brasília: Embrapa.

BAUMGARDNER MF. 1999. Soil databases. In: SUMNER ME. (Ed.). Handbook of soil science. Boca Raton: CRC Press. p.H1-H4.

BOOCH G et al. 1998. The unified modeling language user guide. Massachusetts: Addison Wesley Longman.

CHAGAS CS et al. 2004. Estrutura e organização do sistema de informações georreferenciadas de solos do Brasil (SIGSOLOS – VERSÃO 1.0). Revista Brasileira de Ciência do Solo 28: 865-876.

CEN. 1996. Geographic information – data description – conceptual schema language. Brussels: EUROPEAN COMMITTEE FOR STANDARDIZATION. (Relatório).

COOPER M et al. 2005. A national soil profile database for Brazil available to international scientists. Soil Science Society

of America Journal 69: 649-652.
ELMASRI R & NAVATHE SB. 2017. Fundamentals of Database Systems. 7.ed. Harlow: Pearson.
BHERING SB et al. 1998. Base de informações georreferenciada de solos: metodologia e guia básico do aplicativo SigSolos, versão 1.0. Rio de Janeiro: EMBRAPA – CNPS. 112p. (Boletim de pesquisa 11).
EMBRAPA. 1998. Base de informações georreferenciada de solos: metodologia e guia básico do aplicativo SigSolos, versão 1.0. Rio de Janeiro: CNPS. (CD-ROM - Boletim de pesquisa 11).
FAO. 1996. The digitized soil map of the world including derived soil properties. Rome: FAO. (CD-Rom).
FURLAN JD. 1998. Modelagem de Objetos através da UML. Análise e Desenho Orientados a Objeto. The Unified Modeling Language. São Paulo: Makron Books.
LISBOA FILHO J. 2000. Projeto conceitual de banco de dados geográficos através da reutilização de esquemas, utilizando padrões de análise e um framework conceitual. Tese (Doutorado em Ciência da Computação) Porto Alegre: UFRGS. 212p.
LISBOA FILHO J et al. 2000. Modelagem conceitual de banco de dados geográficos: o estudo de caso do projeto PADCT/CIAMB. In: Carvão e Meio Ambiente. Porto Alegre: UFRGS. p.440-458.
MACDONALD KB & KLOOSTERMAN B. 1984. The canadian soil information system -CANSIS: general users manual. Ottawa Land Resource Research Institute, Research Branch, Agriculture Canada. 56p. (Manual Técnico).
NETO SLR. 2000. Um modelo conceitual de sistema de apoio à decisão espacial para gestão de desastres por inundações. Tese (Doutorado em Engenharia de Transportes). São Paulo: USP. 208p.
NETO SLR et al. 2006. Modelagem conceitual de geodados com técnica orientada a objetos para gestão de recursos hídricos. Revista Brasileira de Recursos Hídricos 11: 235-244.
RUMBAUGH J et al. 1991. Object-oriented modeling and design. New Jersey: Prentice-Hall.
SOIL SURVEY STAFF. 1991. National soil information system (NASIS): soil interpretation and information dissemination sub-system. Draft requirements statement. Lincoln: USDA, Natural Resources Conservation Service, National Soil Survey Center. 67p.
VAN ENGELEN VWP & WEN TT. 1995. Global and national soils and terrain digital databases (SOTER): procedures manual. Wageningen: UNEP-ISSS-ISRIC-FAO. 129p. (Manual Técnico).