Geo-DBMS harmonization of subsurface three-dimensional geo-information related to infrastructure

Research idea paper

Ludvig Emgård

Table of Contents

1. Introduction

This paper contains a suggestion for research in the area of subsurface 3D geometry. Interviews performed within phase 1 in the RGI-029 part of the current GIMCIW project (Tegtmeier, 2006) have clearly shown that harmonization of data created in different types of software is a major problem in handling geo-information in large civil engineering infrastructure works. However, some of the problems addressed in the interviews will not be treated in the RGI-029 project. Centralized management of information throughout the lifecycle of an infrastructure project is a problem that is mentioned, but not in the scope of the evolving project. The proposed research addresses the challenges of modelling geometries in 3D geo-information management and exchange where natural conditions and construction meet. By enabling the possibility to store, analyze and visualize subsurface 3D data in a spatial DBMS, experts could have access to a tool where it is possible to co-represent and interact with geometries from all the fields that are involved in construction under ground.

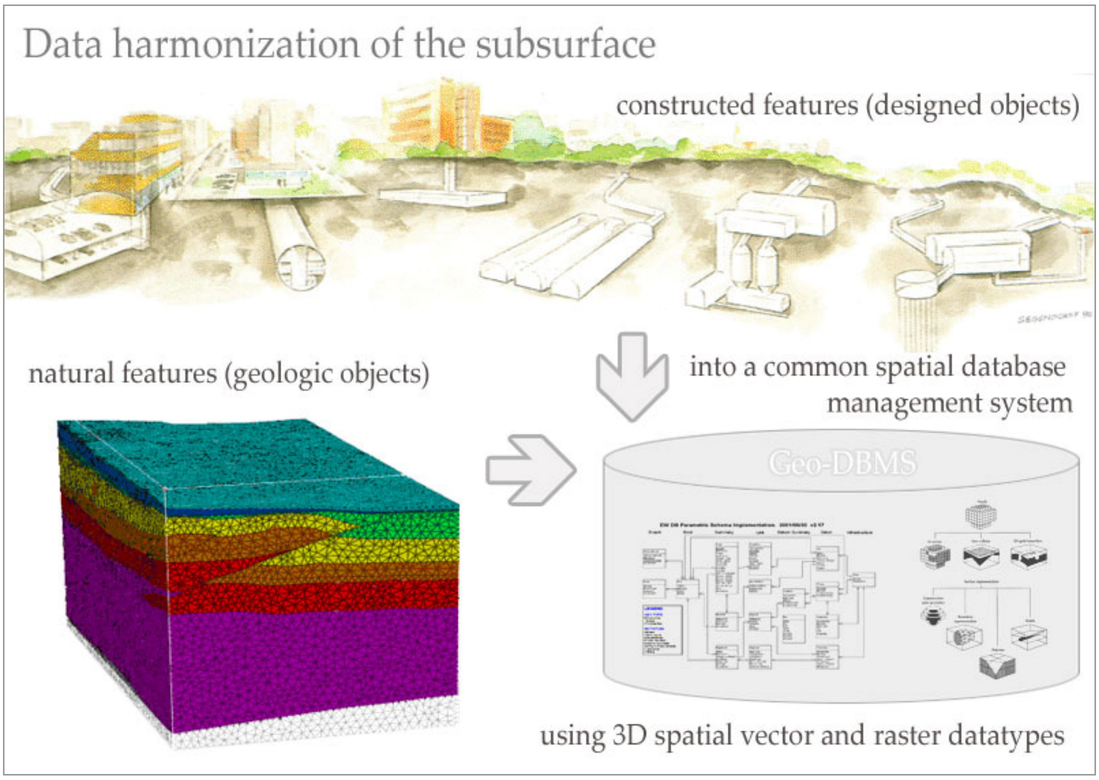

Figure 1.1: Data harmonization of the subsurface (Bergsprängarkommittén & subsurface real world representation in GOCAD).

By strengthening the knowledge about DBMS data types, that can most effectively be used to represent subsurface features in 3D, researchers can more rapidly develop common frameworks and standards to harmonise geo-information created in different software platforms.

The International Tunnelling Association (ITA), consisting of tunnelling associations from 53 nationalities, is promoting advances in planning, design, construction, maintenance and safety of tunnels and underground space (ITA, 2006). In 2000, working group no. 4 of the ITA proposed that state governments and local authorities should collect and collate information on underground geological conditions and existing underground structures, and interpret this information for use in short- and long-range urban planning. The intention of the idea was to integrate discrete databases and modelling technologies so that user-friendly, three-dimensional underground 3D maps of the uppermost 30-60 m beneath an entire urban area can become a reality. Realization of the idea would involve database development, data interchange and identification of minimum data sets for modelling (ITA, 2000).

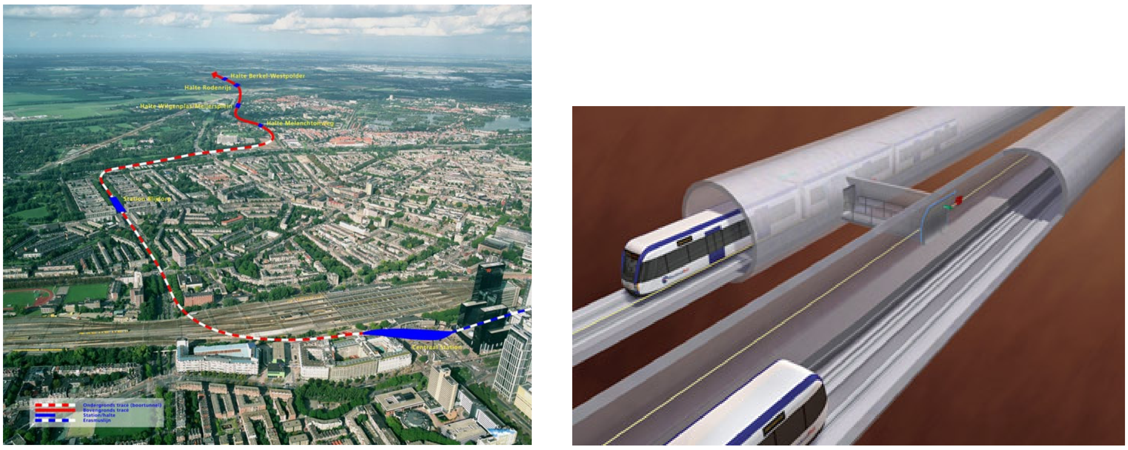

Figure 1.2: A large tunnelling project, The Rotterdam randstad line

There are several features to be represented underground. Subsurface constructions (design features) can be divided into following usage areas: passenger transport, car parking, vehicular goods transport, non-vechicular transport such as cabling, piping etc, goods storage, storage of oil, gas and chemicals, storage of hazardous waste, residential, business and services, small-scale manufacturing, technical research, retailing, entertainment facilities, culture, indoor sport and recreation (ITA, 2000).

2. Background

There are several problems with current methods for storage and interpretation of subsurface data. In this chapter, special attention is given to problems concerning representation, integration and standards within underground spatial data management.

2.1. Problems with existing representations

Features underground can be separated into two groups 1) existing natural and man-made formations (traditionally represented in GIS) and 2) engineering structures (traditionally designed in CAD) (Orlic, 1997; Oosterom et al., 2006).

Today, natural subsurface features are traditionally represented in two dimensions in geologic maps with a specific symbology, for example the Swedish SGU standard, NADM and NATMAP standards. The maps are managed in a GIS system consisting of a user application environment and a database management system (DBMS) (Apel, 2004). Data is collected by point observations and the product of data interpretation is a 2D geologic map. A problem is that the symbology used in the map is difficult to interpret for a person that is not familiar with the specific symbology of geology. The interpretation requires that the interpreter is able to combine a large number of map symbols and imagine how the geologic features are situated in 3D (Orlic 1997; Brunzell 2006).

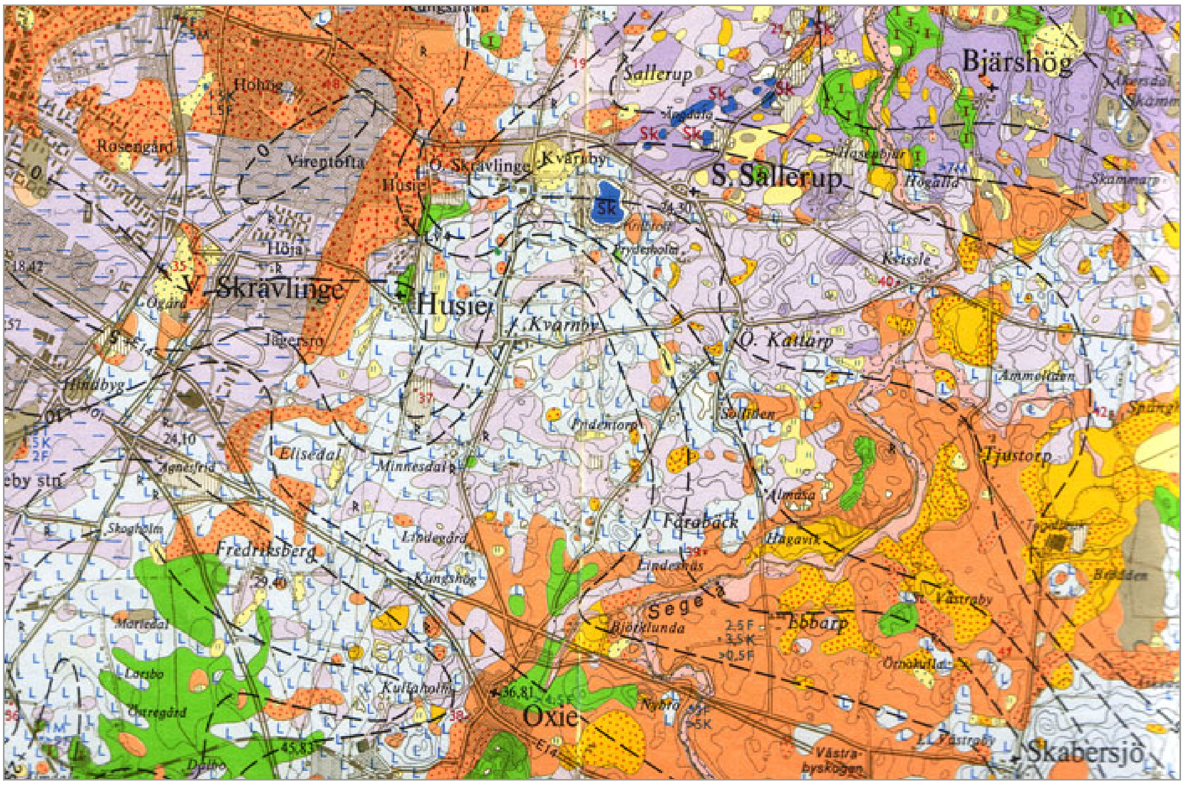

Figure 2.1: Geologic map of Malmö surroundings, SGU Standard

Another, more practical problem is that map data is often scattered in several map documents on paper or in data files. Because of the insecure reliability in interpretation of the geological map and the effort to gather scattered information of previous observations, the decision is often to acquire new observation data of the area for each project. This is inefficient since previous acquired data is not completely reused (Brunzell 2006).

Also design objects (engineering structures) are mostly stored in 2D representations. Even though advanced 3D construction software is available on the market 3D physical objects are still designed in 2D with the aid of linear profiles and cross sections. One reason for this is that contractors and builders are accustomed to 2D drawings and it would require more skills to create drawings in 3D (van Oosterom et. al, 2006).

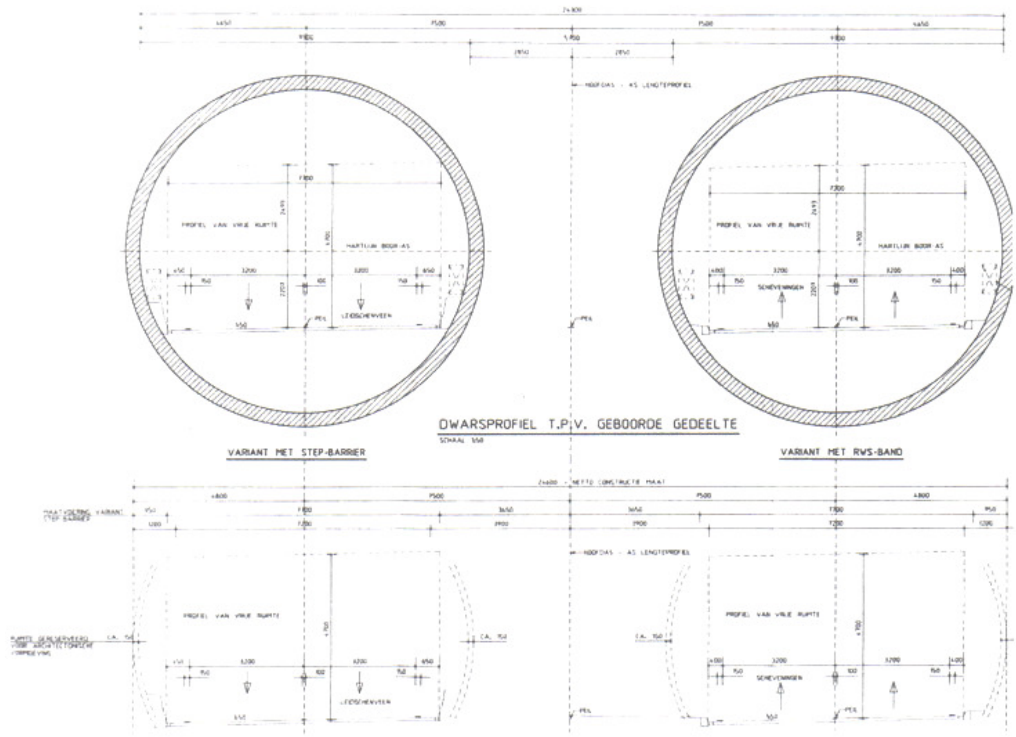

Figure 2.2: CAD model designed for the Hubertus tunnel in The Hague, Courtesy of Bouwdienst, Rijkswaterstaat. (van Oosterom et. al., 2006).

Whenever 3D data is created in an infrastructure project it suffers from many problems (van Oosterom et. al, 2006). For example, the CAD models are often primarily created for construction purposes and created with a too large amount of detail (also including non-geo data e.g. measures such as length and diameter) for storage purposes.

2.2. Problems with data integration

Because of the previous mentioned interpretation problem of 2D maps describing the subsurface, software packages for 3D interpolation, analysis and visualization have been developed within geology, mining and infrastructure. Examples of software packages are Gocad, Petrel (Schlumberger IS, Houston/TX), EarthVision (Dynamic Graphics, Alameda/CA), MVS (Mining Visualisation System, C-tech), Microstation (Bentley Systems) and NovaPoint (Vianova Systems). These software packages are used for different purposes and uses individual formats based on internal specific data models for storage of data which makes it impossible to directly access all data created in one system from another.

A problem experienced in subsurface infrastructure projects is that the results of data analysis from two software platforms can not be completely used and combined in one of the two platforms. For example, interpolated raster data from the water management calculation system MikeShe (DHI Software) can not be used in combination with interpolated boundary representations of geologic layers from the mining visualization system MVS (C-tech Systems). An overlay operation of the two datasets would be valuable, but is not yet available in any of the two systems (Brunzell, 2006).

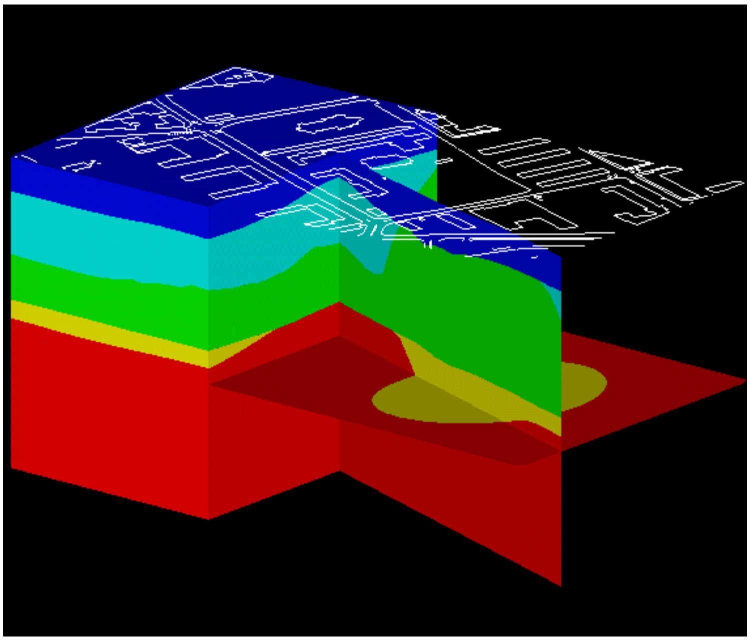

Figure 2.3: Visualization of Malmö created with the MVS System (Holmqvist, 2004)

Today, the method of transferring data between two systems is restricted to exchange by commonly used file formats. Even though one specific system data representation contains plenty of valuable and structured information, only parts of this information can be transferred via one of these formats. For example in the MVS system, the geometry of an externally created body (vector represented) can be imported and used as a boundary in an interpolation. However, the geologic property of such an object can not be included and integrated in the interpolation process (Brunzell 2006). The size of the datasets generated by interpolation is also causing problems because of limits in file size when exporting and importing between systems (Holmqvist, 2006). In addition, another difficulty with file based storage within geology is the lack of consistent multi-user access, and low data safety in comparison with storage in database management systems (Apel, 2004).

The problems described could partially be solved by spatial DBMS storage of the datasets where the systems have the possibility to acquire and store more intelligent data created in the internal system. This kind of method has been widely discussed but still no integrated solution exists. Currently, complex design objects from the CAD domain are usually stored in files. A reason for this is that not until recently, spatial DBMSs have had the possibility to represent features with more than point, lines and polygons in 2D (Breunig and Zlatanova, 2006). Because of the recent arrival of 3D data types, the integration and harmonization of existing data representations using the new data types has not yet been fully investigated.

2.3. Problems with existing data models and standards

A number of standards have been developed for geometric and semantic descriptions for natural features as well as design features. Geologic data is as usually described with national 2D geologic map data models such as the North American Geologic Map Data Model (NADM) or the Swedish SGU equivalent from Swedish Geological Survey. The NADM model is being continuously developed to provide a conceptual basis for the design of various schemas for databases and data interchange formats for geological observation data (NADM, 2006). The NADM 43a is the currently most complete, standardized conceptual data model for geological data. Based on the conceptual observation data model from the NADM 43a, Apel (2004) have created an extended observation point data model in UML where elements have been extended, restricted or substituted to fulfil the requirements of special user groups and tasks

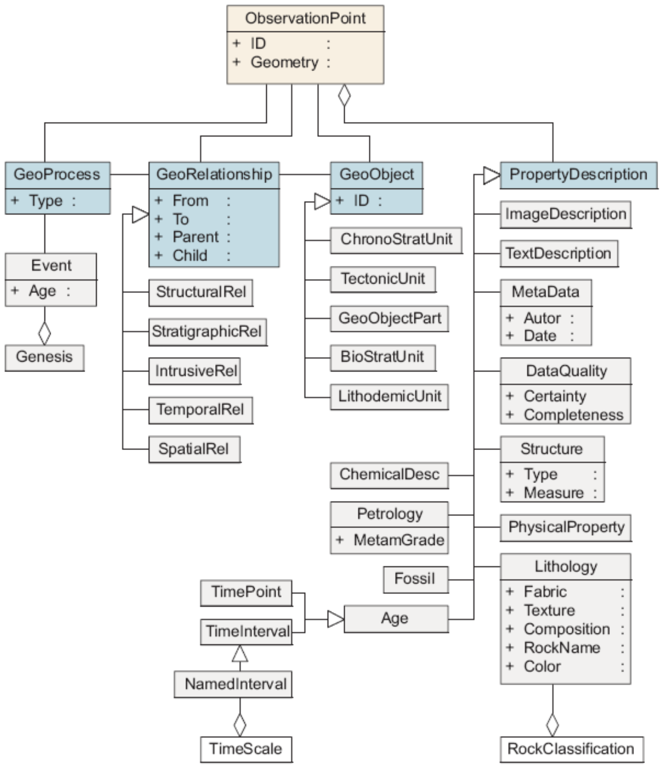

Figure 2.4: Apels' observation point model in UML (Apel, 2004)

In addition to the observation point model, Apel has also developed a new Gocad- and GML-conform spatial geological data model, synthesized by adding the concept of observation points to the BRep-based data model of the GoCad software.

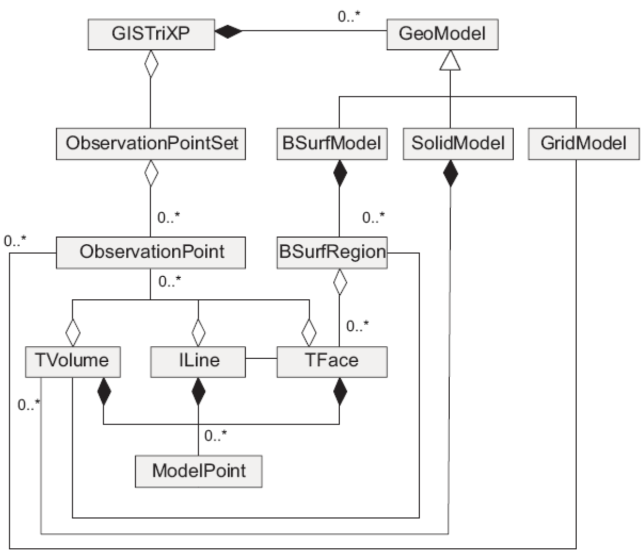

Figure 2.5: UML diagram of Apels' geological data model

The geomodel is created in order to fulfil the following criteria:

- representation of all 3d geomodels and observational data which are related to one geological situation

- conformable with the GML specification for geospatial data exchange using XML format

- compatible with 3d geomodeling software (GoCad).

An international initiative for harmonization of the national geologic models is under progress by the Commission for the Management and Application of Geoscience Information (CGI). The efforts of CGI are resulting in the GeoSciML (GeoScience Markup Language) that provides a data model and format capable of supporting transfer of geology data from multiple environments (CGI, 2006). Currently, a proposal for further development within the project is to deliver GeoSciML in a format that allows incorporation into GeoModeller (Gocad) which would show the use of GeoSciML as a machine to machine data exchange format. Another similar project within the mining industry is the XMML (eXploration and Mining Markup Language) project, initiated with support from participants in the mineral industry (service providers and mining companies) (XMML, 2006).

For design objects, at least two standards describing geometry is essential to be mentioned; Simple feature specification ISO 19170 from the Open Geospatial Consortium and the IFC standard, ISO/PAS 16739 created by the building industry. The ISO 19107:2003 specifies conceptual schemas for describing the spatial characteristics of geographic features, and a set of spatial operations consistent with these schemas. It treats vector geometry and topology in up to three dimensions (ISO 19171, 2003). The schema is used in the XML based standardized interchange format GML3 (Geography Markup Language) which is also a foundation for the proposed standard extension CityGML (Gröger et al. 2005) with an object model describing urban environments in three dimensions. The CityGML object model does currently not include subsurface features (Kolbe, 2006).

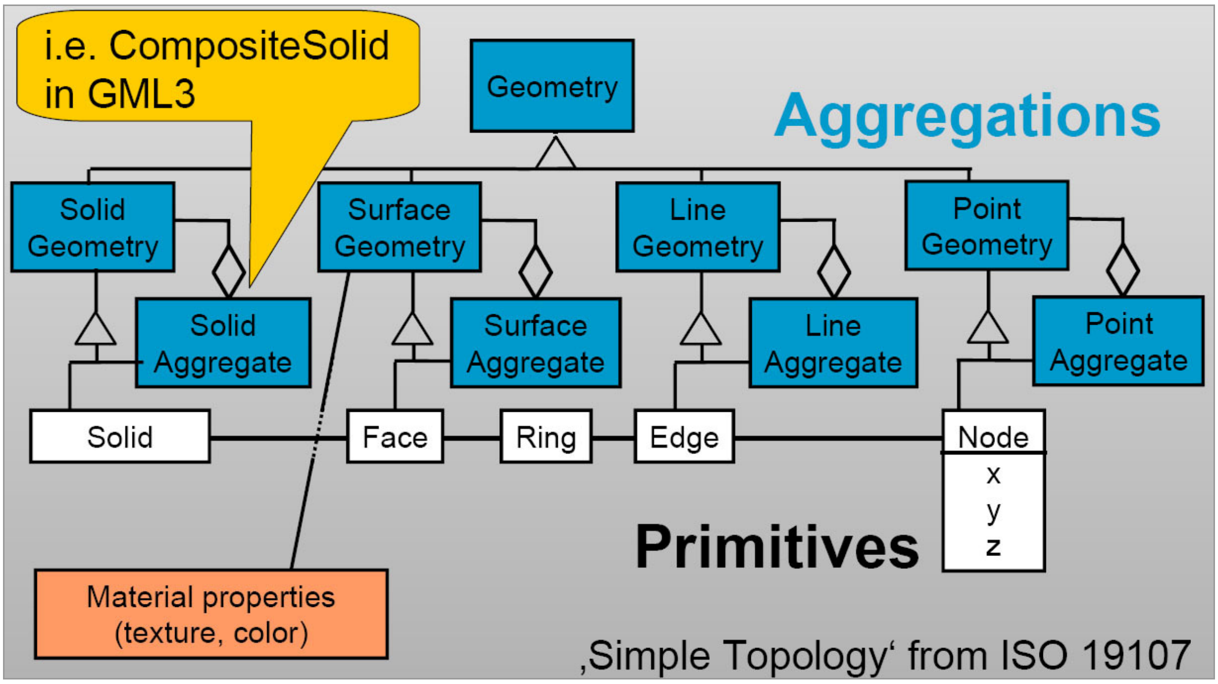

Figure 2.6: ISO 19170 used in CityGML (Gröger et. al, 2006).

The IFCs (Industry Foundation Classes) represents a data model structure for sharing construction and facility management data across various applications used in the building domain. IFCs have been endorsed by the International Standards Organisation as a Publicly Available Specification (PAS) under the ISO label ISO/PAS 16739. The standard is intended to improve communication, productivity, delivery time, cost, and quality throughout a whole building life cycle. Since the standard is developed from the building industry (CAD domain), some geometric objects of IFC (e.g. freeform curves) are more complex than the simple feature objects (point, lines, polygons and volumes) (IFC, 2006).

The standards and data models mentioned above supplies frameworks for storage of geologic observation data and vector representations of designed objects above ground. Anyhow, none of the models is at present sufficient for describing both design objects and natural objects underground expressed by three dimensional geometries.

3. Current methods to store 3D entities in a spatial DBMS

As described, a number of different geometry models are used to describe the reality, including both design objects and natural features. In general, there is a large difference of how we represent the two types of features. Design objects are well defined using geometric primitives or complex curves and surfaces while the geo-scientific objects are revealed by limited samples, or by indicative data that is highly irregular and complex with many more parameters (Lattuada 2006). Surface representations (i.e. grids, parametric surfaces, TINs, etc.) are suitable for describing geometric characteristics of objects by surface entities (e.g. Harbaugh and Merriam, 1968; Muller, 1988; Fried and Leonard, 1990), and volume representations (i.e. tetrahedral, hexahedral, prismatic, etc.) are suitable to characterize an object in terms of its internal properties, which can vary from one element to the next or from one element node to the next node (e.g. Requicha, 1980; Meier, 1986; Bak and Mill, 1989; Jones, 1989). Subsurface objects often require features of both, surface and volume, representations (Lattuada 2006). Today, some of the surface representations can be described and stored in 3D in a spatial DBMS using the multipolygon approach where a number of polygons are grouped into a body element and visualized as one object. Still there are some drawbacks with this approach that stresses research on "true" 3D data types. For example, a disadvantage is the redundant storage of coordinates. Moreover, the representation can not be recognized by a database as a volumetric object, neither be validated and the objects can not be indexed as 3D volumetric objects (Zlatanova & Breuning, 2006).

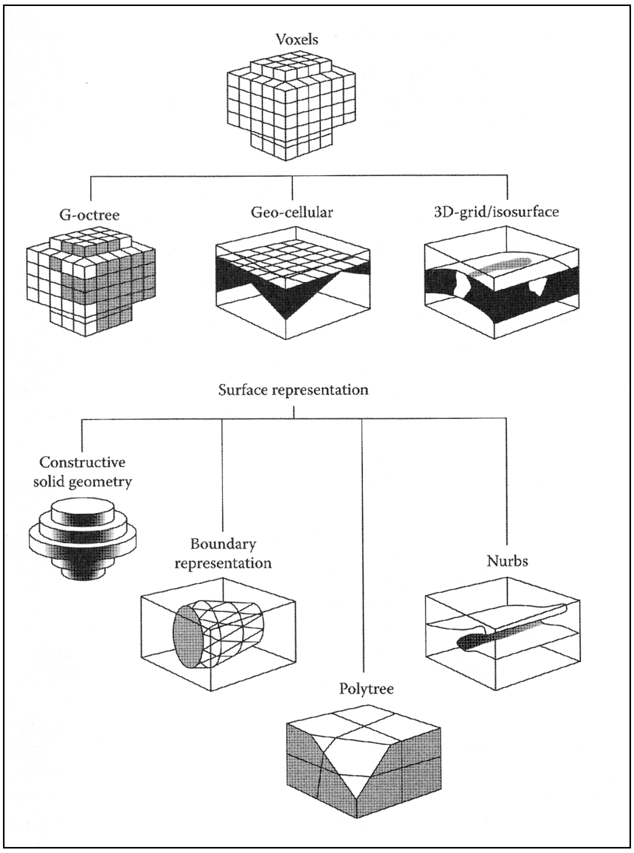

Figure 3.1: 3D spatial representations (Lattuada, 2006)

3.1. Raster representations

The real world objects are as mentioned often stored in X,Y observation points containing samples at specific Z-values. From this model a 3D raster (voxel) representation can be created by an interpolation (Bruzell 2006). Each voxel in the 3D raster can be described in the database with 3D point object. A collection of points represents a continuous volume where one or many attributes attached to each point is describing the material at this specific point. The resolution of the raster is given by the distance between two adjacent points. This representation is suitable for illustration of fuzzy objects underground, for example soil or rocks where many types of material is mixed. Anyhow, some geologic features (for example a lens shaped soil object or a rock) are distinct and may be more suitable to describe with its boundary (Brunzell, 2006).

3.2. Vector representations

Design objects are usually created using a vector boundary representation (B-reps). A closed B-rep (manifold) object can also be considered as a solid or a volume. Recently, at the Section of GIS Technology at TU Delft, two types of solid vector objects have succeeded to be stored in a database; tetrahedron (Penninga 2006), polyhedron (Arens et al 2005). In the Full 3D TEN model approach by Penninga (2006) the entire space is represented by non overlapping volumes described by TENs. A boundary representation containing point, lines and polygons can be derived from the volumetric model. Research has also been carried out in which a 3D polyhedron primitive was implemented in an Oracle Spatial DBMS (Arens et. al 2005).

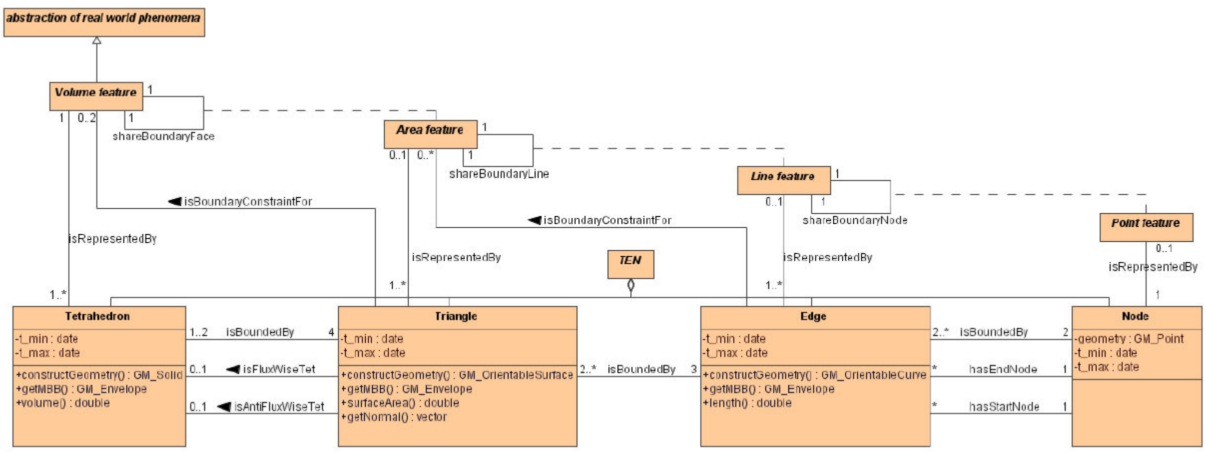

Figure 3.2: Full 3D TEN Model (Penninga, 2006)

3.3. Freeform shapes

Since also freeform shapes (parametrically stored curves and surfaces) have been used in geological modelling for representing underground surfaces and bodies it is interesting to further investigate how this kind of representation can be integrated with features that are represented by discrete b-reps and volumetric representations. At the moment, this data cannot be automatically converted to into the primitives that are available in a spatial DBMS (van Oosterom et. al, 2006). One recent achievement (concerning freeform curves) is that NURBS have recently succeeded to be stored in a spatial DBMS (Pu & Zlatanova 2006).

4. Proposed Nordic case studies, datasets and software

SWECO is the Nordic region's largest provider of consulting services in all fields of infrastructure with a staff of 3,800 employees (SWECO, 2006). Within the company group, a main problem is that specialists from the different domains use tools and data storage methods that are not compatible with each other. Research within harmonisation of subsurface geo-information at a DBMS level is of great importance to improve the interoperability between various numbers of software used by tunnel constructors, geologists, water recourse experts and architects.

In addition to Dutch case studies, two existing large infrastructure projects in the Öresund area of Scandinavia namely

- Malmö City Tunnel Project, planned to be finished 2011 and
- Helsingborg Tunnel Project, recently started

are proposed for case studies of the interoperability between involved software and methods of involved specialists.

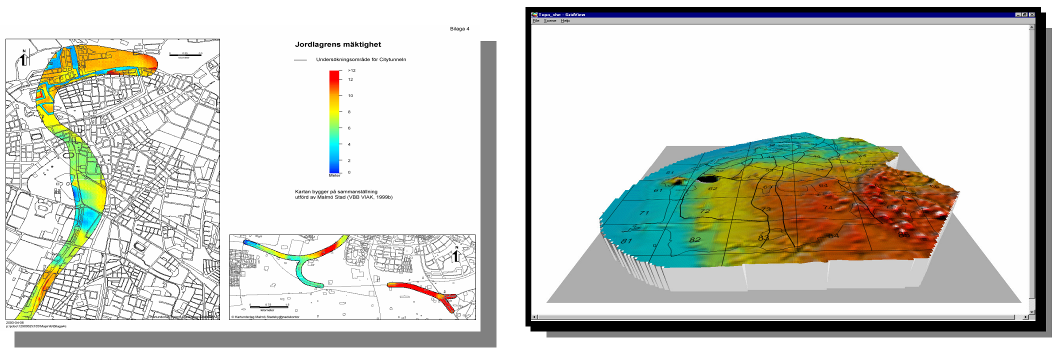

Figure 4.1: GeoAtlas System Interfaces (Brunzell, 2006)

The municipality of Malmö, Sweden has since the early nineties continuously collected geologic drilling data in a large dataset called GeoAtlas. This dataset is particularly interesting due to its size and amount of metadata connected to each observation. The GeoAtlas today contains 30 000 observation points and is widely used inside the infrastructure projects in Swedish parts of Öresund. A case study on this unique dataset is further proposed in the methodology section. Another case study on the GeoAtlas with focus on uncertainty is suitable to be performed in the RGI-029 of the GIMCIW project.

Brunzell (2006) from SWECO VIAK (geology and water resources) is responsible for 3D interpolation and visualisation of the GeoAtlas geology within the project area using the high-performance geology software MVS (Mining Visualisation System) from C-tech. According to Brunzell, a case study on DBMS storage of interpolated 3D data is of large interest and importance for the VIAK department at SWECO.

5. Research question

The main research question proposed for this study is:

- How can 3D vector and raster data representing subsurface geometry be harmonized to fit into a common DBMS model of underground features, concerning both real world objects and design objects?

The goal of the research is to:

- Develop a DBMS framework for harmonization of geometries describing subsurface features, on the basis of existing standards, requirements and methods within subsurface infrastructure projects.

5.1. Sub-division of research

The research can be further divided into following questions:

- How can geo-information related to an infrastructural project (real world objects for example geologic objects and design objects for example tunnels) be represented using DBMS datatypes (existing and recently developed at TU Delft).

- Is it possible to extend the geologic XML DBMS approach by Apel, (2004) or GeoSciML DBMS model with DBMS data types like for example, polyhedron and tetrahedron and extend the approach to also include raster geometries?

- Are existing and evolving data types sufficient for storage of both real world objects and design objects?

- How can the developed DBMS model be mapped to the two high-end geologic systems MVS and Gocad?

- Is it suitable to store freeform curves and surfaces (e.g. from IFC) in a DBMS as freeform data types or is it a better solution to map the freeform objects to vector and raster data types

- What kind of conceptual model and spatial data schema is suitable for the management of the various types of geo-information in a DBMS?

- Which functions representing spatial relationships between geological and design objects is appropriate to be available at DBMS level?

6. Research Methodology

Objects and data types as used in the various representations (i.e. subsurface, surface and design representations) and, hence, throughout the various disciplines (i.e. Geotechnology, GIS technology, Civil Engineering) will be compared in order to arrive at a good solution for the harmonization of geo-information.

The research on a new framework for DBMS storage of subsurface data will be carried out by investigations of existing standards within geology (for example NADM, GeoSciML & XMML) models and standards for geographic information like OGC specifications. Also the INSPIRE initiative will be followed and considered. Moreover, available literature on representations, data models (schemas), and data structures will be studied. Effort will be put on a broad investigation of existing and evolving DBMS data types.

Recent research with geology XML DBMS storage system approach by Marcus Apel at Freiberg Technical University will be examined. Investigations will be made in how to link the XML data types with recently developed data types at TU Delft (the polyhedron and tetrahedron).

The both high-end software packages, MVS and Gocad will be examined with respect to the possibility to store output data from interpolations from the both systems in a DBMS with the data types described above. Specialists at SWECO as well as, other company partners in the GIMCIW project will be visited for an insight in methods within subsurface 3D data software. A case study on the two Swedish infrastructure projects will be performed as well as a case study on how the GeoAtlas dataset can be represented with 3D data types.

By examination of coming generation of DBMS systems (Oracle 11), coming usage of new DBMS data types and DBMS system functions is going to be tested and investigated on the specified use cases.

7. Expected results

The expected results of the first phase of the project will be a description report of a determined framework for storing subsurface features in a DBMS.

8. Dissemination

The aim of securing the knowledge and performing the communication is to disseminate international experience, research status and the research results to consultants as well as to researchers and to help integrating results in the practical work of developing the geo-information management infrastructure in the GIMCIW project.

The results derived during the project will be presented at several conferences such as:

- 11th Scandinavian Research Conference on Geographical Information Science, 5-7 September 2007, As, Norway
- GIN Congress 2007, 21-23 November 2007, Amsterdam, The Netherlands
- 26th Urban Data Management Symposium, 2008
- GeoInfo, Bejing 2008

The work will also be presented to company partners within the GIMCIW project to spread the knowledge of database storage to specialists within geology and (sub)surface construction.

In workshops with database and software developers working at the geo-information infrastructure, the practical consequences of the research results for the further research phases will be determined.

9. References

Apel, M. (2004) A 3d geoscience information system framework. PhD thesis, Technischen Universit at Bergakademie Freiberg

Arens, C. , Stoter, J. & van Oosterom, P. (2005): Modelling 3D spatial objects in a geo-DBMS using a 3D primitive

Bak, R. & Mill, A. (1989): Three dimensional representation in a Geo-scientific Resource Management System for the mineral industry. In: Raper, J. (ed.): GIS - Three dimensional applications in geographic information systems. Taylor&Francis, London.

Breunig, M. & Zlatanova, S. (2006): 3D Geo-DBMS. In: Zlatanova, S. and Prosperi, D. (eds.): Large-scale 3D data integration - Challenges and Opportunities. Taylor&Francis, London, pp. 87-115.

Fried, C.C. & Leonard, J.E. (1990): 3D in depth: Petroleum models come in many flavours. Geobyte, 5, pp. 27-30.

Gröger, G., Kolbe, T, Czerwinski, A. (2006): OpenGIS CityGML Implementation Specification.

Harbaugh, J.W. & Merriam, D.F. (1968): Computer applications in stratigraphic analysis. Wiley&Sons, New York.

Jones, C.B. (1989): Data structures for three-dimensional spatial information systems in geology. International Journal of Geographical Information Systems, 3, pp. 15-31.

Lattuada, R. (2006): Three-dimensional representations and data structures in GIS and AEC. In: Zlatanova, S. and Prosperi, D. (eds.): Large-scale 3D data integration - Challenges and Opportunities. Taylor&Francis, London, pp. 57-86.

Meier, A. (1986): Applying relational database techniques to solid modelling. Computer Aided Design, 18, pp. 319-326.

Muller, J.P. (1988): Digital image processing in remote sensing. Taylor&Francis, London.

Oosterom, P.J.M. van; Stoter, J. & Jansen, E. (2006): Bridging the worlds of CAD and GIS. In:

Orlic, B. (1997): Predicting subsurface conditions for geotechnical modelling. PhD thesis, ITC, The Netherlands.

Penninga, F., van Osteroom, P. & Kazar, B. (2006). A Tetrahedronized Irregular Network Based DBMS approach for 3D Topographic Data Modeling.

Requicha, A. (1980): Representations for rigid solids: theory, methods, and systems. ACM Computing Surveys, 12, pp. 437-464.

Pu, S. & Zlatanova. ,S (2006) Integration of GIS and CAD at DBMS Level

Tegtmeier, W. (2006), Report on the activities within Work package 3 (phase 1) in RGI-029 within GIMCIW project

Zlatanova, S. and Prosperi, D. (eds.): Large-scale 3D data integration - Challenges and Opportunities. Taylor&Francis, London, pp. 9-36.

Zlatanova, S & Breuning, M. 3D Geo-DBMS In: Zlatanova, S. and Prosperi, D. (eds.): Large-scale 3D data integration - Challenges and Opportunities. Taylor&Francis, London, pp. 87-115.

Private communication

Kolbe, T. (2006), Technical University of Berlin, September 2006.

Brunzell (2006), SWECO Regional Office, Malmö, November 2006.

Holmqvist, J. (2006), SWECO Regional Office, Malmö, November 2006.

Internet references

CGI, Commission for Management and Application of Geoscience Information (2006) http://www.bgs.ac.uk/cgi_web/welcome.html

IFC model reference at International Alliance for Operability (2006) http://www.iai-international.org/Model/R2x3_final/index.htm

International Organization for Standardization, ISO 19171:2003 Specification (2006) http://www.iso.ch/iso/en/CatalogueDetailPage.CatalogueDetail?CSNUMBER=26012&ICS1=35

ITA, International Tunneling Association (2006) http://www.ita-aites.org/cms/115.html

NADM, The North American Geologic Map Data Model, (2006) http://www.nadm-geo.org/

SWECO Consultant Company Group, Official Website, (2006) www.sweco.se

XMML - the eXploration and Mining Markup Language (2006) http://www.opengis.net/xmml/