The problem of uncertainty integration and geo-information harmonization

W. Tegtmeier[1, 2], R. Hack[1], S. Zlatanova[2] & P.J.M. van Oosterom[2]
[1]*International Institute for Geo-Information Science and Earth Observation (ITC), Enschede, The Netherlands*

[2]*University of Technology Delft (TU Delft), Research Institute OTB, Delft, The Netherlands*

ABSTRACT: In infrastructural projects, communication between involved parties is difficult. This is, among other things, caused by lacking quality and uncertainty information concerning collected data and derived real world representations. Particularly in subsurface geotechnical representations uncertainties are high, since only sparse information is available for the interpretation. This leads to the introduction of "interpretational uncertainties" into the representation; that are, uncertainties introduced by the expert using own knowledge and experience for the data interpretation. That is what, in addition to the variety of data and information types, makes a harmonization of geo-information extremely difficult. This paper summarizes available methods and software packages as used by different professionals in infrastructural development for the representation of real world and design objects as well as for the management of geo-information. Furthermore, it emphasizes existing problems and gaps towards the harmonized handling of geo-information including uncertainty estimations; with focus on 'interpretation uncertainties' in subsurface representations.

1 INTRODUCTION

Around the world people are busy with the realization of infrastructural projects. Different tasks must be accomplished – infrastructural projects planned, designs changed, existing structures maintained or abandoned, etc. – all asking skills of a number of professionals (e.g. civil engineers, engineering geologists, GIS technologists, etc.). Thereby, various problems need to be tackled, often requiring the combination of a variety of data and knowledge as collected by the different professionals involved in the project.

Generally, these projects for the development of large civil-engineering constructions are characterized by a long lifetime (i.e. tens of years). This lifetime can be subdivided into six main stages that are namely:

1. Exploration
2. Planning
3. Design
4. Realization
5. Maintenance
6. Abandon

Large quantities of geo-information (e.g. GIS-, CAD-, and other data sets) are collected, generated and meant to be (re-) used in this lifecycle and the main problem, as known today, is the difficulty regarding data harmonization; that is the process by which different parties adopt a common (ideally standardized) way of working with geo-information in infrastructural development. The problems regarding this process are, among other things, caused by uncertainties concerning data as well as real world representations. Since large parts of the collected data and retrieved real world representations are not equipped with individual information about data origin, collection method, quality and possible uncertainties, the communication between the different professionals is difficult and, thus, the (re-) use of the information aggravated. This missing uncertainty information regarding various types of geo-information and real world representations, and also the use of different types of data structures and geo-information management systems are, thus, the main obstacles when trying to achieve data harmonization in large infrastructural projects. Consequently, the question is: How can geo-information be harmonized and equipped with uncertainty estimations?

2 THE QUALITY ASPECT OF GEO-INFORMATION

Since many people involved in infrastructural projects make use of available geo-information in order to take decisions, their work strongly relies on the quality of this information. This makes it an important aspect of geo-information and, to allow an effective use of collected data and information, it is, according to Dilo (2006), necessary to know its quality. But, before we can start and determine the quality of various types of geo-information, it is important to understand what the word "quality" actually implies.

Numerous definitions can be found throughout the literature. Harvey & Green (1993), for example, in their pioneering paper explored the nature and usage of quality in relation to higher education, where they conclude that quality is often referred to as a relative concept. In general, there are two senses in which quality is relative. First, quality is relative to the user of the term and the circumstances in which it is invoked. Second, is the 'benchmark' relativism of quality. In some views, quality is seen in terms of absolutes. In other views, quality is judged in terms of absolute thresholds that have to be exceeded to obtain a quality rating. Rather than try to define one notion of quality, Harvey & Green (1993) argued that they could be grouped into five discrete but interrelated ways of thinking about quality. Harvey (1995) provides the following overview of the five categories:

1. The exceptional view of quality sees quality as something special.
2. Quality as perfection sees quality as a consistent or flawless outcome.
3. Quality as fitness for purpose/use sees quality in terms of fulfilling a customer's requirements, needs or desires.
4. Quality as value for money sees quality in terms of return on investment.
5. Quality as transformation is a classic notion of quality that sees it in terms of change from one state to another.

The main definition of quality as used by many engineers and scientists and as defined in the ISO standards (e.g. ISO 9001:2000) is a version of quality as fitness for purpose, namely quality as satisfying needs: "Quality: The totality of features and characteristics of a product or service that bear on its ability to satisfy stated or implied needs. Not to be mistaken for 'degree of excellence' or 'fitness for use' that meet only part of the definition."

As described by Dilo (2006), several factors may in the end affect data quality and cause imperfections in the data. In general, deficiencies in data quality lead to different kinds of imperfection. According to Smets (1996), main aspects of imperfect data are *imprecision*, *inconsistency*, and *uncertainty*. Imprecision and inconsistency are properties of the data: either more than one or no world is compatible with the available information, respectively. If, on top of imprecision, we attach weights to the worlds to express our opinion about which might be the actual world, then we are confronted with uncertainty.

As it is a complex problem to consider together all factors that can affect the quality of geo-information and real world representations, it will foremost be focused on uncertainties in geo-information and real world representations of the subsurface, since for the interpretation of subsurface data usually limited information is available and the knowledge and experience of the interpreter has, thus, a significant influence on the outcome.

2.1 *Uncertainty as part of the quality aspect*

Regarding the fact that numerous different companies are involved in different phases of the process of infrastructural development, the quality and (possible) uncertainties in the geo-information or any of the real world representations must be properly defined and communicated to ensure an unhindered cooperation during infrastructural development. According to Foody & Atkinson (2002), it is, despite this apparent need to reduce uncertainty from an end-users and decision-makers perspective, however, still not possible to completely eliminate this factor of uncertainty.

In general terms, uncertainty can be described as a measure of the difference between estimation and reality (e.g. the difference of the thickness of soil layer as determined with CPTs compared to the situation in reality; expressed in percentage). This general description comes close to the definition as used in statistics, where uncertainty is defined as ‘the estimated amount or percentage by which an observed or calculated value may differ from the true value’.

Similar to the many aspects of imperfect data, different types of uncertainty (e.g. uncertainty to spatial prediction, uncertainty resulting from site investigations/ surveys/ measurements, uncertainty resulting from geological interpretation etc.) can be determined regarding, for example, the process of developing real world representation for infrastructural projects. These different types of uncertainties in geo-information can in the main be distinguished as spatial, temporal and semantic (i.e. thematic, such as classification and value) uncertainties.

Nowadays, several methods for the determination of uncertainty associated with spatial prediction (i.e. spatial uncertainties) as well as temporal uncertainties, as components of spatial data quality, are already available. In the particular case of spatial uncertainties, the choice of the uncertainty estimation technique is for one depending on the quality and quantity of available data, but also, as described by Zhang and Goodchild (2002), on the type of object, for which the uncertainty needs to be determined. Zhang and Goodchild (2002) distinguish two types of objects, for which uncertainties associated with spatial prediction might occur; namely uncertainties in continuous variables and uncertainties in categorical variables. In order to estimate uncertainties in continuous variables, geostatistical methods, such as kriging, indicator kriging or geostatistical simulations are most frequently used (e.g. Orlic 1997). Uncertainties in categorical variables, on the other hand, are mainly estimated using probability-based uncertainty calculations (e.g. Zhang and Goodchild 2002).

In addition to spatial and temporal uncertainties, there is, however, still the semantic uncertainty that might play an important role in infrastructural development, since many different companies are involved in one project and semantic uncertainty might hinder communication and also data and information exchange. Proper definitions of objects and their properties are of major importance in order to achieve a clear communication between all parties involved in the same project.

Regarding infrastructural development, there are two other sources of uncertainty that one should constantly be aware of throughout any characterization, although it seems to be difficult to define them. These additional sources of uncertainty are most prominent in case of subsurface real world representations, since only limited access and information is available for the interpretation of subsurface conditions. The first source of uncertainty concerns the potential for investigation errors (e.g. locational errors or measurement errors caused by wrongly calibrated machines). The second concerns the uncertainty inherent to any interpreted information (i.e. uncertainty introduced by the expert during the interpretation, depending on experience and prior knowledge of interpreter) (Houlding 1994).

According to Houlding (1994), there is generally little one can do about the potential for investigation error in sample and observation values without comprehensive research into each of the wide variety of investigative techniques in common use. There is, however, a general scarcity of information in this regard, aggravating further research and a possible solution of this certain problem. As indicated by Sides (1992) and Houlding (1994), the potential for error and uncertainty resulting from interpretation of geological features, on the other hand, is largely subjective. Up till now, there is no way of incorporating it into a computerized approach unless we are prepared to quantify it ourselves during the interpretation process. Unfortunately, this would essentially require a subjective estimate of the possible variation (in three dimensions) of each geological feature of concern.

Thus, various estimation techniques for the determination of spatial (and also temporal) uncertainties (e.g. geostatistical methods) have been studied carefully and are frequently applied in practice. But, in order to solve the problem of uncertainties in (sub-) surface representations completely, more research needs to be undertaken. What is certainly still missing in the study of uncertainty is, next to a solution concerning possible semantic uncertainties, the influence of the so-called ‘interpretation uncertainties’; that are the uncertainties introduced into the representation by the geology experts themselves. These interpretation uncertainties are, especially in geological (subsurface) representations, a dominant source of uncertainty, which is among other things caused by limitations in data quality and quantity and largely influenced by the knowledge and capability of the interpreter.

Ideally, the various types of geo-information and real world representations should, in the future, be equipped with information about possible uncertainties, that is best in the form of metadata. This ‘associated’ data documentation, or data about data, which is often referred to as metadata, plays an important role regarding the improvement of geo-information exchange and (re-) use and might, in the future, be a useful tool for the communication of quality and uncertainty information.

3 THE DEVELOPMENT OF VARIOUS TYPES OF GEO-INFORMATION

For the development of infrastructural projects an economic planning is of major importance. To achieve this, easy communication as well as data and information exchange needs to be guaranteed. With it, accurate data integration and the harmonization of geo-information is essential and it is of major importance to be able to integrate complex systems (see also Figure 1 and 2) that are consisting of 1) existing natural and man-made formations and 2) engineering structures (traditionally designed in CAD) (Orlic 1997, Oosterom et al. 2006).

In Figure 1 and 2, two examples of real world representations are shown. In these Figures various objects are represented; namely roads, bridges, buildings, street lamps, vegetation (i.e. grass, trees, etc.), etc. For the development of these representations, various experts (e.g. civil engineers, GIS-technologists, etc.) are using various types of software packages (e.g. CAD-, GIS-systems, etc.) in order to achieve a proper real world representation of the area of interest. In civil engineering, often CAD-based systems are used for the design of a bridge, tunnel or road. In geodesy, GIS-based systems are used to represent existing structures, such as buildings, trees, etc. Still today, the integration of the various representation types into one representation is rather difficult. This problem is, among other things, increased by the fact that, in most cases, the quality and (possible) uncertainties of each representation are unknown. As described earlier in this paper, these uncertainties can be caused by various factors, such as inaccuracies in data collection or interpretation, etc. During the process of data integration, for example, the use of varying reference or coordinate systems in the different representations might also lead to additional uncertainties. In such case, the position of a planned bridge, for instance, might be erroneous (e.g. the bridge in Figure 1 might not be placed at the right position). This could then lead to additional project costs for the verification of the correct bridge location and on basis of that even to additional site investigation or a re-design of the bridge.

One of the main problems in creating such a harmonized system capable of offering a wide variety of functionalities is the amount of different data types. A (common) final representation should ideally contain knowledge about reality, so we have to consider the different types of real world objects it must represent (Raper 1989).

Figure 1. Representation of man-made formations (by courtesy of Grontmij).

Figure 2: Representation of engineering structures (by courtesy of DHV).

3.1 *Techniques for the development of real world representations*

Regarding the diverse characteristics of different types of real world objects, we must consider applying different techniques for the representation of each object. In the following subsections, the various representation techniques and software packages as applied for the representation of subsurface, surface and design objects; that are the main types of real world objects as met in infrastructural development; will be presented.

3.1.1 *Subsurface objects*

Subsurface objects can have different dimensionality and can be represented by points, lines, surfaces and volumes (e.g. Orlic 1997). Computer representations of subsurface objects have generally been subdivided in two classes that are namely: 1) surface (= boundary) representations (see Fig. 3) and 2) volume representations (see Fig. 4).

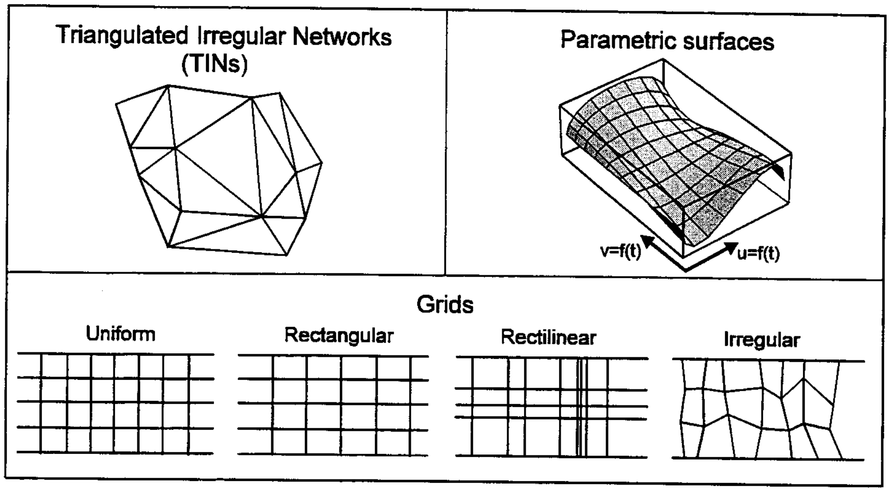


Figure 3. Surface representations (Orlic 1997).

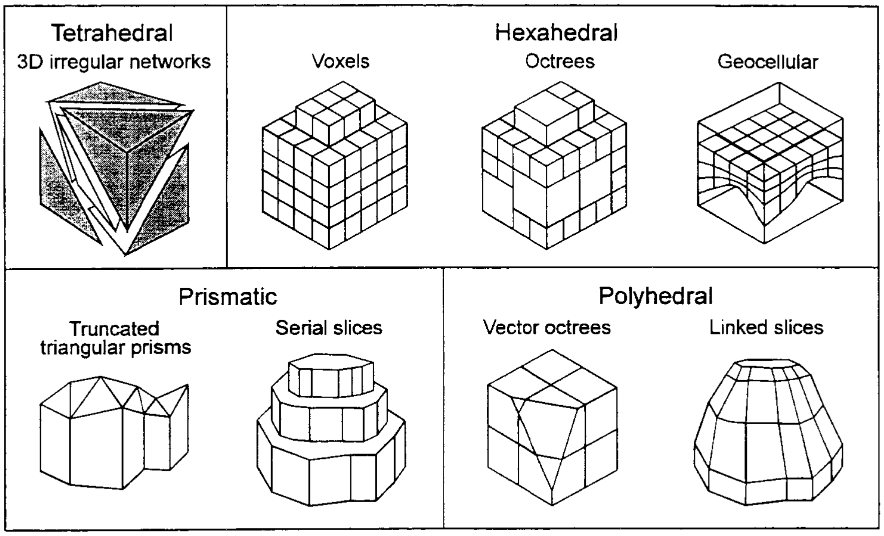


Figure 4. Volume representations (Orlic 1997).

Surface representations are suitable for describing geometric characteristics of objects by surface entities (i.e. assuming the described volume is homogeneous) (e.g. Harbaugh & Merriam 1968, Muller 1988, Fried & Leonard 1990), and volume representations 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. Meier 1986, Bak & Mill 1989, Jones 1989). Yet, subsurface objects often require features of both, surface and volume, representations, and most products today reflect these requirements by including elements of both types of representation techniques (e.g. Lattuada, 2006).

Regarding the robustness of surface and volume representations (see Figure 3 and 4) in relation to uncertainty, it is rather difficult to compare these two representation types. Considering the fact that only sparse information is available for the development of subsurface representations in general, it can be said that all of these representation types are rather sensitive towards the influence of uncertainties. It can be assumed that surface representations (see also Figure 3), on the one hand, are quite sensitive towards the influence of uncertainties, since they are used to describe the geometry of objects. The determination of rock and soil layer boundaries as well as the geometry of any other geological feature is difficult and usually largely dependent on the quality and quantity of data available, the experience of the interpreter and his knowledge about the specific geological setting. Volume representations (see also Figure 4), on the other hand, which are used to describe the internal properties of a geologic object, can also be considered sensitive towards the influence of various types of uncertainties. This is caused by the fact that for a good description of geological properties, a proper sample of the material must be taken from a representative location under appropriate conditions, etc. Furthermore, proper handling of the sample as well as the testing (i.e. laboratory) devices must be ensured. Considering these factor, it can be said that the correctness of both surface as well as volume representations depends on many factors and, thus, both representations can be considered rather sensitive with regard to uncertainties.

Usually, in engineering geological and geotechnical studies it is normal to make a 3D model of the distribution of the geotechnical properties of the subsurface. Such a model generally consists of a boundary model that gives the boundaries between the different defined geotechnical units and a property model for the distribution of the geotechnical properties within the geotechnical units. In principle, the model has to be 3D and able to represent changes of geotechnical properties over time, i.e. the model should be 4D with time as the fourth dimension (Hack et al. 2006).

In the past, dedicated computer programs for modeling of the subsurface were developed resulting in the fact that, at present, various software packages exist for the representation of real world objects in the subsurface. In addition to the commonly used GIS packages (e.g. ArcGIS, ArcView, etc.), GIS and GIS-like packages (e.g. Lynx, Rockworks, GOCAD, etc.), which are specifically developed for the applications in the field of geology and engineering geology, are frequently used in order to achieve reasonable representations of the real world situation in the subsurface. These software packages are not only used to produce two-dimensional representations such as maps or cross-sections, but also and even more increasingly for representations of the subsurface in three dimensions. Also, an increasing number of calculations and analyses are possible with the help of these software packages (e.g. basic statistics, layer thickness computations, etc.).

3.1.2 *Surface objects*

Surface objects are usually represented in 2D and 2.5D with simple primitives, such as point, line and polygon (i.e. in vector representations) or as segmented areas (i.e. in raster representations). In the third dimension real-world objects can be described by four elementary objects (i.e. point, line, surface and volume). As suggested by the ‘Open Geospatial Consortium OGC’ (Open Geospatial Consortium Inc. 2004), the simple primitives can be organized in geometric (simple features) or topological (complex features) data structures (see Fig. 5).

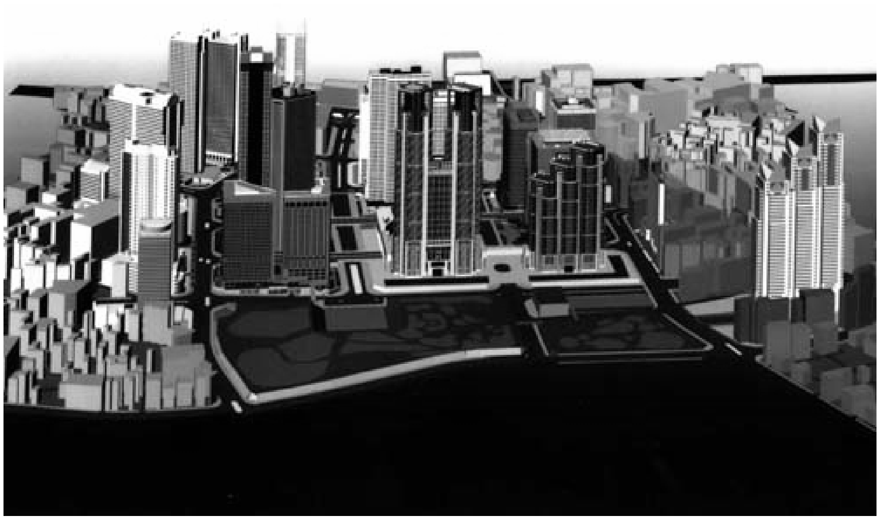


Figure 5. 3D representation of surface objects in GIS (http://shiba.iis.u-tokyo.ac.jp/pub/3d_forum/3d_forum.jpg).

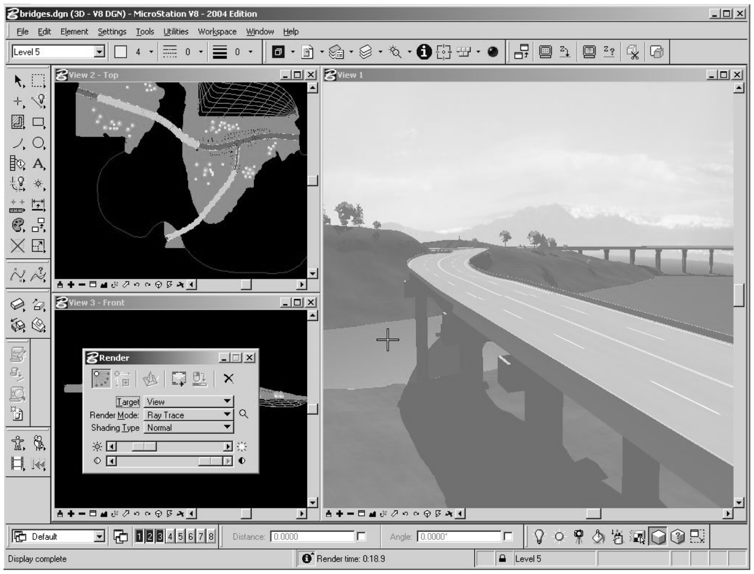
The topological data structures (i.e. Formal Data Structure FDS, Tetrahedral Network TEN, etc.), as reported in the literature (e.g. Molenaar 1989, Molenaar 1990, Pilouk 1996), can be subdivided in two main groups; that are, structures maintaining objects (object-oriented; i.e. relationships between objects must be derived) and structures maintaining relationships (topology-oriented; i.e. representation of objects must be derived). Frameworks (e.g. 9-intersection model, Egenhofer & Herring 1990) to formally describe relationships between different objects (independently of the data structure) are also available (see also van Oosterom et al. 1994, Zlatanova et al. 2004).

For the representation of real world objects on and around the earth's surface GIS systems are most widely used. Generally, the functions of a GIS are specified as follows (Raper & Maguire 1992): 1) data capture, 2) data structuring, 3) data manipulation, 4) data analysis, 5) data presentation & distribution. These systems are able to maintain information about spatial phenomena and provide means to analyze it and thus, gain knowledge of the surrounding world and are as such useful tools for the representation, management and analysis of existing structures on the earth's surface.

3.1.3 *Design objects*

Design objects can be represented by various methods. A number of representational forms for three-dimensional objects have been developed in computer-aided design. Some of these arise from applications and the data structure is determined wholly by the representational strategy. The factors that the representation generally tends to determine are 1) the data structure (i.e. the form of the processing algorithm and the design of fixed program hardware), 2) the cost of processing an object, 3) the final appearance of an object, and 4) the ease of editing the shape of an object. In computer graphics, the most popular method for representing an object is the polygon mesh (triangular) representation (i.e. objects are approximated by a net of planar polygonal facets). In addition, the 'bicubic parametric patches', that are freeform curves and surfaces (i.e. objects are represented exactly by nets of elements called patches), 'constructive solid geometry' (i.e. objects are represented exactly by a collection of elementary parametrical objects, such as spheres or boxes), and 'space subdivision' (i.e. objects are embedded in a space where points in the space are labeled according to object occupancy) techniques are frequently used for the representation of design objects. With it, polygon mesh and space subdivision are approximate representational forms, whereas the other two techniques are exact representational forms. Furthermore, with the polygon mesh and bicubic parametric patch methods boundary representations are derived, whereas the other two techniques deliver volume representations (e.g. Watts 1993).

For the representation of design objects, such as engineering structures, CAD oriented systems are most widely applied; that are typical computer graphics tools for 3D design, which are used, for example, for car, machinery, the construction industry, and architecture (see also Fig. 6).



Level 5

Target: View
Render Mode: Ray Trace
Shading Type: Normal

Distance: 0.0000
Angle: 0.0000°

Display complete
Render time: 0:18.9

Figure 6. 3D Bridge Design in MicroStation V8
(http://www.bentley.com/bentleywebsite/files/corporate/ydb/MicroStation-II.jpg).

CAD systems are focused on the geometric aspect of the object and its 3D visualization and are generally using geometric primitives to design, evaluate, edit, and construct various objects. Classical CAD methods are used to design interactively curves and surfaces (Lattuada 2006) and are generally focused on the geometric aspect of the object and its 3D visualization.

3.2 *Management of various types of geo-information*

Nowadays, the general understanding for GIS and CAD systems is changing. Besides analyses, GIS is becoming an integration of strong database management (i.e. ensuring data consistency and user control) and powerful editing and visualization environments (i.e. inheriting advanced computer graphics achievements); and also CAD vendors are trying to provide means to combine spatial as well as thematic data and organize topologically structured layers. Resulting from these developments in the field of spatial data management is, according to Zlatanova et al. (2002), also a changed role of Database Management Systems (DBMSs).

3.2.1 *Database Management Systems (DBMSs)*

A Database Management Systems (DBMS) can generally be described as a computer program (or more typically, a suite of them) designed to manage a database (i.e. a large set of structured data) and to run operations on any collection of compatible, and (in case of a relational DBMS) ideally normalized, data of a particular application or problem (i.e. as requested by the numerous clients). According to Date (1995), the functions a DBMS should offer are 1) data definition, 2) data manipulation, 3) data security and integrity, 4) data recovery, and 5) data dictionary.

At present, the so-called Geo-DBMSs, for example, can manage temporal, thematic and spatial data and they are providing spatial data types and functions that define the spatial functionality of this Geo-DBMS. A Geo-DBMS knows simple and composed data types, unfortunately, most of the spatial data types are still only two-dimensional; i.e. point, line, and polygon (Breunig and Zlatanova, 2006). In the near future, however, it is to be expected that also 3D data types will be supported inside the DBMSs (e.g. in the new version of Oracle Spatial; Oracle 11g). Furthermore, according to the Abstract Specifications of the OpenGIS Consortium (OGC 2006), a spatial object can generally be represented by two structures in the DBMS; that are namely, geometrical structure, i.e. simple feature, and topological structure, i.e. complex feature. Each spatial object is then maintained in the DBMS environment following the Implementation Specifications of the OpenGIS Consortium (e.g. OGC 1999).

Nowadays, DBMSs can already be integrated within GIS systems and important parts of data analyses are performed by the DBMS during the execution of database queries. In contrast to GIS systems, in the field of CAD, the use of DBMS functionality was considerably more restrictive (Breunig & Zlatanova 2006). However, an increasing number of CAD systems (e.g. MicroStation, GeoGraphics) have, at present, also developed extensions that make use of spatial data structures and functionality provided by Geo-DBMS. In that way, DBMSs can also be used as a bridge between GIS and CAD applications. But, in order to be able to provide a stronger management of objects from CAD and GIS, Geo-DBMSs have to extend their spatial support to accommodate design objects and real-world objects.

However, the question ‘who is responsible for the spatial analysis’ (i.e. front-end applications or spatial DBMS) is still open and even extensively discussed. Generally, it can be said that GIS functionalities that are not specific to a certain application belong in the DBMS and not in GIS (or CAD) front-ends (Zlatanova and Stoter, 2006). Nevertheless, it should not be forgotten that Geo-DBMSs, in the first place, are DBMSs; i.e. the location for data storage and management. Thus, the 3D functionality should not be completely taken away from GIS and CAD applications. 3D Geo-DBMSs should only provide the basic (simple) 3D functions (e.g. as computing volumes and finding neighbours as basis needed for querying) and application specific (i.e. as needed for complex analysis) should still be attributed to the GIS/CAD systems (Breunig and Zlatanova, 2006).

What should become clear is that in infrastructural development a number of varying types of geo-information and software packages are used in order to arrive at a proper real world representation sufficient for the purpose of infrastructural planning. A suitable integration and harmonization of these geo-information and representation (i.e. subsurface, surface and design) types, however, is still far from being possible. The various systems frequently used may well have features in common (e.g. all are mainly concerned with geometry); however, they also differ in many respects such as dimensions, storage, analysis, semantics, etc.

4 CRITICAL RESEARCH ISSUES AND DEVELOPMENTS

4.1 *The current situation*

In infrastructural development a number of various objects (i.e. subsurface, surface and design objects) need to be determined and represented in order to get thorough information about the situation at and around the construction site. Regarding the various representation techniques used for the illustration of the different real world objects, significant progress could be realized throughout the last years.

As regards the representation of subsurface geotechnical objects, various software packages exist, which were especially developed regarding the specific requirements met in geological and geotechnical engineering. The development and increasing use of computer based modeling for geo-engineering purposes, however, seemed to stall during the past decade. Full 3D programs are now sometimes used for modeling the subsurface for very large projects, but more often they are just 2.5D. In even more projects only a 2D program is used to make horizontal and vertical sections or no computer program is used at all and just old fashioned but trusted paper methods are applied for the preparation of subsurface models and sections. This is mainly caused by the fact that the quality and quantity of subsurface data in geo-engineering projects is often limited and, thus, the added value of using a digital 3D system not always obvious. Furthermore, not all required tasks can be executed in one combined program; but a number of sophisticated computer programs must be applied for the special tasks and, consequently, the geotechnical experts cannot be familiar with all existing packages.

Concerning the representation of surface objects, GIS systems are most frequently employed. These software packages can be used for the (re-) construction, management and analysis of existing (geographical) objects, about which only sparse and incomplete information is available. For a long time, these functions of the GIS systems were focussed on 2D. Nowadays, 3D GIS aim at the same functionality as 2D GIS, but in 3D space. Therefore, important developments in the field of 3D GIS are the improvement of 3D data collection techniques as well as developments concerning hardware, such as processors or memory and disk space devices, which have become more efficient in processing large data sets. Furthermore, elaborated tools to display and interact with 3D data are evolving. GIS software-tools have also made a significant movement towards 3D GIS and the major progress achieved in 3D GIS is on improving 3D visualization and animation. Nevertheless, 3D functionality, such as generating and handling (querying) 3D objects, 3D structuring or 3D manipulation and analyses, is still limited (or sometimes even lacking).

As to the representation of design objects, CAD systems are usually applied for the development of new constructions. In the past, CAD systems were solely focused on the geometric aspects of a certain objects as well as on the problem of 3D visualization. No topologic or attribute information were included in the representation of the object and the CAD systems were, thus, allowing hardly any analysis on the data. As these analytical capabilities of computer software are, however, of importance for most geoscientific disciplines, CAD systems for engineering applications are currently changing and, nowadays, tend to be called AEC systems (i.e. architecture, engineering and construction). Already now, many CAD systems (e.g. Bentley, AutoCAD) offer more and more extended tools to represent and organize real world objects. Vendors dealing with either spatial or thematic information attempt to achieve GIS functionality already for years and CAD vendors (e.g. AutoDesk, Bentley), nowadays, provide means to combine 2D and even 3D spatial data to thematic information and organize topologically structured layers.

In the main, it can be said that spatial objects of reality, nowadays, can well be visualized in the form of graphics in two- as well as three-dimensional representations. Unfortunately, the developments concerning data analyses were, up till now, less successful and they are still limited to 2D and 2.5D. In most geoscientific disciplines significant developments are, thus, taking place regarding the capabilities of the various software packages that are frequently applied. Unfortunately, these developments are essentially limited to specific scientific disciplines and not meant to improve a possible integration of different types of real world representations developed throughout infrastructural projects. A possible integration of the various types of real world representations is, thus, still missing.

The combined use of the various software packages often results in problems caused by non-compatible data structures or differences in scale and functional level. In order to facilitate the data exchange and communication between different parties involved and also to achieve an economic planning of infrastructural projects, a harmonization of the various types of geo-information must be approached and, ideally, a conceptual model for the semantics of data frequently used in infrastructural development build up, which would also lead to a possible decrease in semantic uncertainty. In addition to the problems caused by the combined use of the various software packages, missing information about the quality and possible uncertainties of data in the same way as representations intensifies the problems in infrastructural development.

In Figure 7 the various components (and their relationships) of importance in infrastructural planning are shown. This figure is meant to give an overview about the problems that have to be faced in most infrastructural projects.

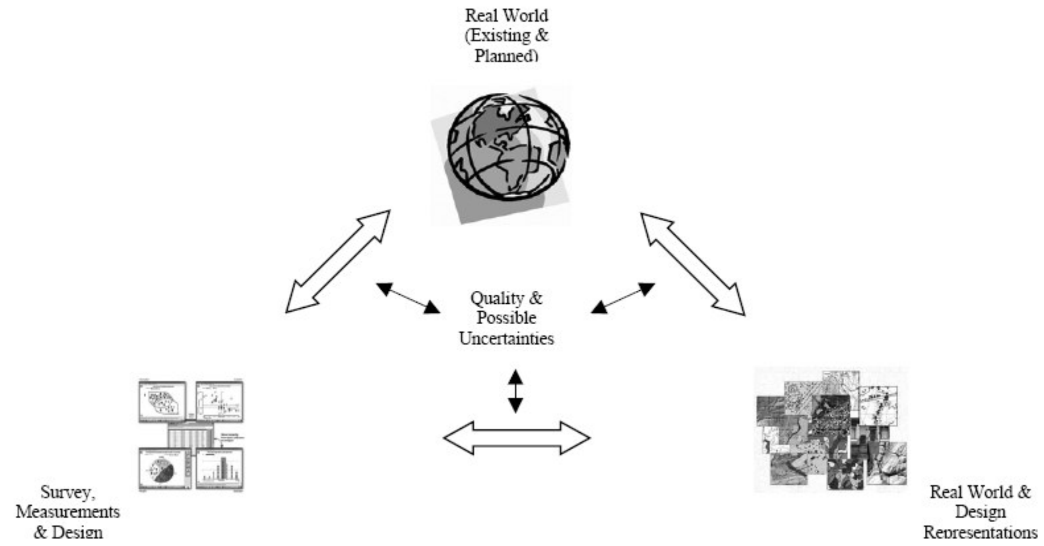


Figure 7. Problems (and their relationships) to be tackled in infrastructural development.

What should become clear is that during the development of infrastructural projects, large quantities of data and information are collected and used for the creation of various types of real world representations; that are mainly subsurface, surface and design representations. The harmonization of these various types of geo-information is still difficult (if not impossible) and, in addition, usually no information is attached to the geo-information giving an indication on the quality and possible uncertainties regarding data and representations, which is, however, of importance for future (re-) use of this information. This leads to communication problems between companies involved in the infrastructural project and, often, makes an economic planning of the project difficult (or even impossible). It is, thus, desirable to solve the problems of data harmonization and uncertainty determination in infrastructural development.

4.2 *Future Research*

Throughout this research, an approach will be made to solve a number of problems as mentioned earlier in this paper. On the subject of quality and uncertainty of data, information and representations, this research will be focused on the determination and communication of the so-called ‘interpretation uncertainties’; that are uncertainties introduced into the representation by the geology experts themselves. These interpretation uncertainties are, especially in geological (subsurface) representations, a dominant source of uncertainty, which is among other things caused by limitations in data quality and quantity and largely influenced by the knowledge and capability of the interpreter With regard to the harmonization issue, the focus will be put on the harmonization of geo-information, and there especially concerning the meaning of data (i.e. thematic semantics of data).

4.2.1 *Determination and communication of ‘interpretation uncertainties’*

As described earlier in this paper, uncertainties in data sets and real world representations in infrastructural development can be of various nature, namely spatial, temporal and semantic. In addition, uncertainties in geo-information can also be caused by investigation and interpretation errors. Regarding the fact that spatial as well as temporal uncertainties in geo-information are quite thoroughly investigated, the focus of this specific research will be put on the determination of the so-called ‘interpretation uncertainties’ in subsurface geotechnical representations.

This topic alone is still quite complex; however, a first step will be made towards the determination and communication of interpretation uncertainties in subsurface geotechnical representations. The goal within this research is to arrive at a description of the level of the interpretation uncertainties in a certain geotechnical representation on a scale of, for example, 1 to 5; with 1 a low level of interpretation uncertainty and high reliability of the subsurface representation and 5 vice versa. In this research, a weighting system will therefore be build up and applied in order to arrive at scalable values indicating the interpretation uncertainties to be expected in a certain geotechnical representation as well as their influence on the construction and maintenance of the infrastructural project. Aspects that will be taken into account in the weighting system are, for example, the quantity of the collected data, the quality of the collected data, the extent/size of the construction site, the expected impact of the civil construction on the geology (i.e. type/size/etc. of construction) and the experience of the geotechnical expert executing the interpretation (i.e. familiarity with geology around the construction site, number of representations made in this area, etc.). Each of these aspects will then be given a factor depending on the conditions met in a certain project. Additionally, the values are weighted depending on their influence on the final interpretation uncertainty to be expected in this geotechnical representation. In the end, this uncertainty information will be included in the metadata of the geotechnical representation and, ideally, be equipped with additional information regarding the implications of this interpretation on the construction of the infrastructural project.

To achieve the scaling of the possible interpretation uncertainties and their influence on the quality of geotechnical representations, (engineering) companies and agencies involved in infrastructural projects will be questioned about their use of uncertainty information in subsurface real world representations and case studies will be will be analyzed in order to acquire information about the influence of the expert knowledge on the quality of a real world representation.

4.2.2 *Harmonization of geo-information*

Regarding the harmonization aspect of geo-information, it will in this research be focused on the ‘meaning of data’; that is the thematical semantics of data. Similar semantics must be used for the representation of the various objects and real world representations should be equipped with sufficient metadata describing their meaning and implications for the development of the project in a language understandable by all different parties. Consistent application of terms is thereby a prerequisite for successful implementation and unambiguous adoption of legislation, regulations, guidelines and interpretations and should also decrease possible semantic uncertainties caused by inconsequent applications of terms and definitions.

To achieve this, a glossary shall be established to define the meaning of those terms regarding geographic information that are used regularly within infrastructural projects. This will ensure that such terms are consistently and correctly interpreted, as far as is practicable, at all stages of the lifecycle of an infrastructural project. Various (engineering) companies will be visited and, in addition with information gathered with the help of a questionnaire, information about commonly used semantics, attributes, definitions, standards, etc. gathered. Finally, a concept will be developed for the harmonized use of common semantics together with additional metadata including information about possible risks in the representations as well as the impact of these representations on the design and security of the infrastructure.

REFERENCES

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*. London: Taylor&Francis.

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

Corbea Diaz, P.P. 1996. Modelling and Visualization of 3D Geo-Spatial Data. MSc thesis, ITC, Enschede, The Netherlands.

Date, C.J. 1995. *An introduction to Database Systems*. New York: Addison-Wesley Publishing Company.

Dilo, A. 2006. Representation of and reasoning with vagueness in spatial information–A system for handling vague objects. PhD thesis, ITC, Enschede, The Netherlands.

Egenhofer, M.J. & Herring, J.R. 1990. A mathematical framework for the definition of topological relationships. *Proceedings of the 4th International Symposium on SDH*. Zurich, Switzerland, pp. 803-813.

Foody, G.M. & Atkinson, P.M. 2002. *Uncertainty in Remote Sensing and GIS*. West Sussex: John Wiley & Sons Ltd.

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

Hack, R., Orlic, B., Ozmutlu, S., Zhu, S. & Rengers, N. 2006. Three and more dimensional modelling in geo-engineering. *Bulletin of Engineering Geology and the Environment* 65(2): 143-153.

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

Houlding, S.W. 1994. Uncertainty, Sampling Control and Risk Assessment. In: Houlding, S.W. (ed.), *3D Geoscience Modeling–Computer Techniques for Geological Characterization*: 185-200. Berlin: Springer-Verlag.

ISO 9001:2000 Quality Management Standard.

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

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

Lucieer, A. 2004. Uncertainties in Segmentation and their Visualisation. PhD thesis, ITC, Enschede, The Netherlands.

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

Molenaar, M. 1989. *An Introduction to the theory of spatial objects modelling*. London: Taylor&Francis.

Molenaar, M. 1990. A formal data structure for 3D vector maps. In: *Proceedings of EGIS'90.*, Amsterdam, The Netherlands, Vol. 2, pp. 770-781.

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

Oosterom, P.J.M. van; Vertegaal, W.; Hekken, M. van & Vijlbrief, T. 1994. Integrated 3D Modelling within a GIS. *International GIS workshop AGDM'94 (Advanced Geographic Data Modelling)*, Delft, The Netherlands, pp. 80-95.

Oosterom, P.J.M. van; Stoter, J. & Jansen, E. 2006. Bridging the worlds of CAD and GIS. In: Zlatanova, S. & Prosperi, D. (eds.), *Large-scale 3D data integration–Challenges and Opportunities.*: 9-36. London: Taylor&Francis.

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

Pilouk, M. 1996. Integrated modelling for 3D GIS. PhD thesis, ITC, Enschede, The Netherlands.

Raper, J. 1989. *GIS–Three dimensional applications in geographic information systems*. London: Taylor&Francis.

Raper, J. & Maguire, D.J. 1992. Design models and functionality in GIS. *Computers & Geosciences* 18: 387-394.

Sides, E.J. 1992. Reconciliation studies and reserve estimation. In: Annels, A.E. (ed.), *Case histories and methods in mineral resource evaluation*: 197-218. Geological Society Special Publication.

Smets, P. 1996. Imperfect information: Imprecision, and uncertainty. *Uncertainty Management in Information Systems:* 225-254.

Watt, A. 1993. *3D Computer Graphics*. Wokingham: Addison-Wesley.

Zhang, J. & Goodchild, M. 2002. *Uncertainty in Geographical Information*. London: Taylor & Francis.

Zlatanova, S. 2000. 3D GIS for urban development. PhD thesis, ITC, Enschede, The Netherlands.

Zlatanova, S.; Rahman, A.A. & Pilouk, M. 2002. Trends in 3D GIS Development. *Journal of Geospatial Engineering* 4: 1-10.

Zlatanova, S.; Rahman, A.A. & Shi, W. 2004. Topological models and frameworks for 3D spatial objects. *Journal of Computers & Geosciences* 30: 419-428.

Zlatanova, S. & Stoter, J. 2006. The role of DBMS in the new generation GIS architecture. In: Rana, S. & Sharma, J. (eds.), *Frontiers of Geographic Information Technology*. Berlin: Springer-Verlag.