Back

The determination of interpretation uncertainties in subsurface representations

W. Tegtmeier[1, 2], R. Hack[1] & S. Zlatanova[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 geo-engineering re-use of existing data and real world representations is, at present, limited to non-existent. This is mostly due to difficulties regarding the use of data obtained by a number of various professionals. The main problem in this respect is a lack of standardization of data ("lack of data harmonization") and, underlying, the often-unknown quality of the collected data and derived real world representations. Particularly in geological interpretations, uncertainties are high, since only sparse information is available for the interpretation process. The paper presents a methodology, which will be applied in order to determine the influence of so-called "interpretation uncertainties" on subsurface representations and to develop an appropriate way to include quality and uncertainty expressions in the metadata of the subsurface representation. In addition, an outlook will be given regarding the problem of data harmonization and standardization within the process of infrastructural development.

1 INTRODUCTION

Increasing mobility is of major importance in today's society. To be able to cover the needs of the world's citizens concerning their unobstructed movement, sufficient infrastructural capacities (e.g. highways, railways, airports, etc.) in the same way as infrastructure security and transportation safety must be ensured. Clearly, this requires the occupation of a number of specialists, each of them facing different problems, which have to be solved. With it, new civil infrastructures must be planned, designed, and built and existing structures monitored, maintained and eventually be abandoned. These diverse processes are commonly spread over the whole lifecycle of civil infrastructures with a duration of commonly tens of years. The lifecycle can generally be subdivided into six main stages that are namely:

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

For the execution of the various tasks during infrastructural development, the skills of a number of various professionals (e.g. civil engineers, engineering geologists, GIS technologists, etc.) are needed. Large quantities of geo-information (e.g. GIS-, CAD-, and various other data sets) are collected, generated, (re-) used, managed and exchanged throughout the lifecycle of a civil infrastructure and the main problem as identified 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 problem of data harmonization is partly caused by the lack of information about qualities and possible uncertainties regarding the collected data as well as derived real world representations. Still, at present, large parts of the data as well as representations are not equipped with quality or uncertainty information. This aggravates the communication and also co-operation between the different parties involved in infrastructural development and intensifies the problems concerning the (re-) use of geo-information as delivered by diverse companies and experts. This missing uncertainty information regarding various types of geo-information and real world representations, and also the use of different types of data structures, geo-information management systems and software packages are, thus, the main obstacles when trying to achieve data harmonization in large infrastructural projects (Figure 1). Consequently, the question is: How can geo-information be harmonized and equipped with uncertainty estimations?

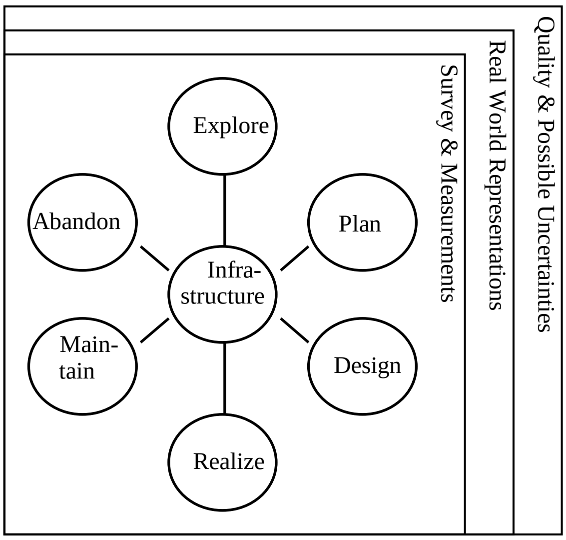

Figure 1. The lifecycle of civil engineering infrastructures.

2 QUALITY AS PART OF THE GEO-INFORMATION

Many people of different professions are involved in infrastructural projects. Regarding the fact that they have to rely on the correctness of the work that is delivered to them by other experts in order to (re-) use this information for further planning and decision making, the quality aspect of geo-information plays an important role in infrastructural development. This makes quality, thus, an important aspect of geo-information and to be able to make successful use of collected data as well as derived representations and interpretations, it is important to receive indications about their quality (Hack 1997, Dilo 2006). Before one can start to determine the quality of the diverse types of geo-information, however, it is important to understand the meaning behind the term "quality".

Countless definitions can be found in the literature, varying for each profession (e.g. car industry, medicine, education, engineering, etc.) they have been especially defined for. In their pioneering work, Harvey & Green (1993), for example, determined the nature and usage of quality in relation to higher education, where they conclude that quality is often referred to as a relative concept. First, quality is described to be relative to the user of the term and the circumstances in which it is invoked. Then again, regarding higher education, is the "benchmark" relativism of quality, where, on the one hand, quality is to be seen in terms of absolutes and, on the other hand, quality is to be judged in terms of absolute thresholds that have to be exceeded to obtain a quality rating. Following, Harvey & Green suggested that quality should rather be grouped into five discrete but interrelated ways of thinking, rather than being described by only one meaning.

The main definition of quality, however, as used by many engineers and scientists and as defined in various international standards (e.g. ISO 9001:2000) is derived from the meaning of quality as fitness for purpose; that is namely quality as satisfying the determined needs of the user. In these definitions, it is stated: "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."

Many factors can affect the quality of data and representations and, eventually, lead to imperfections in the data as well as in the resulting work of the various companies. Different kinds of imperfection in data have been defined in the work by Smets (1996). The main aspects are, accordingly, imprecision, inconsistency, and uncertainty. Thereby, imprecision and inconsistency are properties of the data, whereas uncertainty is introduced into the data by attaching weights to the worlds in order to express our opinion about which might be the real world situation.

Since the problem of data and representation quality together with the numerous factors influencing this quality is too complex to be covered all at once, this specific part of the research is focused on uncertainties in geo-information and real world representations concerning the geotechnical (subsurface) part of infrastructural development. This seems to be most appropriate considering the fact that, usually, only sparse information is available for the interpretation as well as representation of the geological situation at the construction site and, thus, the knowledge and experience of the geo-engineers has a significant influence on the final result.

2.1 *The uncertainty aspect of quality in geo-information*

Uncertainty in geo-information plays an important role throughout the development of infrastructural projects, because it can affect the future (re-) use and processing of geo-information, and also, most importantly, the process of decision making in these large projects. Despite the number of initiatives trying to reduce the uncertainty from an end-users and decision-makers perspective, it is, still, not possible to completely eliminate this factor of uncertainty (Foody & Atkinson 2002).

Often, uncertainty is described in rather general terms as "…a measure of the difference between estimation and reality". This, for example, might be the difference between the thickness and extent of a sand lens as determined via an interpretation of borehole and CPT data as compared to the real world situation; expressed in percentage. A definition similar to this rather general description is used in statistics, where the uncertainty is defined as "the estimated amount or percentage by which an observed or calculated value may differ from the true value".

In the same way as the quality aspect, uncertainty as part of this quality aspect is determined by different types of uncertainty (Figure 2). These are, for example, uncertainty with regard to spatial prediction, uncertainty resulting from site investigations/ surveys/ measurements, or uncertainties resulting from geological and geotechnical interpretations (i.e. mainly caused by limited amounts of data).

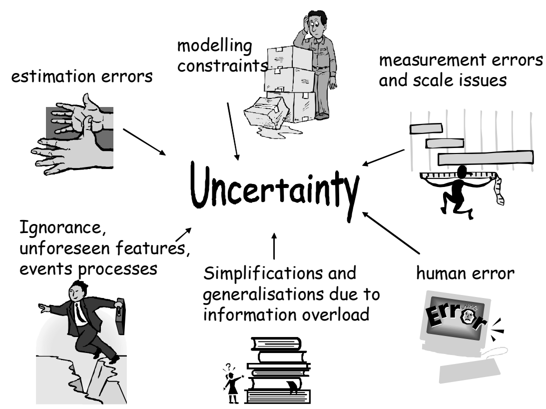

Figure 2. The different types of uncertainties in geo-engineering (Hack et al. 2006).

These days, numerous techniques are available for the determination of uncertainty resulting from the process of spatial prediction. Depending on the quantity and quality of available data, but also on the type of object (i.e. continuous or categorical), for which the uncertainty must be determined, different estimation techniques are frequently used; that are namely geostatistical simulations, kriging and probability-based methods (Orlic 1997, Zhang & Goodchild 2002).

As described by Hack et al. (2006), also in geo-engineering work it is (or should be) common practice to make an estimation of the errors/possible errors in the geotechnical properties of the subsurface and the influence of these errors on the engineering structure to be built in or on it. Different methodologies, such as the "geotechnical base-line methods"(Staveren & Knoeff 2004), probability studies and Monte Carlo simulations (Viseur & Shtuka 1997, Hack 1998, Hack et al. 2003), are applied to give a certain amount of quantification of possible errors in the design of an engineering structure due to uncertainty regarding the subsurface properties. Statistical routines exist, thus, in extenso, to calculate the temporal-spatial distribution of properties in a unit (see also Deutsch & Journal 1998, Houlding 2000).

Next to the uncertainty associated with spatial prediction or the prediction of geotechnical properties, there are, however, mainly two other sources of uncertainty one should constantly be aware of, since these types of uncertainties are less prominent and, thus, difficult to be defined. Due to the scarcity of data, these additional types of uncertainty are most prominent in geo-engineering and, accordingly, in subsurface real world representations. As described by Houlding (1994), these two sources of uncertainty are:

1. The potential for investigation errors (i.e. locational errors or measurement errors caused by wrongly calibrated machines)
2. The potential for interpretations errors (i.e. uncertainty introduced by the expert, depending on the experience and prior knowledge)

Unfortunately, there is little one can do about these specific types of uncertainty in geo-information. To be able to quantify the uncertainties regarding investigation error in sample and observation values, comprehensive research would be necessary into each of the common investigation techniques in use. In the same way, it is rather difficult to determine uncertainties that are caused by errors made during the interpretation of geological features. This is a rather subjective procedure and up till now, there is no way of incorporating it into a computerized approach unless we are prepared to quantify ourselves during the interpretation process.

Thus, numerous estimation techniques for the determination of uncertainties associated with spatial prediction have been developed and are frequently used in practice. However, especially regarding the problem of uncertainties in subsurface real world representations, more research is still to be undertaken and especially the so-called "interpretation uncertainties"; that are uncertainties introduced into the representation by the experts themselves; must be determined and communicated, as they form a dominant source of uncertainty in geo-information.

3 CRITICAL RESEARCH ISSUES AND DEVELOPMENTS

Due to the fact that for the representation of the subsurface (geotechnical) situation at a construction site only sparse information is available, the knowledge and experience of the interpreter plays an important role regarding the outcome of the interpretation. The quality of his experience and "a priori knowledge" that is of major importance for the interpretation process can, however, not be qualified at present. If the engineer/geologist is good, this will result in a good and reliable geotechnical representation. If the engineer/geologist is not as good, it will result in higher uncertainties and, thus, a poor geotechnical representation. Today, many up-to-date analyses are available describing all sorts of uncertainties in measurable properties. Without an indication on the level of interpretation uncertainties to be expected in the representation, it is, however, difficult to rely on any geotechnical representation and to use it for further planning and decision-making.

Therefore, part of the research will be focused on the determination and communication of the so-called "interpretation uncertainties" in subsurface geotechnical representations.

As this problem cannot be completely solved in this limited amount of time, a first step will be made towards an acceptable solution. The goal within this research is to arrive at a description of the level of interpretation uncertainty to be expected in a certain interpretation or representation of subsurface conditions. This level of interpretation uncertainty in geotechnical representations is, at this time, intended to be described 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. For the determination of the level of interpretation uncertainty, a weighting system will be developed 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 measures as needed for 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, these factors are weighted depending on their influence on the final interpretation uncertainty to be expected in this geotechnical representation.

In order to get insight in the present use of uncertainty information, it will be co-operated with various engineering companies throughout this research. A number of companies involved in infrastructural development will be visited and questioned about their use of uncertainty information in subsurface real world representations and case studies will be analyzed in order to acquire information about the influence of the expert knowledge on the quality of a real world representation.

Finally, the newly determined uncertainty information will, ideally, be included in the metadata; that is “data about data”, additional information that is used to provide further information to, for example, attribute tables; of the subsurface (geotechnical) representation and, if possible, be equipped with supplementary information regarding the implications of this interpretation on the construction of the infrastructural project.. This should, significantly improve the communication between the companies involved in infrastructural development and facilitate the (re-) use of the geo-information.

4 CONLUSIONS & FUTURE RESEARCH

In addition to the missing information concerning possible uncertainties in real world representations, the numerous types of geo-information as used in infrastructural development, a lack of standardization and, especially, harmonization of the geo-information makes the different working steps in civil engineering projects difficult.

In infrastructural development, a number of different experts are involved in the lifecycle of the civil infrastructure. Thereby, different types of data, file formats, software packages, etc. are used for the representation of the real world. Depending on the specialization, also different representation techniques for the representation of the diverse real world objects are available. During the last years, several initiatives have been followed in order to integrate the various types of geo-information (Oosterom et al. 1994, 2006; Zlatanova et al. 2002). The whole problem of geo-information harmonization is, however, too complex to be solved in a short time. Thus, more work still needs to be done to achieve a solution to this problem.

In order to increase the data harmonization and to improve the communication and co-operation of the different parties involved in infrastructural works, the second part of this research will be focused on the topic of data harmonization; with its main focus on the “meaning of the data” (the thematical semantics of data). With it, it is desirable to use similar semantics for the representation of the various objects. Furthermore, 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.

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. Therefore, various (engineering) companies will be visited and, together 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.

REFERENCES

Deutsch, C.V. & Journal, A.G. 1998. *GSLIB: geostatistical software library and user’s guide*. New York: Oxford University Press.

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

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

Hack H.R.G.K. 1997. Digital data for engineering geology: disaster or benefit? In: European Science Foundation, *“Virtual environments for the Geosciences”, Space-time modelling of bounded natural domains*. Rolduc, the Netherlands.

Hack H.R.G.K. 1996, 1998. *Slope Stability Probability Classification*. ITC publ. No 43, Enschede, the Netherlands.

Hack R., Price, D. & Rengers N. 2003. A new approach to rock slope stability - a probability classification (SSPC). *Bulletin of Engineering Geology and the Environment*. Vol. 62: article: DOI 10.1007/s10064-002-0155-4. pp. 167-184 & erratum: DOI 10.1007/s10064-002-0171-4. pp 185-185

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.

Harvey, L. & Green, D. 1993. Defining Quality. *Assessment and Evaluation in Higher Education*. 18(1).

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.

Houlding, S.W. 2000. *Practical geostatistics: modelling and spatial analysis*. New York Berlin Heidelberg: Springer-Verlag.

ISO 9001:2000 Quality Management Standard

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.

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

Staveren, M.Th. van & Knoeff, J.G. 2004. The geotechnical baseline report as risk allocation tool. In Hack, R.; Azzam, R. & Charlier, R. (eds), *Engineering geology for infrastructure planning in Europe: a European perspective.* Lecture notes in earth sciences, vol. 104. Berlin Heidelberg New York: Springer-Verlag.

Viseur, S. & Shtuka, A. 1997. Advances in stochastic boolean simulation of channels. *15th GOCAD Meeting Report.*

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.