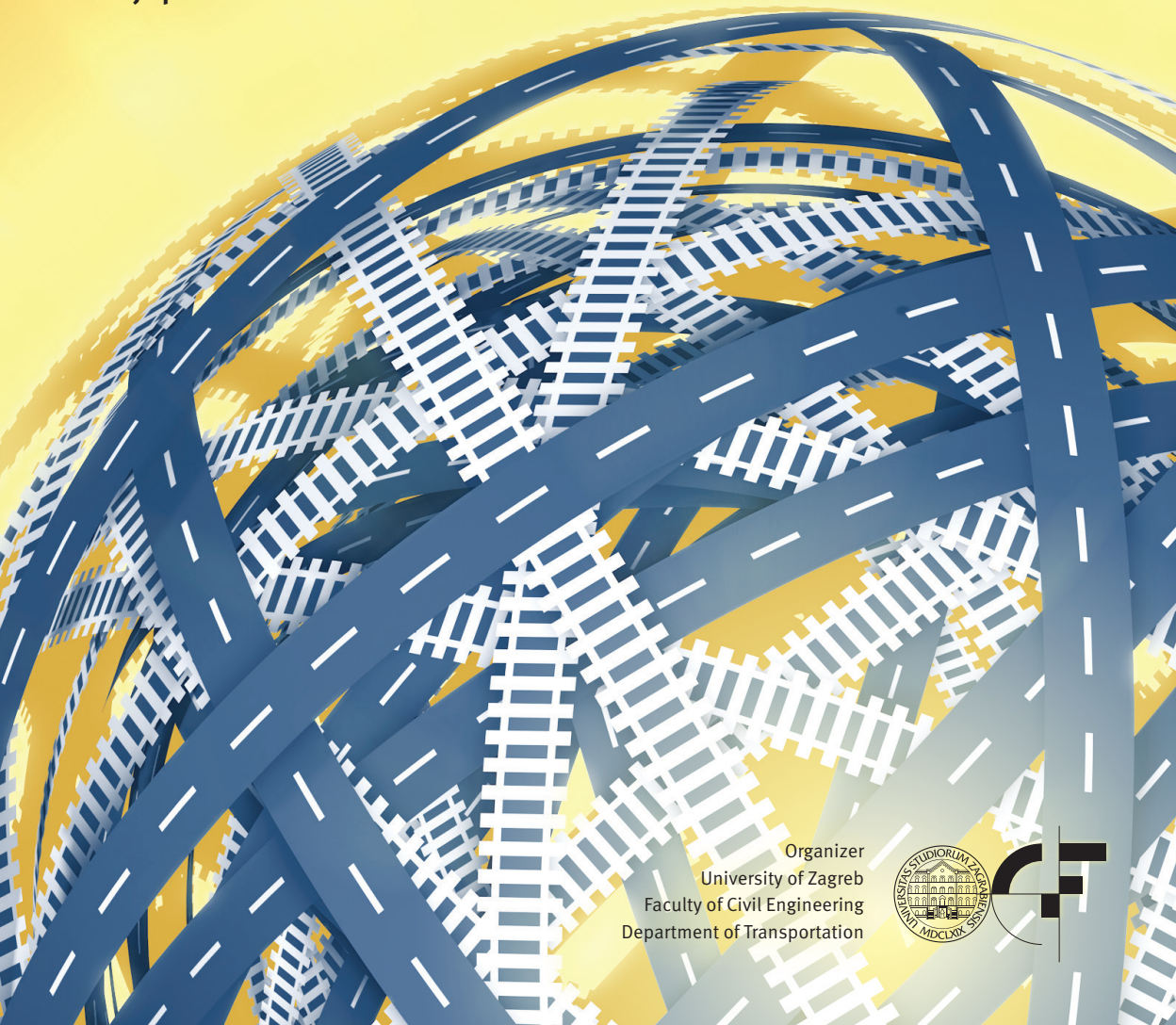


CETRA 2016

4th International Conference on Road and Rail Infrastructure
23-25 May 2016, Šibenik, Croatia

Road and Rail Infrastructure IV

Stjepan Lakušić – EDITOR



Organizer
University of Zagreb
Faculty of Civil Engineering
Department of Transportation



CETRA²⁰¹⁶

4th International Conference on Road and Rail Infrastructure
23–25 May 2016, Šibenik, Croatia

TITLE

Road and Rail Infrastructure IV, Proceedings of the Conference CETRA 2016

EDITED BY

Stjepan Lakušić

ISSN

1848-9850

PUBLISHED BY

Department of Transportation
Faculty of Civil Engineering
University of Zagreb
Kačićeva 26, 10000 Zagreb, Croatia

DESIGN, LAYOUT & COVER PAGE

minimum d.o.o.

Marko Uremović · Matej Korlaet

PRINTED IN ZAGREB, CROATIA BY

“Tiskara Zelina”, May 2016

COPIES

400

Zagreb, May 2016.

Although all care was taken to ensure the integrity and quality of the publication and the information herein, no responsibility is assumed by the publisher, the editor and authors for any damages to property or persons as a result of operation or use of this publication or use the information's, instructions or ideas contained in the material herein.

The papers published in the Proceedings express the opinion of the authors, who also are responsible for their content. Reproduction or transmission of full papers is allowed only with written permission of the Publisher. Short parts may be reproduced only with proper quotation of the source.

Proceedings of the
4th International Conference on Road and Rail Infrastructures – CETRA 2016
23–25 May 2016, Šibenik, Croatia

Road and Rail Infrastructure IV

EDITOR

Stjepan Lakušić
Department of Transportation
Faculty of Civil Engineering
University of Zagreb
Zagreb, Croatia

CETRA²⁰¹⁶

4th International Conference on Road and Rail Infrastructure

23–25 May 2016, Šibenik, Croatia

ORGANISATION

CHAIRMEN

Prof. Stjepan Lakušić, University of Zagreb, Faculty of Civil Engineering
Prof. emer. Željko Korlaet, University of Zagreb, Faculty of Civil Engineering

ORGANIZING COMMITTEE

Prof. Stjepan Lakušić	Assist. Prof. Maja Ahac	All members of CETRA 2016 Conference Organizing Committee are professors and assistants of the Department of Transportation, Faculty of Civil Engineering at University of Zagreb.
Prof. emer. Željko Korlaet	Ivo Haladin, PhD	
Prof. Vesna Dragčević	Josipa Domitrović, PhD	
Prof. Tatjana Rukavina	Tamara Džambas	
Assist. Prof. Ivica Stančerić	Viktorija Grgić	
Assist. Prof. Saša Ahac	Šime Bezina	

INTERNATIONAL ACADEMIC SCIENTIFIC COMMITTEE

Davor Brčić, University of Zagreb
Dražen Cvitanić, University of Split
Sanja Dimter, Josip Juraj Strossmayer University of Osijek
Aleksandra Deluka Tibljaš, University of Rijeka
Vesna Dragčević, University of Zagreb
Rudolf Eger, RheinMain University
Makoto Fujii, Kanazawa University
Laszlo Gaspar, Institute for Transport Sciences (KTI)
Kenneth Gavin, University College Dublin
Nenad Gucunski, Rutgers University
Libor Izvolt, University of Zilina
Lajos Kisgyörgy, Budapest University of Technology and Economics
Stasa Jovanovic, University of Novi Sad
Željko Korlaet, University of Zagreb
Meho Saša Kovačević, University of Zagreb
Zoran Krakutovski, Ss. Cyril and Methodius University in Skopje
Stjepan Lakušić, University of Zagreb
Dirk Lauwers, Ghent University
Dragana Macura, University of Belgrade
Janusz Madejski, Silesian University of Technology
Goran Mladenović, University of Belgrade
Tomislav Josip Mlinarić, University of Zagreb
Nencho Nenov, University of Transport in Sofia
Mladen Nikšić, University of Zagreb
Dunja Perić, Kansas State University
Otto Plašek, Brno University of Technology
Carmen Racanel, Technological University of Civil Engineering Bucharest
Tatjana Rukavina, University of Zagreb
Andreas Schoebel, Vienna University of Technology
Adam Szeląg, Warsaw University of Technology
Francesca La Torre, University of Florence
Audrius Vaitkus, Vilnius Gediminas Technical University



VIRTUAL MANAGEMENT OF COMPLEX INFRASTRUCTURE: INFORMATION SYSTEMS IN THE AGE OF BIG DATA

Timo Hartmann

TU Berlin, Civil Engineering Institute, Systems Engineering Chair

Abstract

Currently a fundamental shift in how infrastructure managers and engineers obtain and make use of virtual information is taking place. In recent years, vast amount of information has become available in distributed virtual systems, be it publicly available or within the information infrastructure of a single organization, which is often labelled as BIG DATA. The availability of BIG DATA makes it now possible that information that is required for many decisions is somewhere available, but needs to be meaningfully mined, fused, and visualized. While only some years ago, much work of infrastructure managers and engineers was concerned with specifying the information to make a decision that then was carefully collected by inspecting and measuring, nowadays, most decision making tasks start with a broad search of virtual information that is already somewhere available. Obtaining and making use of information in this new style of working has become much more dynamic and unpredictable, but at the same time much quicker and richer. This paper shows a range of examples from our ongoing research activities from the area of managing road and rail infrastructure systems that shows the potential of this new way of working. Based on these examples, the paper then discusses some of the implication this new way of working has for setting up information management strategies at large organizations, such as engineering companies and public authorities. In particular, the paper argues that modern information management systems need to be open, agile, and distributed to fully support the newly emerging practices.

1 Introduction

Access to information has always been a crucial aspect of infrastructure engineering work, be it during the design and construction of new infrastructure or during the maintenance of existing infrastructure systems. Large parts of engineering work is then also concerned with obtaining information ranging from social aspects, such as traveling behaviour or preferences of abutters, to engineering aspects, such as material strengths or specification of parts, to physical aspects of the surrounding, such as the existing soil conditions or weather data. In traditional practice, engineers carefully considered information requirements upfront for which they then data was carefully collected by inspectors and surveyors before engineering or planning decision making activities started.

Recently, this traditional practice is changing. This change is caused by more and more data sets about existing natural and physical conditions, but also about the behaviour of infrastructure users are becoming available. Examples for such data sets are, for example, Google maps [1] or Open Street Maps [2] that make information available about the existing infrastructure networks in most countries, Google Street View [3] providing photographic imagery, the Dutch AHN (actueel hoogtebestand) [4] that provides detailed 3D points of the entire Netherlands, or different initiatives of European communities to share data about infrastructure [5], such as the city of Enschede's open GIS data repository, or the city of Berlin's or Seattle's 3D Cesium models [6].

Next to these trends in open data, many public and private organizations are starting to collect large amounts of data. These data are often collected without a specific goal or decision in mind. Examples for this new emerging practice are the efforts of many organizations to collect laser scan data of their existing structures, the equipment of bridges and tunnels with sensors and monitoring systems, or the equipment of construction machinery with GPS and other tracking devices.

All these efforts have change engineering practice in the last years significantly. Engineers traditionally started new design efforts with drafting a data collection plan, then collecting these data, and only then starting to design. Today, engineers, more and more, start designing much earlier using existing data and only collect additional data, if some information is not available. The most illustrative example for this practice is probably the many engineering efforts that start on maps printed out of Google Maps.

A close look at these emerging practices shows that engineering work is becoming much more interactive, agile, and iterative. This in particular holds for the application of supporting tools within this practice. To make full avail of the available data by combining different sources and formats, engineers start to develop many customized data fusion and visualization methods. Many of the existing proprietary software programs and database systems do not support this practice well. The missing support, in turn, challenges many of the, at private and public organizations existing, IT systems and IT implementation strategies.

This paper provides a number of examples of data fusion and visualization methods from railway and roadway engineering. All examples show how data from different sources were combined to support decision making and design. Based on these examples the paper will discuss how existing IT systems and IT system implementation strategies need to change to support such combination of data from different sources. The paper is structured as follows: The next section will provide a brief introduction about data fusion and BIG DATA as it applies to infrastructure design and planning. Afterward the paper will provide a number of data fusion examples from road and railway design. After these examples we will provide our argument of how current IT system and implementation needs to change to allow for benefiting from the vast amount of data that is becoming available.

2 BIG DATA driven simulation and visualization systems

To support their design work engineering and planners make use of simulation and visualization models. These models abstract and formalize knowledge about the characteristics of a specific physical situation and a number of social, technical, or natural processes within that situation [7, 8]. For example, spatial models abstract an urban area using a geographical information model and different behavioural processes of humans that live in the area. Another example are finite element models of bridges that abstract the structural behaviour of the bridge. Through abstraction simulation and visualization models allow designer to experiment with perceived future conditions by changing certain parameters of the physical situation, in the examples above some geographical or structural characteristics that reflect certain transformation possibilities. This allows designers to explore alternatives for designing and planning inquiries that are too complex to be meaningfully represented by physical prototypes or mathematical models.

If simulation and visualization models provide adequate abstractions of the existing natural, technical, and social conditions, models become a none-replaceable part of all engineering and planning decision making work throughout the entire life-cycle of an infrastructure system. This potential has been discussed by many. Research for example has shown that data driven simulation and visualization models can bridge knowledge boundaries between different specialists [9, 10] or enabling multi-disciplinary design teams to generate new knowledge in the form of creative design ideas [11,12].

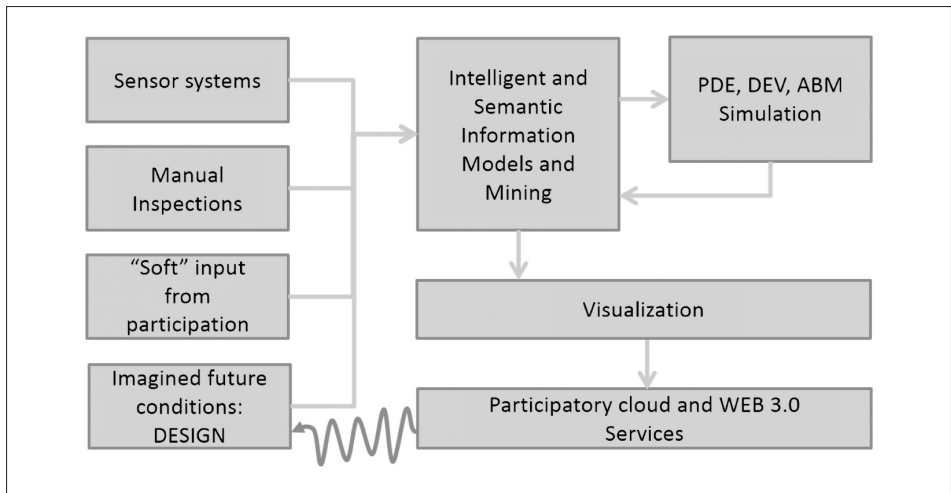


Figure 1 Basic architecture of a decision support system

Figure 1 depicts the core architecture of such data driven simulation and visualization systems. They all rely on input from a variety of different data sources. As design and planning work for infrastructure systems always need to account for the status quo, much of these sources are related to information about the existing natural, physical, and social conditions around an infrastructure network. Traditionally, large parts of the data input describing the existing conditions relied on manual inspections and surveying exercises.

With the fast improvement of sensing technology much of the labor intensive manual inspection work is slowly replaced by sensor systems. The use of sensor systems can replace much of the labor intensive field work. At the same time, however, the adequate use of sensors still requires thorough upfront planning. Additionally, cleaning and analyzing sensor data is still very labor intensive.

Additionally to data about the existing conditions, soft data about the requirements, wishes, and constraints of stakeholders involved around a design or planning effort become increasingly important. This data is often less standardized, but as important as input to design and planning efforts as data about the existing conditions. After all, the re-designed infrastructure network's purpose is to support the social and economic needs of citizens.

A final important source of data are the different design ideas. As planning and design relies on input from many different specialists, every design and planning effort always involves vast amounts of data representing these different inputs. Additionally, planning and design is also an iterative process. Data representing design ideas ideally therefore captures different concurring possibilities describing different alternatives for future actions. Obviously, capturing different alternatives increases the amount of data significantly.

To allow for meaningful decision making around these different types of data, information models are required that allow to meaningful convert the raw data into meaningful information. To this end, information models provide a computer based representation of data in the form of objects that allow humans, as well as, computers to reason about the data. For example, the physical characteristics of structures can be represented by using parametric object building information models that describe structures as building elements, such as beams, columns, or walls. Additional information can then be attached to these elements. Some well-known building information schema are the Industry Foundation Classes (IFC) ISO standard or the CityGML standard. Moreover, most 3D based parametric CAD systems use their own representation schema.

Another example are geo-spatial data that most of the existing GIS applications represent using polygons. Again different information ranging from the demographics of a specific area to natural and geographic characteristics of an area can be attached to the basic polygonal object representing a specific area. Next to these geometrical ways to represent data meaningfully with the computer, many other possibilities exist, ranging as wide as meaningfully named SQL tables or NoSQL based document storage.

Oftentimes, the transformation of raw input data into meaningful information models is not easy. Manual conversion is often error prone and time consuming. Therefore, much recent research and development work has focused on the development of advanced data mining methods applying state of the art machine learning methods. These methods automate the cumbersome work of categorizing input data of different formats, such as from images, point clouds, or historical databases. Next to machine learning methods, the conversion of raw data into meaningful information models also relies on data fusion mechanisms that allow to combine raw data from different sources.

Once data is meaningfully represented using semantic information models, computer applications can be developed that use the data to meaningfully simulate possible scenarios of future conditions. Often used simulation methods in the area of infrastructure design and planning are methods based on partial differential equations, discrete event simulations, or agent based simulations. What all of these methods have in common is that they take data meaningfully represented through information models as input. Then the data is used to simulate a number of possible future alternatives. The outcomes of these simulations, in turn, come in the form of additional data. If different alternatives and scenarios are simulated, the amount of these data is often larger than the initial input data. Again, these output data need to be meaningfully converted and stored in semantic information models to allow computers and humans to reason with the simulation output.

To support meaningful decisions all data, be it data about existing conditions or simulation output, have to be represented to decision makers. To this end, purposeful data visualizations need to be generated. This generation again requires the fusion and mining of the existing data with the help of meaningful semantic information models.

To close the cycle, designers and planners then make decisions about the future configuration of an infrastructure network, about maintenance cycles, and so forth. Again these decision are oftentimes modeled and can serve as data input for the next decision cycle.

In summary, the meaningful introduction of simulation and visualization models in engineering design processes requires a tight integration with large amounts of data representing a wide range of different aspects. Because design by definition involves the creation of something new, it is important for simulation and visualization models to provide an accurate model of the existing physical situation for a meaningful departure from this status quo. Modeling this existing condition, in turn, requires accurate input data that can stem from a large variety of different sources, ranging from existing drawings and documents, to manual collection, to sensor based measurements using, for example, laser scanners or photogrammetric methods. At the same time, these data have to be meaningfully combined with computer based representations of design requirements and suggested future design alternatives. To illustrate this practice, the following section introduces four examples of such data driven decision support systems from the domains of road and railway design and planning.

3 The cases

Without going into detail the paper will provide brief descriptions of four cases of previous or ongoing research work. More detail about each of the cases can be found, with the exception of the second case, in more detailed journal publications. I would like to refer the interested reader to these publications to learn more about the specific system implementations of each of the cases. The four cases are:

- 1) An illustration of how to support construction activities for railway stations. This case shows how different input data are fused to coordinate different design and construction tasks on two major European train stations: Amsterdam Centraal and Arnhem Centraal [13]
- 2) The second example from railway planning and design was concerned with using the existing data at a major European railway organization to use machine learning techniques to predict delay propagation based on historical schedule data.
- 3) The project to illustrate the possibilities to support road construction activities was conducted within the context of an asphalt paving living lab. The simulation and visualization system developed was concerned with tracking and controlling asphalt paving operations in real time to ensure continuous process improvements and long term quality of the constructed road [14, 15, 16].
- 4) An example to illustrate long term planning efforts for road network management was conducted within a study to explore the possibilities of machine learning methods to understand deterioration of roads based on design parameters and environmental conditions [17, 18].

3.1 Refurbishment design and construction planning of train stations

The inner city traffic infrastructure is crumbling. In the European context this is particular visible at the existing major train stations. To overcome the problem with the existing Dutch stations the Dutch government has started a program to renovate its major train stations for example in Amsterdam, Delft, Arnhem, or Rotterdam. One major issue during the renovation and upgrade of train stations is that new design ideas need to be integrated with old existing structures and architecture. Engineering the integration of such new ideas is often one of the most time consuming processes during the overall planning, engineering, and construction process. However, the required coordination work can be significantly improved by decision support systems that allow to clearly understand how new proposed renovation measures will fit with the existing conditions.

In the case of the Amsterdam railway station renovation it was therefore decided to obtain a laser scan of the existing conditions. It was then planned to match the scan with the new proposed design of all architects and engineers that were involved on the project. The laser scans provided a large amount of data that had to be integrated meaningful in the engineering process. To this end, the project team developed a specific system. The basis for the system was an accurate Building Information Model (BIM) that was established based on the data from the laser scans. This BIM model transformed the semantically poor description of the existing condition from the laser scans – laser scans simply provide a so called point cloud that describe existing geometry in the form of a large number of three dimensional points – into an accurate semantic description of the existing train station. This semantic description, for example, provided information about the type of objects, e.g. column, ceiling, wall, about the object's material and its function. All this information was essential to support the engineering process on this project.

This as-built BIM was then used in two different ways to support the engineering process of the project. For one, the model provided the engineers with an accurate description of the existing situation that they could use to develop creative ideas for how to best engineer the transformation of the train station. Additionally, after specific alternative solutions had been

engineered, these solutions could automatically be compared with the existing situation. In this manner, the engineers on the project were able to identify a number of conflicts between the existing conditions and their envisioned future condition that would have probably required costly adjustments of the design during the later construction phase, Figures 2 and 3.

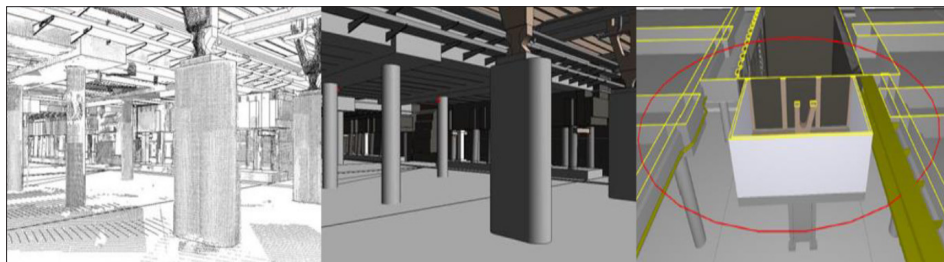


Figure 2 Understanding the as-built situation in relation with proposed engineering designs. From point cloud to clash detection (left to right)

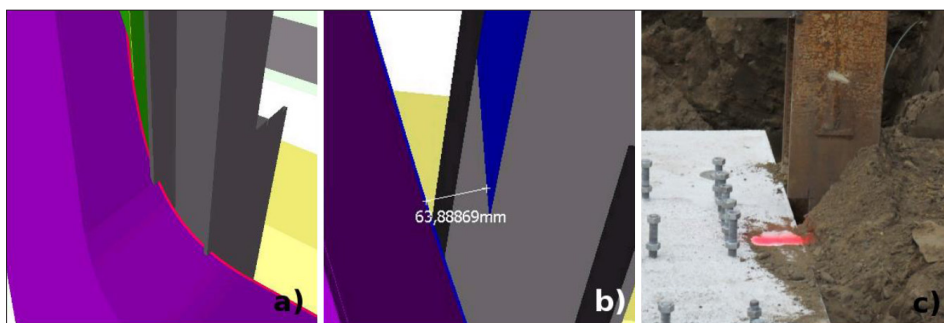


Figure 3 Use of the 4D CAD model to resolve a conflict between the foundation and a temporary structure after construction work had been started. Left screen-shot: Conflict between a temporary column and a to be installed permanent column. Middle screen shot: Identifying the exact dimension of the conflict using the 4D CAD software's measurement function. Right screen shot: The existing situation on the construction site before the installation of the new permanent column

On the Arnhem Centraal project, the project management team even went a step further. On this project, it was decided that all design input from all involved parties should be provided in 3D format. These 3D models were then merged in a common model to understand the different conflicts, not only with the existing conditions, but also within the different design suggestions itself. Moreover, a construction schedule was integrated to establish a so called 4D model to understand possible conflicts and clashes during the construction of the project, Figure 4.

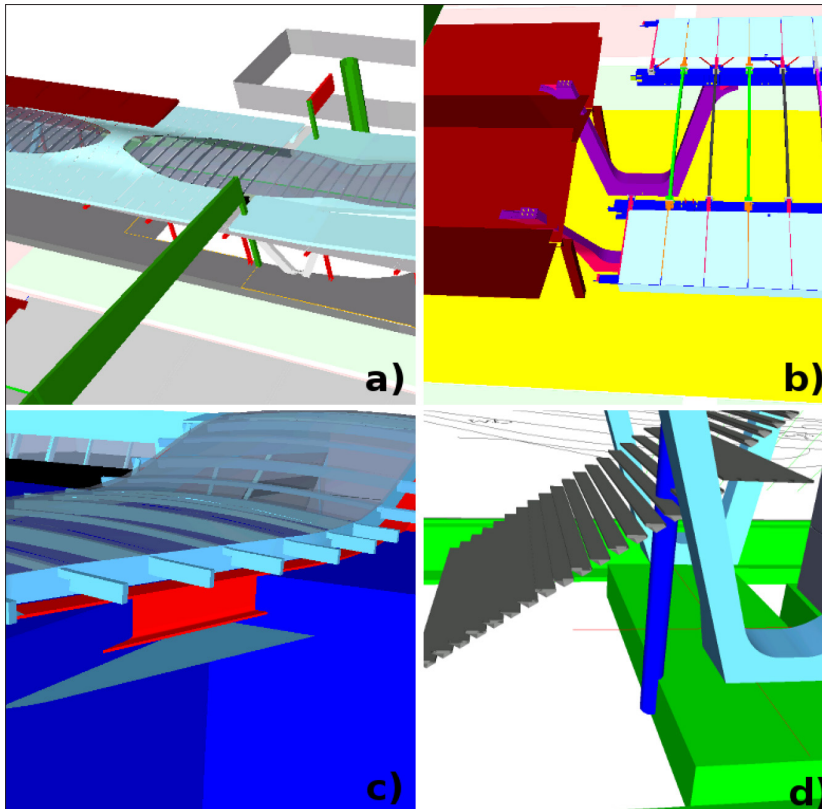


Figure 4 Screen-shots showing different conflicts caused by misaligned construction sequences that were found using the 4D CAD model. The top-left screen-shot (a) shows a conflict between the overhead railway electrification system and the roof construction. The top-right screen-shot (b) shows a conflict between the demolition and installation sequence for the steel structure. The screen-shot on the lower-left (c) shows a conflict between the old and new roof structure. The bottom-right screen-shot shows a conflict between the foundation, temporary stair structure, and the new steel columns

3.2 Predicting delay propagation on railway networks using historical train operations data

Delay propagation is one of the hardest phenomena that railway operations managers face. The delay of a single train within a railway network will often cause delays of a large number of other trains. Predicting how delays will propagate throughout a network is complex, while the effect of delay propagation on travellers is significant. Better understanding the delay propagation is therefore a crucial part of improving the management of railway operations. This case describes the combination of two different databases available at Irish Rail to learn more about delay propagation. In particular, a system was developed that combined information of infrastructure elements, such as tracks, switches, crossings, buffer stops, with the railway networks topology and locations from GIS and CAD systems, with the historical timetable and operational data, such as the arrival and departure time of each train and all delays more than 300 seconds. Most of this information was combined using the existing information model standard RailML that allows to meaningfully model all the above information. Using these different data sources a model to predict delay propagation within the Irish network was developed using artificial neural networks. The model used the delay times, train maximum speed, route maximum speed, speed homogeneity, route departures, vertex

departures, edge length, and percentage of double tracks as input variables to predict the amount of delay propagation.

This case shows the potential of the purposeful combination of different data available at railway management organizations. In particular, the combination of data about the railway infrastructure with historical data about operations seems to be a fruitful endeavor. The developed model allows first predictions of first-order knockdown delays. This model can then be used to understand different problematic sections on the network that are main drivers for delay propagations. Railway management organizations can then use this knowledge to devise measures to mitigate delay propagation, be it through an adjustment of time-tables or through the improvement of the physical railway infrastructure.

3.3 Construction and maintenance of road networks

The road infrastructure is a vital component of any urban transportation system that addresses the travel demands to transport people and goods. In 2009, more than 73% of the total amount of the Tonne-kilometre of the inland goods transportation in EU employed roads as a transportation means [19]. Given the vital role for transportation, in particular for urban regions, there is a clear need for the search of ways to improve the road infrastructure. To accommodate this improvement most roads in urban areas are subject to constant construction work with the aim to rehabilitate existing roads, extend road capacity and to build new roads. Within the urban context such construction work is usually highly disruptive to the economic and social urban processes. Accelerating road construction processes is therefore of utmost importance for the seamless transformation of cities.

Despite this pressing need, most acceleration efforts are currently still hampered by missing understanding of how the construction process influences the quality of the finished asphalted road. This holds in particular for the compaction process. This case presents research and development work towards a data driven real-time decision support system to provide machine operators and site managers with adequate process information about ongoing road construction projects. The system is based upon measurements of asphalt temperature and GPS measurements of all involved construction plant, Figure 5. The measured data is then fused by aligning the measured time stamps and geo-referencing all information. Additionally, simulations about the predicted cooling of the asphalt temperature was integrated in the system.

In this way, the system could provide real-time visualizations of the road construction process to site managers. This allowed these managers to streamline ongoing road construction work while safeguarding quality. In particular, it was now possible to understand when to best compact a certain part of the newly asphalted road in relation to the asphalt temperature. Such understanding is an important factor for safeguarding the final road quality, because if asphalt is compacted while it is too hot, roller compactor will damage the road deck as they will sink into the too viscous asphalt. On the other hand, roller compactors will damage the road deck if the asphalt is too cold as the hardened road deck will crack under the introduced compaction energy.

In a next step this project also started to experiment with understanding the failure of road network in relation to the tracked data from the road construction. To this end, the measured data from the operations were fused and visualized together with the at the Dutch road authority existing GPS based road damage data. Figure 6 shows a first impression of this work.

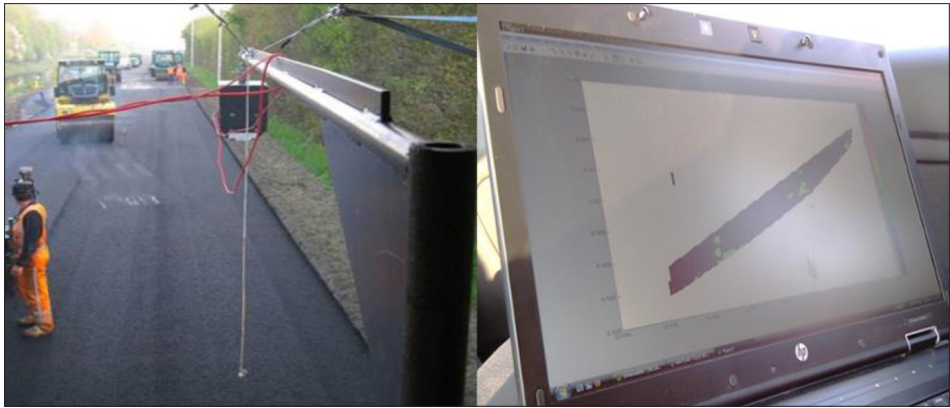


Figure 5 On site temperature measurement with an industrial temperature line scanner (left) and on-site information provision to site management staff during ongoing operations (right)

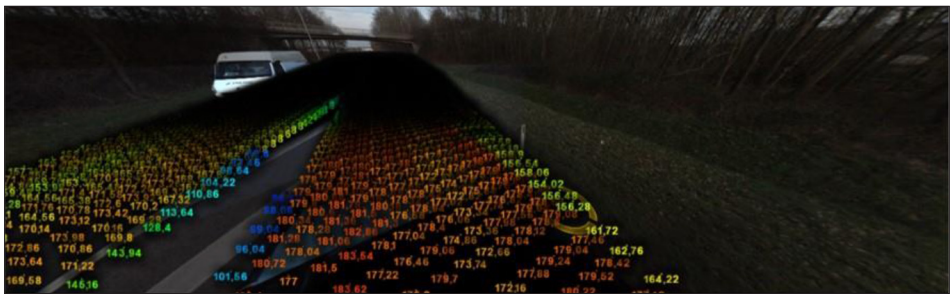


Figure 6 Overlay of temperature measurements from a paving effort with GPS referenced video data of a highway [16]

3.4 Long term prediction of road network quality using historical data

Authorities responsible for planning and designing road networks, use pavement management systems to support their decision making procedures with respect to maintaining pavements in serviceable and functional conditions throughout their life-cycle. A key component of these systems, pavement prediction models play an important role. This case provides an example of how publicly available data from the US national highway administration was used to develop a machine learning prediction method to understand the factor that effect pavement deterioration.

As input for the machine learning mechanisms data from the US long term pavement performance database was used. As specific input variables the study used surface thickness, percentage of asphalt content, percentage of air void and hot mix asphalt, and unit weight of hot mix asphalt as characteristics of the asphalt. Additionally, equivalent single axial load, average annual daily traffic, and average annual daily truck traffic were used as load related input variables. Finally, with annual average precipitation, annual average temperature, and annual average freeze index two climate based variables were also used.

Using regression analysis and artificial neural networks the influence of these parameters on the international roughness index in the short and long term were predicted. Different models were developed to be used within pavement management systems to predict the performance of a specific pavement. The models also showed that the two factors of annual

average precipitation and average annual daily truck traffic had the most influence on pavement deterioration.

All in all, this case shows the possibility to implement machine learning mechanisms to make use of three different source of data available at most road management authorities: design characteristics of paved roads, traffic data about road use, and openly available weather data. By a meaningful combination of these data within an information model and by mining this information model with state of the art machine learning techniques, better prediction models for the deterioration of assets can be developed. These prediction models, in turn, will allow road management agencies to more objectively devise maintenance and repair strategies and plans.

4 Discussion

The above cases all describe applications of data driven decision support systems within infrastructure management and design organizations. What is common between all these cases is that they make use of available data, fuse this data meaningfully, mine the fused data, and then visualize it to support decision making tasks. What is also common between the four cases is that all of these applications were developed in a bottom-up manner driven by the decision support needs at hand. In our earlier work, we argued that only such bottom-up development will make the sound implementation of decision support systems within organizations possible [20-23]. The above cases stress the importance of this IT management strategy.

In what follows we will provide a number of suggestions for the development of information models, service oriented patterns for the interoperability between different information models, and a management framework all targeted at providing the flexibility to quickly develop and implement decision support systems as mentioned above. The recommendations are targeted to overcome some of the existing problems with information management at organizations that oftentimes is too inflexible to support the development of targeted data-driven decision support systems.

4.1 Information Models

An information model can be best defined as a simplified representation of a small, finite subset of the world. As such each of the objects within a model corresponds to some real or abstract object that might exist in the world or within the state of mind of a group of persons or an individual [24, 25]. Each information model can ultimately only represent a certain aspect of the world using a certain categorization and a certain level of detail [24]. Therefore, before an information model can be developed and the above described decisions about the specific form of the information model can be made a specific universe of discourse [24] and a purpose for the model needs to be identified.

After identification of the overarching universe of discourse, decisions need to be made about the level of detail to represent the object, about the degree of decomposition of objects to be represented, about how to represent the relations between objects, and about how to deal with changes to objects over time [25]. How these decisions are made does not depend on any natural law, but is largely determined by arbitrary decisions of the designers of a specific information system. These decisions will be made according to the required kinds of information for the specific purpose the information model should support and due to considerations about how a specific information model needs to be maintained while in use [25].

While making these choices it is important to consider that every information models is a computer based knowledge representation. As such it has to follow a number of within the literature well established principles [26]:

- An information model is a set of ontological commitments. The designers of a database need to agree upon categories for the things that the information model should represent. It is

important to realize that the information model only remains useful if all users of the model follow this initial ontological commitment.

- Every information model is a medium for efficient computation. It is important that every information model, not only encodes knowledge, but ensures that computers can easily process the information model and that the information model can be easily implemented with existing technology.
- An information model is also a medium of human expression. Next to ontological commitment, an information model should also support human communications of those using the information model.

From these three generally accepted principles, three different types of information models can be derived. These three types are summarized in Figure 7, labelled as semantic representation, conceptual completeness, as well as, ease of implementation and querying. The figure also depicts that each of the three quality characteristics are to a certain extent exclusive. The figure depicts that a shift of an information model towards any of the three qualities will inevitably require the reduction of the quality within the other categories. Exclusiveness, in turn, means that for different decision purposes, different information models need be established, even if they describe the same type of data.

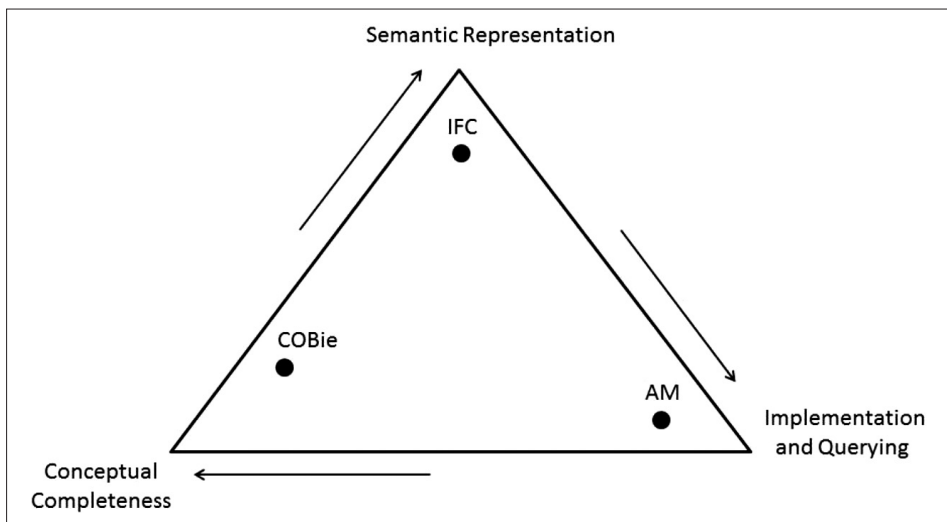


Figure 7 Three types of information models. The three types are to a certain extent exclusive. For example, a shift of a specific information model towards a higher quality of semantic representation will most likely reduce the conceptual completeness and the ease of implementation and querying (1). Similarly, an increase in conceptual completeness will reduce the possibilities of semantic representation and the ease of implementation, while an increase in the ease of implementation will decrease the possibilities for semantic representation and the conceptual completeness

At the outset, and probably the most important type of information models, are models that are highly semantically accurate. Every information model is an abstract representation of reality. As such, it can be seen as a computer based sign system that represents something in the real world. This sign system, in turn, has to be interpreted by somebody who wants to learn something about the real world based on the representation in the computer [27, 28, 29]. The semantic accuracy of the model to represent real world objects allowing for their adequate interpretation from various different points of view at different stages within a design process seems therefore the most important characteristic of an information model.

However, to design meaningful data driven decision support systems two other characteristics of information models are similarly important.

A second type of information models is characterized by high conceptual completeness, or in other words, the degree to which the combined system of signs can represent a specific universe of discourse at hand in its entirety and exactly. Other than semantic accuracy, conceptual completeness requires information models to closely focus on a specific engineering task often around the accurate combination of data from two different sources. To enable such specific engineering tasks a conceptually complete model needs to provide a highly accurate representation of the specific universe of discourse at hand. Redundancies in possibilities to store data related to the information models decomposition need to be avoided. This holds for the chosen categorization of objects and their respective level of detail. Additionally, highly conceptually complete models need to allow domain expert to quickly comprehend the model.

A third type of information models, finally, is targeted at optimizing computational characteristics, in particular, with respect to ease of implementation and ease of model querying. This is important because a specific information model, after all, will need to be implemented in a software system to become of any use. Here question about how well the chosen information model structure can be effectively implemented within database systems or other software and how well a specific information model can be queried and updated move to the forefront. Important then is that IT management of organizations needs to consider this requirement for different types of information models. With the necessity that any one decision support system will need to custom tailor their underlying model to the needs at hand, the question arises how organizations can manage the multiple information models that are required to support their different decision support processes ranging from the strategic to the operational level. The next sub-section provides some recommendations for the required interoperability that need to be established between different models.

4.2 Interoperability

According to Sowa [30], there are a number of patterns to deal with data interoperability between different data sources. The easiest pattern is a “file gateway”, a pattern quite often used within infrastructure design and planning. This pattern relies on a central gateway that is able to read files exported in the format of one information model and translates it to the other. The problem with this pattern is that there is no immediate response, it also adds complexity and additional administrative requirements. For example, file storage and versioning systems have to be designed to organize the meaningful file exchange.

A slightly more advanced pattern is the “data model transformation” pattern. In this pattern an independent software service running on a server is responsible to translate between two information models. The advantage of this pattern with respect to the “file gateway” is the immediate availability of the translated data. However, the development of good “data model transformation” services is difficult and requires a significant effort. A different service has to be developed for each pair of decision support systems that need to exchange data. Additionally, maintaining a large number of these services running in production requires additional administrative effort. Also the additional run time performance overhead should be considered, in particular, with respect to large data sources.

The “schema centralization” pattern makes use of the necessary overlap of information between different information models. It selects overlapping sub-schema from different decision support systems and decouples these sub-schema from the original information models into a central repository. Each of the different decision support systems of an organization can then make use of these common central repositories. Other than the earlier described two patterns the complexity that “schema centralization” introduces is of an organizational nature. It is very difficult to understand which sub-schema of the different deci-

on support systems can be shared within an organization. Additionally, political discussion about ownership and rights of shared schema might form a barrier to their implementation. Finally, the “canonical schema” pattern tries to develop one single information model schema to serve all purposes of an organization. While this schema is often touted as the “holy grail” of information modelling, in particular, in the current discussion about building information models, one should always bear in mind the different purposes of information models introduced in the previous section. Even in the rare case that an organization could make use of a single information model, the implementation of a canonical scheme often cannot be realized. The introduction of a “canonical schema” requires a significant governance effort and a cultural change throughout the entire organization. Designing the schema is also highly complex as all needs of the organization’s decision support systems need to be accounted for. Finally, the implementation of a “canonical schema” will inhibit the innovative potential within the organization to continuously improve its decision support systems. More often than not innovative efforts in organizations that have implemented “canonical schema” in the past were implemented outside the “canonical schema”.

Independent of what combination of schema an organization wants to implement, the infrastructure of decision support systems in relation to the different databases implementing the above patterns need to be managed. The next section, introduces a concept and framework for this management, labelled “Cooperate App Store”.

4.3 The corporate app store

Organizations who want to make full avail of the possibilities that data driven decision support systems offer, need to provide an IT infrastructure that is open, flexible, secure, and that support collaboration within the organization and with third parties. The concept of the corporate app-store provides a framework for the development of an IT infrastructure that fulfils all these criteria. The corporate app-store is schematically presented in Figure 8.

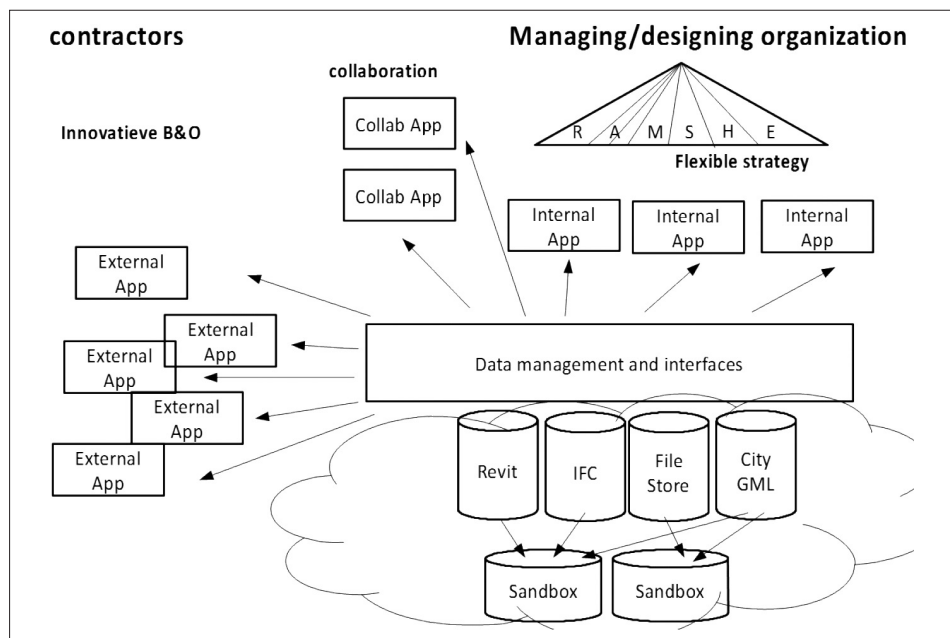


Figure 8 Architecture of a corporate app store

The main purpose of the corporate app store should be the support of the organization's strategy in terms of design and planning of the infrastructure systems. This support should account for the highest amount of flexibility of this strategy, so that the organization will be able to quickly react to changes in its environment.

Different decision making tasks at the strategic, tactical, and operational level can be supported by small self-standing applications each of which present a specific data-driven decision support system. Additionally, applications can be developed that support the collaboration with external parties. Finally, services that the organization outsources can be supported by application that are used only by sub-contractors, but that nevertheless tie into the organization's app-store.

The provision of these different applications on the basis of a central IT infrastructure is made possible by a distributed cloud based data repository. In this repository all data available within the organization should be stored in a distributed database system. At the outset, it is not important in what format the data is stored here. Different information systems and different data storage technologies should be supported, ranging from file based storage to state-of-the-art NoSQL and graph based database systems. Important is only that the data schema and schema implementation is provided openly as source code, as well as, human understandable documentation documents that clearly explain the underlying storage mechanisms as well as the information model is documented. Important is also that all data of the entire organization is stored and documented here.

On top of this distributed cloud based data storage platform, an interface layer is then constructed through which the different applications can access the data. Similar to the distributed data storage, the assumption here is that the interface level can grow evolutionary with the flexible requirements of the organization and their applications. Additionally, the interface should also be made available openly and all interfaces should be documented. Interfaces can then provide different data stored in the distributed repository to application. Interfaces can also take care of certain interoperability requirements using any of the patterns discussed in the previous section.

Having this IT architecture in place allows organizations to flexibly extend their application environment with new data driven decision support systems whenever the need arises. Because of the open nature of the app-store it is possible to develop applications internally or to outsource the application development to outside software developers. Importantly, however, is that the open character prevents the organization to become dependent on one single software company. Further, the open and evolutionary growing character of the app-store prevents organizations from implementing systems that in a number of years will be outdated, so called legacy software.

The app-store's data interface layer allows app-store architects to implement data security mechanisms. None of the distributed data stores should be directly accessed by an application, but only through a well documented and open interface. This allows for the highest amount of control and flexibility in terms of data security.

5 Conclusion

This paper argues that a shift in how engineers and planners work is occurring currently. Practice is moving from targeted information retrieval towards applying the vast amount of existing databases. It supports this argument by providing four illustrative cases of data driven decision support systems for the design and planning of road and railway networks and their components. Establishing the possibilities of data driven decision support systems, the paper then provides a number of recommendations for how organizations can design their IT infrastructure to support this changing practice and to make full avail of the newly arising possibilities. In particular, the paper discussed different purposes of information models,

it introduces different interoperability patterns, and finally it suggests the concept of the corporate app store as a IT implementation and management strategy.

I hope that the paper can provide some fresh insights into the quickly evolving current practice of infrastructure planning and design. Further, it is hoped that the paper can provide some fresh thoughts about how to overcome the problems of current IT management practice that is still heavily dominated by legacy systems, silo-thinking, and closed data systems. To overcome this practice in the upcoming years will be a challenge for most large organizations working in the area of infrastructure management. As it is argued in the paper a new IT management culture that operates flexible, open, transparent, and secure is required to truly support the emerging design and planning practice around data driven decision support systems.

References

- [1] Miller, C.C.: A beast in the field: The google maps mashup as gis/2, *Cartographica: The International Journal for Geographic Information and Geovisualization* 41 (3) (2006), pp. 187-199.
- [2] Haklay, M., Weber, P.: Openstreetmap: User generated street maps, *Pervasive Computing, IEEE* 7 (4) (2008), pp. 12-18.
- [3] Anguelov, D., Dulong, C., Filip, D., Frueh, C., Lafon, S., Lyon, R., Ogale, A., Vincent, L., Weaver, J.: Google street view: Capturing the world at street level, *Computer* (6) (2010), pp. 32-38.
- [4] Swart, L.: How the up-to-date height model of the Netherlands (ahn) became a massive point data cloud, *NCG KNAW* 17.
- [5] Conradie, P., Choenni, S.: Exploring process barriers to release public sector information in local government, in: *Proceedings of the 6th International Conference on Theory and Practice of Electronic Governance, ACM*, 2012, pp. 5- 13.
- [6] Mahdavi-Amiri, A., Alderson, T., Samavati, F.: A survey of digital earth, *Computers & Graphics* 53 (2015), pp. 95-117.
- [7] Küppers, G.: *Computer simulation: Practice, epistemology, and social dynamics*, in: *Simulation*, Springer, 2006, pp. 3-22.
- [8] Hartmann, T., Olde Scholtenhuis, L., Zerjav, V., Champlin, C.: Mindfully implementing simulation tools for supporting pragmatic design inquiries, *Engineering project organization journal*, 5 (2015) 1, pp. 4-13.
- [9] Henderson, K.: Flexible sketches and inflexible data bases: Visual communication, conscription devices, and boundary objects in design engineering, *Science, technology & human values* 16 (4) (1991), pp. 448-473.
- [10] Di Marco, M.K., Taylor, J.E., Alin, P.: Emergence and role of cultural boundary spanners in global engineering project networks, *Journal of Management in Engineering* 26 (3) (2010), pp. 123-132.
- [11] Ewenstein, B., Whyte, J.: Knowledge practices in design: the role of visual representations as epistemic objects', *Organization Studies* 30 (1) (2009), pp. 07-30.
- [12] van Amstel, F.M., Zerjav, V., Hartmann, T., van der Voort, M.C., Dewulf, G.P.: Expanding the representation of user activities, *Building Research & Information* (2014), pp. 1-16.
- [13] Trebbe, M., Hartmann, T., Doree, A.: 4D CAD models to support the coordination of construction activities between contractors, *Automation in construction* 49 (2015), pp. 83-91.
- [14] Vasenev, A., Hartmann, T., Doree, A.: A distributed data collection and management framework for tracking construction operations, *Advanced engineering informatics* 28 (2) (2014), pp. 127-137.
- [15] Vasenev, A., Pradhananga, N., Bijleveld, F., Ionita, D., Hartmann, T., Teizer, J., Doree, A.: An information fusion approach for filtering GNSS data sets collected during construction operations, *Advanced engineering informatics* 28 (4) (2014), pp. 297-310.

- [16] Sluer, B., Aalbers, D., Vasenev, A.: Risico's van afwijkingen in asfaltproces te kwanti_ceren, CROW Infradagen (2014), pp. 18-19.
- [17] Ziari, H., Sobhani, J., Ayoubinejad, J., Hartmann, T.: Analysing the accuracy of pavement performance models in the short and long terms: GMDH and ANS methods, *Road Materials and Pavement Design* (2015), pp. 1-19.
- [18] Ziari, H., Sobhani, J., Ayoubinejad, J., Hartmann, T.: Prediction of IRI in short and long terms for flexible pavements: ANN and GMDH methods, *International Journal of Pavement Engineering* (2015), pp. 1-13.
- [19] E. U. R. Federation, *European road statistics* (2011).
- [20] Hartmann, T., Fischer, M., Haymaker, J.: Implementing information systems with project teams using ethnographic-action research, *Advanced Engineering Informatics* 23 (1) (2009), pp. 57-67.
- [21] Hartmann, T., Van Meerveld, H., Vossebeld, N., Adriaanse, A.: Aligning building information model tools and construction management methods, *Automation in construction* 22 (2012), pp. 605-613.
- [22] Hartmann, T., Levitt, R.E.: Understanding and managing three-dimensional/four-dimensional model implementations at the project team level, *Journal of construction engineering and management* 136 (7) (2009), pp. 757-767
- [23] Hartmann, T.: Goal and process alignment during the implementation of decision support systems by project teams, *Journal of Construction Engineering and Management* 137 (12) (2011), pp. 1134-1141.
- [24] Turk, Z.: Phenomenological foundations of conceptual product modelling in architecture, engineering and construction, *Artificial Intelligence in Engineering* 15 (2) (2001), pp. 83-92.
- [25] Kent, W., Hoberman, S.: *Data and reality: a timeless perspective on perceiving and managing information in our imprecise world*, Technics publications, 2012.
- [26] Sowa, J.F.: *Knowledge representation: logical, philosophical, and computational foundations*.
- [27] Andersen, P.B.: *A theory of computer semiotics: Semiotic approaches to construction and assessment of computer systems*, Cambridge University Press, 1997.
- [28] Hartmann, T., Vossebeld, N.: A semiotic framework to understand how signs in construction process simulations convey information, *Advanced engineering informatics* 27 (3) (2013), pp. 378-385.
- [29] Hartmann, T.: Semiotic user interface analysis of building information model systems, *Journal of computing in civil engineering* 28 (5).
- [30] Erl, T.: *SOA design patterns*, Pearson Education, 2008.