

Management of Data Science: How to Prevent Errors in Work with Data for Sustainable Development?

M. Janakova

Abstract — This paper is focused on designing complex work with available data for improved analysis based on a four-dimensional (4D) model. The aim is a search for unexpected knowledge, error minimization for a large volume of data and better support for sustainable development. Data processing is an activity which changes data on information and knowledge. Knowledge is important for everyday decision-making and increasing a competitive advantage. A competitive advantage influences the market, but a suitable solution for global problems and rapid changes needs collaboration. Optimal collaboration helps solve existing problems. There are numerous approaches with links to extensive data and data science, but errors exist and data science must more talks of existing reality. The reason is the high complexity of the implemented activities and processes. In data science, default layers create computer science, applications, simulations, statistics, analytics and math. These layers are implemented in order to search for a solution in selected fields. This composition (layers and fields) creates the basis for a two-dimensional (2D) model. Good experience provides an extension of the 2D model about relation (dimension) for known intelligences such as Artificial, Business, Computational, Customer and Swarm in a three-dimensional (3D) model. The selection of individual layers and intelligences is based for practical reasons on adopted preferences. This selection (zoom) creates a fourth dimension for the designed model. Based on these four dimensions (layers, fields, intelligences and zooms), a 4D model is designed for a better description of existing reality. The created connection between dimensions breaks unexpected errors and supports collaboration for sustainable development in the global view.

Keywords — Extensive data, data science, errors, sustainable development.

I. INTRODUCTION

SCIENCE is a word used often in global information society. Similarly, companies, organizations and individuals call for innovations and verified approaches. There are a number of analyses, surveys, questionnaires and best examples from practice demonstrating the actual perception and method for information technology (IT) implementation. Many of them use verified methods and methodologies [1] based on achieved knowledge with science support.

Knowledge is needed for proper assistance of implemented

activities. When the IT user has knowledge than he (or she) may decide about the optimal procedure and implemented method. Knowledge is a factor which distinguishes individual approaches and ways of using information technology and its products. Knowledge creates the basis of the implemented activities in a global society in connection with a competitive advantage. [2]

One form of the benefits of a global information society is general accessibility various information via the Internet. The Internet is a medium which provides access to information regardless of time, location, language or preferences. Available data and information is in a wide spectrum of formats. The volume of this data is so extensive that extensive data and data science have to be defined.

Work with extensive data [3] and data science [4] help IT users find unexpected relations between stored data. The default need achieves optimal results quickly, instantly and personally in an optimal quality and format. Good experience brings visualization and support via selected IT products focused on this issue. There are open-source products and also commercial applications and tools. In many examples, IT users have a preference for MS Excel [5] for user-friendly interface and easy control imported data.

In this situation, the question arises as to “Why so many IT users have a preference for MS Excel as opposed to optimized and specialized tools?”, and “Why they do not use the results of data science with verified methods for active work with data?”. Companies and organizations call for new products and services in relation to innovations and a scientific approach. [6] These innovations are based on everyday work with data. These data consist of “treasure” which maps out existing preferences, needs and requirements. The extensive volume of data requires suitable work with data and an optimal approach. Various ways exist and unfortunately a unique solution does not exist.

Knowledge and optimal skills play a key role. A solid orientation in terms of theme and an optimal composition of work is an advantage. Continual education [7] may help with support from on-line learning courses, and of course the finest examples from practice. Science and in particular data science has to provide new knowledge and management work for optimal data processing. There is a need to immediately access practice examples with video support and recommendations

M. Janakova is now with the Department of Informatics and Mathematics, Silesian University in Opava, School of Business and Administration in Karvina, Universitni nam. 1934, 73301 Karvina (corresponding author to provide phone: +420596398111; e-mail: mija@opf.slu.cz).

for breaking errors and mistakes. A similar approach is needed for management of data science and searching for a new approach to current conditions and preferences.

There are of course societal needs and requirements. Sustainable development [8] is one of the important requirements. This natural requirement is based on the need for continual development in all fields of human activities. This development breaks unexpected fluctuations and crises introducing numerous difficult conditions into people's lives.

II. URGENT SUPPORT FOR SUSTAINABLE DEVELOPMENT

Sustainable development is not merely a modern term from science. Sustainable development is a crucial condition for continuous and solid evolution in all disciplines of global information society. This requirement slowly provides, via step by step development of mankind without crises, with the support of collective knowledge. Over just a few years, the needed goals are defined in detail [9]:

- Affordable and clean energy.
- Clean water and sanitation.
- Climate action.
- Decent work and economic growth.
- Gender quality.
- Good health and well-being.
- Industry, innovation and infrastructure.
- Life and land.
- Life below water.
- No poverty.
- Partnership for the goals.
- Peace, justice and strong institutions.
- Quality education.
- Reduced inequalities.
- Responsible consumption and production.
- Sustainable cities and communities.
- Zero hunger.

The format and urgency of the above-mentioned aims vary in terms of locality and actual needs, but all members of a global information society face difficult living conditions which reduce their standard of living. In Europe, there is a lack of jobs which brings many other limits in connection with economic growth, the level of family budgets, and the availability of adequate housing with a suitable role in society. Global unemployment has increased from 170 million (for the year 2007) to nearly 202 million (for 2012). This poverty eradication needs a stable and well-paid job. [10] In the future, 470 million jobs are needed globally between 2016 and 2030. [11]

The primary key for a solution to these weaknesses is an interest in skill development, optimal education and an active life. [12] Total resignation on the part of people is one of the ineffective approaches when solving existing difficulties. How can one break total resignation in people for an optimal role in society? One verified approach is to rely on science and innovations which may prevent such events. Modern global

information society needs:

- Higher levels of economic productivity. [13]
This aim uses diversification, modern technologies and innovation with a focus on sectors with added value.
- High creativity based on collaboration. [14]
Business activities create other jobs with a link to innovation.
- A stable background on sustainable consumption and production. [15]
Harmony between the environment, natural resources and innovation activities of mankind is the basis for further development.
- Stable employment and work for all. [16]
Innovations and IT development provide help for women, men, young people and people with disabilities.
- Stable education and training. [17]
Innovations and new knowledge support job creation for people.

In all the above-mentioned societal needs, the word “innovation” is the focus of interest. Innovation may provide a new view of given reality, new approaches and methods, and new products and services. Suitable innovation [18] is based on knowledge and optimal skills. Science and management science plays a key role for suitable composition of work with analyzed data. Science is not merely for educators and scientists. Science has to live with society and all people. Science has an excellent assumption for this task with links to questions about existing reality, simulations for monitoring the behavior of variable objects, combinations of unexpected events, or dissemination of new knowledge and information.

For these aims, science works with data stored in various sources. There are also numerous disciplines for optimal work with data such as Business Intelligence. [19] There is an increased need for responsible work. Existing reality is more complex with various relations (well-known and hidden). The main focus is on searching for an unexpected relation between stored data. Reason detects previously unknown connection and knowledge [20] with suitable development support.

Responsible work with data has to compare analysed data with numerous relations and therefore there are various intelligences:

- Artificial Intelligence.
- Business Intelligence.
- Competitive Intelligence.
- Computational Intelligence.
- Customer Intelligence.
- Swarm Intelligence.

Data processing is not an intuitive issue. Optimal work is based on principles of extensive data and data science, and these approaches have to be used on all known intelligences for suitable development. It is simple to state that information technology has enough resources for quality support of implemented activities with data. There are number of IT

products (systems, applications, or tools) on the Internet. Every IT user can select by quality priorities, available skills, financial sources and available time. The positive benefit is that IT development is quick with global support for interested teams, developers, analysts and fans. IT users work with various applications and products of information technology, but satisfaction with IT products is problematic. [21] IT users generate various data in numerous analyses, but there are mistakes and errors which cause difficult orientation in the global complex of reality.

III. DATA ANALYSIS AND EXISTING ERRORS

Errors, mistakes, faults or bugs are often dreaded words in implemented activities. We all wish to avoid such dangerous situations bringing unexpected difficulties. It is easy to say “be careful”, “carry out verification”, or “think about feedback” in every step (or phase) of an adopted method. Established methodologies and methods [22] help with an optimal approach to data analysis but there are still errors and difficulties. At first glance, errors are the result of an irresponsible approach to a given activity on the level of:

- Irresponsible Communication.
- Misunderstanding of importance.
- Oversight.
- Underestimating the speed of the ongoing changes.
- Wrong timing.

This list only creates a fragment of the well-known errors and mistakes. Every well verified method and methodology declares available benefits with needed error resiliency. Errors are feared for all solvers of data analyses and needed IT projects but also contribute to better results based on an understanding of the current situation and existing barriers.

For data analysis, measurement errors are ubiquitous. Optimal work with data needs a suitable perception of reality. There is sound advice in this issue:

- Comparing only adequate data. [23]

Sound results provide data segmentation and existing trends in selected period such as the year, six or three months. The required innovation provides an unexpected comparison of various segments, but they must be suitably calibrated. Certain values may tend to duplicities, double counting, or are defined in a confusing way in relations between them.

- Only working with clear graphs. [24]

Graph creation is a task on the basic level of work with information technology. There are numerous IT products for supporting graph creation with needed titles, axes, descriptions and trends. A suitable metric has an important place for a sound orientation in existing changes with relation to defined reasons. The default interest is about performance. Certain users do not have a preference for pie graphs that have no clear influence on how people think. It is difficult to say “it is good” or “it is bad”.

- Setting individual axes (such as time) optimally. [25]

Monitoring performance and existing changes on a day to

day basis is a problem. In this period, it is difficult to define the main changes and find an optimal reaction to them. Good advice involves beginning from a larger period such as months (or weeks) and consequently going into detail in individual weeks and days.

- Selecting a suitable color for the presented data. [26]

Visualization is a great help in data analysis. A large volume of data needs visualization support for improved data interpretation. Default reports with thousands and thousands of rows are hard to read and interpret. Sound results have to be marked with positive colors (green, blue) and bad results with alarming colors (red, yellow). This perception of colors is natural for people.

- Having an optimal data volume. [27]

A large data volume brings difficulties in optimal work for a further prediction of other trends. A small volume of data also brings great difficulties. Data analysis with limited data is irrelevant and such conclusions are inadequate. A better understanding of the adopted initial strategy is of some possible benefit for these cases.

- Repeating important results. [28]

The volume of data and information crosses the borders of people's perception. It is suitable to remember the repetition of important results because people only remember the interesting aspects. One tends to overlook the vast amount of information, graphs, tables and rows with data. It is difficult to find one's way. Important data has to be repeated in various formats with a focus on existing changes.

- Using suitable metrics. [29]

There are verified metrics providing additional benefits than ones in various dashboards, graphs and analysis. These metrics are default average, standard percentages, various ratios and calculated metrics. Suitable metrics provide a better understanding of selected objects in existing reality.

Responsible work with data and implemented analysis has to respect the dynamism of a global information society. Modern society creates numerous changes, but also advice and benefits. For a good requirement, data science has an important place in any work with data. Data science is science concerning data with a needed complexity. The level of required complexity will be higher and higher. For a better mastering of complexity, optimal management science and implemented work play an important role.

IV. A COMPLEX VIEW ON MANAGEMENT OF DATA SCIENCE

Knowledge is attractive for people at all times. He (she) who has knowledge has a solid background for optimal decision-making. [30] One of the verified sources of knowledge is experience from practical work. For data science, these experiences are also very important. This is not enough in modern information and a global society. We need suitable skills for collaboration, communication and also other skills by preferences from practice such as curiosity about hidden trends, a desire for learning new innovations, and

a hunger for unexpected changes.

Additional difficulties are caused by rapid developments in all disciplines of human activities. Information technology brings many new and unexpected solutions based on IT user preferences. Data science needs suitable knowledge from [31]:

- Analytics.
- Applications.
- Computer science.
- Math.
- Simulations.
- Statistics.

Data science currently seeks out connections between data from various sources based on number approaches. This is difficult for little teams. The questions concern the needed timing, quality, details, and also available knowledge with skills. Based on solid requests, there are trends to use various intelligences (Artificial, Business, Computational, Customer and Swarm). Such a wide spectrum of views is needed for optimal specification as to what data means and a recommendation of a method for working with data.

Active intelligence connects extensive data and data science, knowledge and intelligences with relation on suitable development. This approach must use wide spectrum verified methods and available skills from various views on existing reality. This way respects great global complexity. Aim is looking more closely at requires of information society. Key is interactive and active approach to solve defined tasks. There are many available methods for active solution these tasks, but only some lead to optimize or maximize benefits. Selection of the best method may relay on simulation with active monitoring predictive behavior.

By practice experiences, important assumptions for considered model are based on activation of intelligence that influences:

- Combination of achieved results with big data and data science.
- Interconnection various data sources with experienced knowledge from practice and realized analysis.
- Specification of an inside and outside view on risks and opportunities.
- Knowledge management for needed support active intelligence.
- Better understanding ongoing activities in defined environment with optimal reaction on specific conditions.
- Balancing knowledge (development, identification and application) with all available sources for suitable decisions.

These assumptions create limits for design of model with aim to search unexpected information. This model is oriented on analysis to description of future behavior with forecasting probabilities and trends. In time, predictive analysis relies on data, analysis, monitoring and forecasting. Standard question are “What happened and why?”, “What is happening now?”

and “Happen in future?” [32] These questions are still just interesting, but there are a new relations and requirements on available answers with support 4D model.

V. 4D MODEL FOR UNEXPECTED INFORMATION SEARCHING

Searching for unexpected information and connections is essential work with data and available information based on suitable knowledge and skills. Many cases are based on practice and a wide spectrum of views is highly valued with optimal support visualization and a clear description.

IT products provide numerous applications and tools for these goals but working with them needs a suitable background. Standard work is focused on various applications in terms of IT user interest and preferences. The question concerns an optimal composition (management) for implemented work for a better description of reality at the needed level of detail. A natural characteristic is the ability to move between individual levels. For many solutions, an optimal approach is based on abstraction and objects which provide a needed synergy between data and adopted methods. For complex work with extensive data and data science, the following model has been designed with four objects like the dimensions of a 4D model:

- Layers.

Layers create areas such as computer science, applications, simulations, statistics, analytics and math. This dimension is needed for an active use of data science.

- Intelligences.

Intelligences create available intelligences such as Artificial, Business, Computational, Customer and Swarm. This dimension enriches the solution about unexpected relationships between data.

- Fields.

Fields create disciplines where our 4D model is applied such as business, finance or informatics. This dimension defines existing specifics based on demands from practice.

- Zooms.

Zooms create an adopted level of research for a solution of selected problems by defined interests. This dimension creates the needed zoom for selected data. It works like filters.

Designed 4D model (4DM) is well described via the following relation:

$$4DM = (L_i, I_i, F_i, Z_i), \quad (1)$$

where L_i and I_i represent layers and intelligences, F_i is dedicated to disciplines (fields), and Z_i is oriented on available zooms, for $i = 1, \dots, n$. Implementation such 4D model relays on a sequence of executed works and applied applications (tools). Unique advantage must be unlimited switching between individual dimensions (layers, intelligences, fields and zooms) based on needed details according to dynamic changes in global information society with support of active intelligence.

In many cases, there is important to create relations (expected and unexpected) between analyzed data. There is

good knowledge about default expected relations, but unexpected information and its searching needs inspiration via available dimensions (layers, intelligences and fields). Considered switching between dimensions will use well-known processes like abstraction and specification. Adopted level of work will respect range of solution based on defined

aims, existing needs for given discipline (field), requirement on time and finances, or also skills of individual members of team.

For more information, please see Fig. 1.

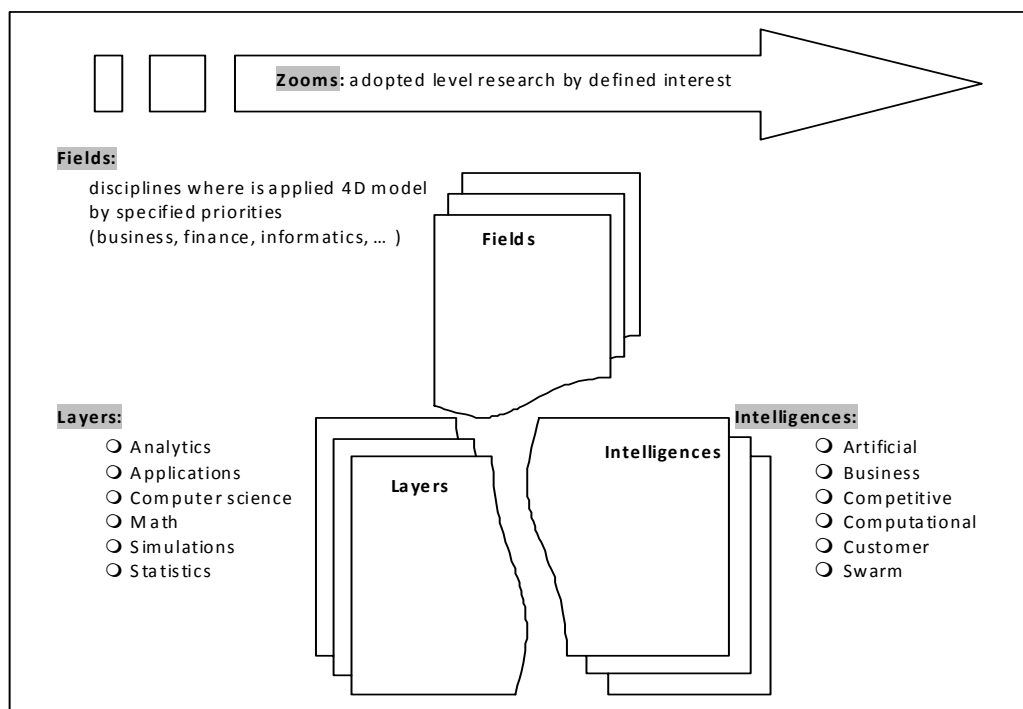


Fig. 1: 4D model for complex work with data and data science in a global information society

Selection needed application is easy task. Information technology offers many applications, systems and tools for active support executed works. Benefit is that this software is available on the Internet with support blogs and discussion groups for other advices and recommendations. Optimal combination will relay on preferences of research team.

Designed model brings difficulty that is oriented on optimal sequence of executed works via individual dimensions (layers, intelligences, fields and zooms). From common view, helper may be swarm intelligence with pheromones. For example, ants define optimal way to food based on pheromones. In designed model, food is information with knowledge oriented on description of future behavior with forecasting probabilities and trends. Optimal way is created adopted sequence of works based on “pheromones”.

There are also other analyses and studies that are focused on multidimensional models by interest:

- Visualization in projects. [33]

The 4D model provides benefits for participants such as architects, designers and contractors to clients. They highly evaluate the visualization of the needed activities and events in the implemented project via its lifetime.

- Project management. [34]

The aim is a discussion of strategies of representation and analyses of the 4D model. This model is focused on planning, controlling and evaluating the importance and applicability of professional vision in industry.

- Information technology tool. [35]

The information is generated based on IT tools in the project. These tools are commonly used in disciplines focused on details. Sound experience provides work at several levels in light of the variety of interfaces such as desktop, web, or virtual reality.

The common denominator for the above-mentioned analysis is an interest in links for all of the needed areas by preferences of teams with support for an optimal reaction to outgoing changes, specified priorities and indicators. The default interest is focused on a quick reaction of the highest quality. These requirements need an optimal user interface for work at various levels and details. In order to practice implementation of this designed model, further work will be focused on the creation of an optimal relation based on the defined indicators. Sound inspiration will swarm intelligence using pheromones

such as indicators to determine an optimal way to solve the existing problems.

VI. CONCLUSION

The complexity of existing reality is one of the difficult complications for an analysis of existing reality. The situation is hampered by the vast volume of data and dynamic changes in global information society. This entropy has a negative influence on implemented activities, processes and human lives. Effective protection is based on continuous monitoring and an actualization of specified preferences in connection with implemented processes and support for sustainable development.

Suitable support for sustainable development needs a thorough view of data with an aim of transferring data on information and knowledge. Seeking out unexpected relations, information and knowledge is of interest for information technology products. There are verified methods for active work with extensive data and data science, although errors and difficulties still exist. In order to avoid unnecessary errors and mistakes, a 4D model for work with data has been designed. This model connects four dimensions such as layers, fields, intelligences and zooms. Layers create the needed areas for data science such as computer science, applications, simulations, statistics, analytics and math. The fields are areas of implementation of the designed model such as business, finance and informatics. Intelligences create known intelligences such as Artificial, Business, Computational, Customer and Swarm. Zooms create needed filters, (focus) of interest, for defined preferences of analysts, designers, developers, managers and other IT users. The benefit is based on collaboration between various views on data with answers on the effect on global information society.

REFERENCES

- [1] *Software Development Methodologies*. (2015). [Online]. Available: <http://project-management.com/software-development-methodologies/>
- [2] N. A. Qureshi, U. Ghani, *Knowledge Management: The Way to Organizational Competitive Advantage*. In: *Oeconomics of Knowledge*, Vol. 7, Issue 1. (2015), pp. 25-35.
- [3] B. Marr, *Big Data: Using SMART Big Data, Analytics and Metrics To Make Better Decisions and Improve Performance*. Wiley, 2015. ISBN-10: 1118965833.
- [4] L. Pierson, *Data Science For Dummies*. For Dummies, 2015. ISBN-10: 1118841557.
- [5] G. S. Linoff, *Data Analysis Using SQL and Excel*. Wiley, 2015. ISBN-10: 111902143X.
- [6] D. J. Eich, *Innovation Step-by-Step: How to Create and Develop Ideas for your Challenge*. CreateSpace Independent Publishing Platform, 2014. ISBN-10: 149928747X.
- [7] Ch. Pappas, *9 Tips To Apply Adult Learning Theory to eLearning*. (2014). [Online]. Available: <http://elearningindustry.com/9-tips-apply-adult-learning-theory-to-elearning>
- [8] J. D. Sachs, *The Age of Sustainable Development*. Columbia University Press, 2015. ISBN-10: 0231173156.
- [9] *Sustainable Development Goals*. (2015, September). [Online]. Available: <http://www.un.org/sustainabledevelopment/sustainable-development-goals/>
- [10] *Global Employment Trends 2014: Risk of a jobless recovery?* International Labour Organization, Geneva, 2014. ISSN 2304-2893.
- [11] *Growth and employment, key to the post-2015 development agenda*. (2013). [Online]. Available: http://www.ilo.org/global/about-the-ilo/media-centre/press-releases/WCMS_222713/lang-en/index.htm
- [12] C. Rusbult, *Effective Learning Skills*. (2011). [Online]. Available: <http://www.asa3.org/ASA/education/learn/203.htm>
- [13] B. van Ark, S. K. Kuipers, G. H. Kuper (Editors), *Productivity, Technology and Economic Growth*. Springer; Softcover reprint of hardcover 1st ed. 2000 edition, 2010. ISBN-10: 1441950036.
- [14] Ch. Jarrett, *9 Facts Every Creative Needs to Know About Collaborative Teams*. (2013). [Online]. Available: <http://99u.com/articles/16850/everything-youve-ever-wanted-to-know-about-teams>
- [15] *Sustainable Consumption & Production: Making the Connection*. [Online]. Available: <http://www.unep.fr/shared/publications/cdrom/WEBx0028xPA/contents/InWEnt%20%28CP-SC%20Jan04%29/BackgroundPaper.pdf>
- [16] T. Heinilä, *Sitra Trends: stable employment becoming a thing of the past*. (2014). [Online]. Available: <http://www.sitra.fi/en/artikkelit/future/sitra-trends-stable-employment-becoming-thing-past>
- [17] L. Unwin, J. Wellington, *Young People's Perspectives on Education, Training and Employment: Realizing Their Potential*. Stylus Publishing, 2001. ISBN: 0-7494-3122-9.
- [18] L. Frishberg, Ch. Lambdin, *Presumptive Design: Design Provocations for Innovation*. Morgan Kaufmann, 2015. ISBN-10: 0128030860.
- [19] R. Sherman, *Business Intelligence Guidebook: From Data Integration to Analytics*. Morgan Kaufmann, 2014. ISBN-10: 012411461X.
- [20] K. King, S. McGrath, *Knowledge for Development?* Zed Books, 2013. ISBN: 1842773259.
- [21] J. James, *10 classic mistakes that plague software development projects*. (2012). [Online]. Available: <http://www.techrepublic.com/blog/10-things/10-classic-mistakes-that-plague-software-development-projects/>
- [22] L. Greiner, *What is Data Analysis and Data Mining?* (2011). [Online]. Available: <http://www.dbta.com/Editorial/Trends-and-Applications/What-is-Data-Analysis-and-Data-Mining-73503.aspx>
- [23] A. Kaushik, *Data Analysis 101: Seven Simple Mistakes That Limit Your Salary*. (2011). [Online]. Available: <http://www.kaushik.net/avinash/data-analysis-101-seven-simple-mistakes/>
- [24] J. H. McDonald, *Guide to fairly good graphs*. (2014). [Online]. Available: <http://www.biostathandbook.com/graph.html>
- [25] G. Soskey, *10 Design Tips to Create Beautiful Excel Charts and Graphs*. (2015). [Online]. Available: <http://blog.hubspot.com/marketing/excel-graph-tricks-list>
- [26] M. Stone, *Expert Color Choices for Presenting Data*. [Online]. Available: <http://www.stonesc.com/pubs/Expert%20Color%20Choices.pdf>
- [27] M. Gershoff, *Big Data or Big Distraction*. (2013). [Online]. Available: <http://conductrics.com/big-data-or-big-distraction/>
- [28] H. Nordmark, *How not to get 'confounded' by data science*. (2015). [Online]. Available: <http://www.itproportal.com/2015/05/12/how-not-to-get-confounded-by-data-science/>
- [29] G. Halevi, L. Colledge, *Standardizing research metrics and indicators – Perspectives & approaches*. (2014). [Online]. Available: <http://www.researchtrends.com/issue-39-december-2014/standardizing-research-metrics-and-indicators/>
- [30] M. J. Kochenderfer, *Decision Making Under Uncertainty: Theory and Application* (MIT Lincoln Laboratory Series). The MIT Press, 2015. ISBN-10: 0262029251.
- [31] *What is a data scientist?* (2015). [Online]. Available: <http://www-01.ibm.com/software/data/infosphere/data-scientist/>
- [32] *Predictive Modeling*. [Online]. Available: <http://www.predictiveanalyticstoday.com/predictive-modeling/>
- [33] F. Castronovo, S. Lee, D. Nikolic, J. Messner, *Visualization in 4D Construction Management Software: A Review of Standards and Guidelines*. In: *Proceedings of the International Conference on Computing in Civil and Building Engineering*. Orlando USA. (2014), pp. 315–322.
- [34] S. Chen, F.H. Griffis, P. Chen, L. Chang, *A Framework for an Automated and Integrated Project Scheduling and Management System*. *Automation in Construction*, 35 (2013), pp. 89–110.

- [35] B. Schwegler, M. Fischer, K. Liston, *New Information Technology Tools Enable Productivity Improvements*. In: North American Steel Construction Conference, AISC. Las Vegas. (2000), pp. 1-19.