

Methodology for risk assessment and corresponding costs in e-government projects

Petra Peharda, Dušan Mundar, Neven Vrček

Faculty of Organization and Informatics

University of Zagreb

Pavlinska 2, 42000 Varaždin, Croatia

{petra.peharda, dusan.mundjar, neven.vrcek}@foi.hr

Abstract. *The main objective of this paper is to introduce methodology for risk assessment and corresponding costs in e-government projects. The methodology is based on application of Bayesian networks. In order to connect risks with e-government service value, typical Bayesian networks based on probabilities were upgraded with costs. A method for calculations of cost estimation for every node (risk) in the network and the project as hole will be introduced and shown on an example. The paper concludes with brief overview on presented methodology and limitations.*

Keywords. Bayesian network, risks, e-government

1 Introduction

E-government is a very important public service tool and its implementation will become imperative in near future. Although very complex, range of its outcomes is very wide with significant impact on governmental agencies and citizens. In order to avoid cancelling or failure, planning of e-government project should not be a trivial task. In phase of planning, among other things, risks have to be very thoroughly considered.

The main objective of this article is to introduce methodology for risk assessment and corresponding costs in e-government projects based on Bayesian network.

At the beginning of this paper state of the art is given. Further, brief overview of Bayesian network, risk management and e-government is given and, finally, methodology for risk assessment and corresponding costs in e-government projects is introduced. The paper concludes with brief overview on presented methodology and limitations.

2 State of the art

Application of Bayesian networks in risk management has been widely studied until today in different context. Although to our knowledge the method was not used in context of e-government.

In [10] authors presented methodology for building an information technology (IT) implementation BN from client-server survey data. Further, authors also demonstrated how to use the BN to predict the attainment of IT benefits, given specific implementation characteristics and activities.

In [14] is presented a scheme for large engineering project risk management using a Bayesian belief network and applies it to the Korean shipbuilding industry. Twenty-six different risks were deduced from expert interviews and a literature review. Results of study demonstrate the difference of risks between large-scale and medium-sized shipbuilding companies, and the relationships among the risk items.

Authors in [4] shown a possible approach for building a BN in the particular case in which only prior probabilities of node states and marginal correlations between nodes are available, and when the variables have only two states.

In [11] authors presented a scheme to incorporate BBNs (Bayesian belief networks) in software project risk management. They defined a theoretical model to provide insights into risk management and based on these insights, they have developed a BBN-based procedure using a feedback loop to predict potential risks, identify sources of risks, and advise dynamic resource adjustment. This approach facilitates the visibility and repeatability of the decision-making process of risk management.

In [18] authors combine Bayesian networks (BNs) and structural reliability methods (SRMs) to create a

new computational framework, termed enhanced BN (eBN), for reliability and risk analysis of engineering structures and infrastructure. Further, in [19] authors present the application of the eBN: to the assessment of the life-cycle reliability of a structural system; to the optimization of a decision on performing measurements in that structural system; and to the risk assessment of an infrastructure system subject to natural hazards and deterioration of constituent structures.

3 Bayesian networks

Bayesian network, also called belief network is a powerful tool for knowledge representation and reasoning under conditions of uncertainty [7]. A Bayesian network is probabilistic graphical model. Its structure is defined by set of nodes and set of directed edges - arcs. Nodes represent variables of interest while arcs represent conditional dependencies between nodes, i.e. causal connections. This part of Bayesian network - graphical model - represents qualitative part.

Each node of Bayesian network can be in various states and it is not limited to two states [2]. States are values that variables can take and the number of states is selected by the risk analyst. Conditional dependencies in the graph are often estimated by using known statistical and computational methods. Hence, BNs combine principles from graph theory, probability theory, computer science, and statistics. [3]. The quantitative part of a Bayesian network, the so-called parameter learning, finds dependence relations as joint conditional probability distributions among variables using cause and consequence relationships from the qualitative part and data of variables [14]. Conditional probability represents the chance that one event will occur given that a second event has already occurred. The probability of any node in the Bayesian belief network being in one state or another without current evidence is described using a *conditional probability table* [6]. Since quantitative analysis requires specification of conditional probabilities, for each transition from one node to another Conditional Probability Table has to be defined. Prior probabilities for all root nodes are estimated while posteriori probabilities for all other nodes are calculated according to Bayes' formula.

$$P(A|B) = \frac{P(B|A) P(A)}{P(B)} \quad (1)$$

where

$P(A)$ – prior probability of A (it does not take into account any information about B)

$P(A|B)$ – conditional probability of A , given B

$P(B|A)$ – conditional probability of B , given A

$P(B)$ – prior probability of B

4 Risk management in e-Government projects

Authors differently define e-government, although in scope of all of them is *using of information and communication technology (ICT) in order to improve services of public sector organizations, and finally to improve interaction between government and citizens*. The purpose of e-government initiative is to extend and provide new opportunities to citizens “putting them on-line instead of in line”. These initiatives are very complex and their implementations take time, but in near future will become imperative. Implementation of e-government leads to some improvements such as: (1) Improvement of collaboration between citizens and government agencies, (2) Improvement of citizens quality of life and quality of public sector services, (3) Reduction in duplication of efforts and costs through provision of products and services electronically, (4) Enhancement of efficiency and effectiveness of public sector services, (5) Higher availability and easier accessibility to services and information 24 hours a day seven days in week. E-government services have their value that is not recognized only in Return On Investment (ROI), but also in some other factors. According to [5] there are six essential factors that have to be measured in order to understand and capture the value of e-service fully. These factors are: (1) direct user value, (2) social value, (3) government financial value, (4) government operational value, (5) strategic/political value and (6) risk. The study has found out that the sixth factor, perceived risk, is a crucial determinant that decreases e-government service value [17]. The risk associated with an investment in an e-government initiative may degrade performance, impede implementation, and/or increase costs [8].

There are different methodologies developed for measuring the value of e-service. Below we will give brief overview on three of them:

(1) Value Measuring Methodology was first articulated in [5] in 2002 for the US Social Security Administration as part of an electronic services project. The purpose of the Value Measuring Methodology (VMM) is to define, capture, and measure values associated with electronic services unaccounted for in traditional Return-on-Investment (ROI) calculations, to fully account for costs, and to identify and consider risk [8].

(2) German WiBe methodology originally developed 1992 by Dr. Röthig from WiBe-TEAM PR as result of a consulting order for the Ministry of the Interior. WiBe is one of the first frameworks for assessment of economic efficiency of federal administrations [20].

(3) MAREVA methodology developed in 2004 by the French ADAE (Electronic Administration Development Agency). MAREVA, methodology of value analysis and return on investment, has been developed for any administration that wishes to better manage its transformation projects portfolio [1].

One of the most important parts of project management is risk management and it must be done during the whole life of the project [15]. According to [13], risk management include planning for risk, assessing (identifying and analyzing) risk issues, developing risk handling strategies, and monitoring risks to determine how they have changed. Below we will give brief overview on risk assessment – risk identification and risk analyzing.

[16] define risk as a discrete occurrence that may affect the project for better or worse and it consists of two components: probability and impact on the project. Accordingly, risk can be written as function of these two variables $risk=f(probability, impact)$.

Authors differently classify risks in e-Government projects. According to VMM it is necessary to identify risks in eight categories: organizational and change management, business, data and information, technical, strategic, security, privacy and project. According to [1] project risks, technical risks, legal risks, deployment risks. According to [9]: societal, technical, economical, political and security risks.

In order to identify risks in e-government project we observed challenges to e-government initiative identified by authors in [12] who reviewed the literature based on e-government. Authors in [12] grouped challenges into five categories according to their core aspect: (1) information and data, (2) information technology, (3) organizational and managerial, (4) legal and regulatory, and (5) institutional and environmental.

5 Methodology for risk assessment and corresponding costs in e-government projects

As authors in [5] specified, one of six essential factors which has to be measured in order to understand and capture the value of e-service fully, is risk. In this article we will present methodology for risk and service value assessment in e-government projects. The methodology is based on application of Bayesian networks. In order to connect risks with e-government service value, typical Bayesian networks based on probabilities were upgraded with costs. A method for calculations of cost estimation for every node (risk) in the network and the project as whole will be introduced. The purpose of this methodology is to assess probability of risk occurrence and corresponding costs in e-government initiatives.

There are several steps in this methodology as follows: (1) Risk identification, (2) Risk categorization, (3) Identification of risk dependencies (formation of Bayesian network), (4) Determination of a-priori probability table associated to each root node in the network, (5) Costs estimation for each risk in the network, (6) Determination of Conditional Probability Table upgraded with costs associated to each transition from one node to another root node in the network (probabilities estimation and costs calculation), (7) Calculation of a-posteriori probability table for each not-root node and not-summary node in the network according to Bayes' formula, (8) Calculation of probability for each summary node in the network according to slightly modified Bayes' formula, (9) Calculation of costs for each not-root node in the network according to slightly modified Bayes' formula.

5.1 Risk identification and categorization

There are different methods for risk identification. In order to identify risks in e-government initiatives we have used challenges for e-government initiatives identified in [12]. Authors in [12] reviewed current literature in information systems research in order to identify factors found to influence the success of IT initiatives. The primary challenges for e-government initiatives were grouped into five categories according to their core aspect: (1) information and data, (2) information technology, (3) organizational and managerial, (4) legal and regulatory, and (5) institutional and environmental. In our case, these challenge categories are considered as risk categories. Further, each challenge or its absence is considered as potential risk. In Table 1. are represented all identified challenges transformed into risks in e-government initiative. Further, for every identified risk was estimated cost of its occurrence which is also represented in Table 1.

Although the most widely used scale in risk assessment is *low-medium-high*, each node in this study can have only two states, (0) - *the risk does not appear* and (1) - *the risk appear*, for easy understanding of the methodology and calculation.

Information and data risks	Information technology risks	Organizational and managerial risks
Information and data quality 3000	Usability 8000	Project size 2000
Dynamic information needs 2000	Security issues 6500	Manager's attitudes and behavior 4000
	Technological incompatibility 2500	Users or organizational diversity 5000
	Technology complexity 8000	Lack of alignment of organizational goals and project 4000
	Technical skills and experience 4500	Multiple or conflicting goals 3500
	Technology newness 2500	Resistance to change 3500
		Turf and conflicts 2500

Legal and regulatory risks	Institutional and environmental risks
Restrictive laws and regulations 4500	Privacy concerns 2000
One year budgets 3500	Autonomy of agencies 2500
Intergovernmental relationships 3000	Policy and political pressures 4500
	Environmental context (social, economic, demographic) 3000

Table 1. Risks in e-government project and estimated costs

5.2 Definition of risk dependencies

After risk identification, follows phase of defining dependencies between risks in each category. The resulting structure has a structure of a mathematical graph, so called Bayesian network where each node represents risk while arc represents influence of one risk on another. The Bayesian network of risks in e-government project is shown in Figure 1. There are 22 variables – risks in the network and 11 of them are root nodes: Dynamic information needs, Technology newness, Technology complexity, Technical skills

and experience, Project size, Users or organizational diversity, Manager's attitudes and behaviour, Lack of alignment of organizational goals and project, Restrictive laws and regulations, Intergovernmental relationships and Autonomy of agencies. Risk categories (*Information and data risks, Information technology risks, Organizational and managerial risks, Legal regulatory risks and Institutional and environmental risks*) represent summary nodes of all risks from each category while *Project Risk Assessment* represent final summary node of all risk categories.

5.3 A-priori Probability Table

A-priori Probability Table has to be defined for each root node in the network. In a-priori Probability Table are defined probabilities for *no appearance of the risk (0)* and *appearance of the risk (1)*. The sum of all possible states for each risk is equal 1, that is if probability of risk appearance is p than probability of risk no appearance is $1 - p$. In order to define a-priori probabilities it is recommended to consult corresponding literature and interview experts. Table 2 represents summary a-priori table for all root node in the Bayesian network of risks in e-government project.

	Dynamic information needs	Technology newness	Technology complexity
0	0,7	0,3	0,5
1	0,3	0,7	0,5

	Technical skills and experience	Project size	Users or organizational diversity
0	0,3	0,75	0,4
1	0,7	0,25	0,6

	Manager's attitudes and behaviour	Lack of alignment of organizational goals and project	Restrictive laws and regulations
0	0,3	0,5	0,4
1	0,7	0,5	0,6

	Intergovernmental relationships	Autonomy of agencies
0	0,5	0,2
1	0,5	0,8

Table 2. A-priori Probability Table

5.4 Costs estimation

The next step in the methodology is to estimate costs of risk occurrence. The following estimation, as presented is Table 2., is made from theoretical aspect and it does not represent real situation.

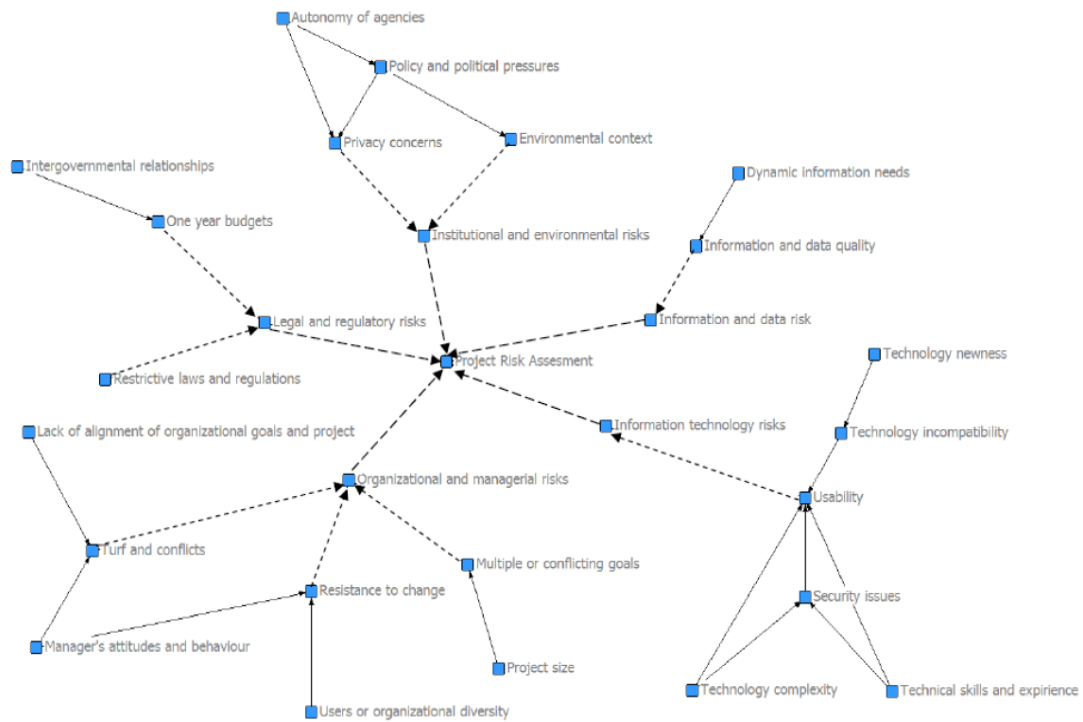


Figure 1. Bayesian network of risks in e-government projects

5.5 Conditional Probability Table

Conditional Probability Table has to be defined for each transition from one node to another, that is for every dependency between risks in the network. In Conditional Probability Table is defined what is probability that one risk in state (0) or (1) will cause another risk in state (0) or (1). Probabilities have to be defined for every possible scenario. The sum of (0) and (1) probabilities for every possible scenario in Conditional Probability Table is equal 1.

It is important to mention that in our case, conditional probability tables have been defined only for transitions between risks while relations between risks and risk categories have not been observed as dependencies. The same case is with relations between risk categories and node Project Risk Assessment.

5.6 A-posteriori Probability Table

After performing all of the previous steps, follows calculation of probabilities for each not-root node and not-summary node in the network. These probabilities are calculated according to previously described Bayes' formula (formula 1).

Example description.

Estimated probabilities for Dynamic information needs risk (DIN) are 0,7 for state (0) and 0,3 for state (1). This risk causes Information and data quality risk (IDQ). Let assume that occurrence of Dynamic information needs risk (state (1)) costs 3000 and occurrence Information and data quality risk (state (1)) costs 2000. If risk does not occur (state (0)) its cost is 0. Regarding previous assumption and previously estimated probabilities Conditional Probability Table can be defined. From Conditional Probability Table is evident that there is 0,9 probability that if node Dynamic information needs is in state (0) that node Information and data quality will also be in state (0). Cost of each scenario in Conditional Probability Table is equal to sum of costs in corresponding states of all observed nodes.

After that follows calculation the a-posteriori probability table for node Information and data quality. The probability that Information and data quality will be in state (0) is calculated according to Bayes' formula. That is,

$$\begin{aligned}
 P(IDQ = 0) &= \\
 &= P(IDQ = 0|DIN = 0) P(DIN = 0) \\
 &+ P(IDQ = 0|DIN = 1) P(DIN = 1) \\
 &= 0,9 \cdot 0,7 + 0,1 \cdot 0,3 = 0,66
 \end{aligned}$$

As we already mentioned, the probability that the node is in state (1) is

$$p(IDQ = 1) = 1 - p(IDQ = 0) = 0,34$$

For every node in the network we were calculating average cumulative costs. Average cumulative costs include probability pondered costs of node (risk) and all child-nodes. We were distinguishing those costs from specific costs of node (costs if risk represented by the node occurred). For example, average cumulative cost for node Dynamic information needs can be calculated according to the formula:

$$c_a = c_c(IDQ = 0) P(IDQ = 0) + c_c(IDQ = 1) P(IDQ = 1),$$

where

$$c_c(IDQ = 0) = (c(IDQ = 0) + c(IDN = 0)) \cdot P(IDQ = 0|DIN = 0) + (c(IDQ = 0) + c(IDN = 1)) \cdot P(IDQ = 0|DIN = 1).$$

5.7 Calculation of probability for summary nodes in the network according to slightly modified Bayes' formula

As we already mentioned, there are some specific nodes in the Bayesian network of risks in e-government project: risk categories and Project Risk Assessment. Those nodes represent summary nodes and their occurrence probabilities cannot be calculated according to Bayes' formula. There are two cases in regarding summary nodes as you can see in Figure 1:

1. *There is only one final node in sequence of risk dependencies in one risk category (Information and data risks, Information technology risks).* In this case, probabilities for summary nodes (Information and data risk and Information technology risk) are mapped from each state of final node in each risk category.
2. *There are more than one nodes in sequence of risk dependencies in one risk category (Institutional and environmental risks, Legal and regulatory risks, Organizational and managerial risks).* In this case, probability of minor risk problems in the category (state 0) is equal to product of probabilities for state 0 of all final nodes in risk category. We have major problems in category (state=1) if one of final nodes is in state 1.

In order to assess project risk it is necessary to consider probabilities of all risk categories as shown in Figure 3. According to previously described cases, this summary node (Project Risk Assessment (PRA)) belongs to second case and on this example will be

explained calculation (Figure 2). Calculation of average cumulative costs will be presented on final project risk assessment.

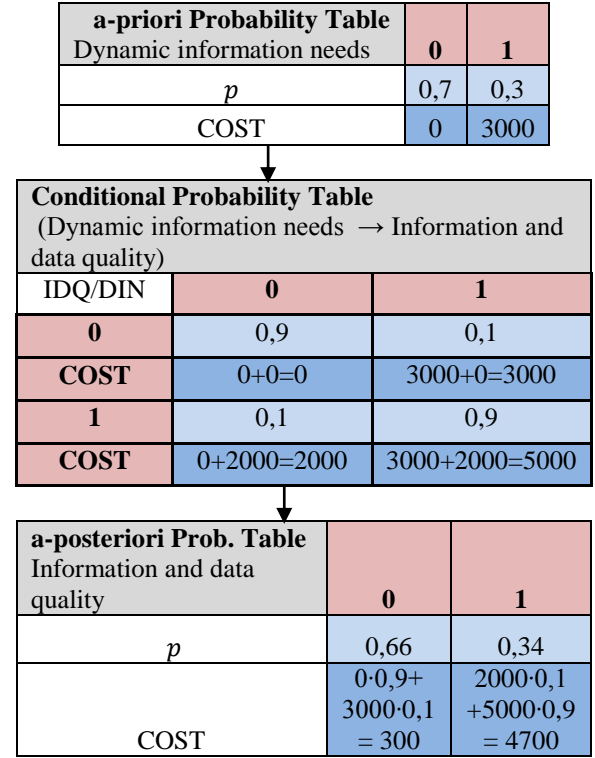


Figure 2. Example - calculation of a-posteriori probabilities and corresponding costs

Calculation:

First we will introduce notations:

$$p(r_1, r_2, r_3, r_4, r_5) = p(IDR = r_1) \cdot p(ITR = r_2) \cdot p(OMR = r_3) \cdot p(LRR = r_4) \cdot p(IER = r_5),$$

$$C_c(r_1, r_2, r_3, r_4, r_5) = c_c(IDR = r_1) + c_c(ITR = r_2) + c_c(OMR = r_3) + c_c(LRR = r_4) + c_c(IER = r_5)$$

Therefore, probability and average cumulative cost can be calculated by the formula

$$p(PRA = 0) = p(0,0,0,0,0) = 0,18 \cdot 0,2123 \cdot 0,66 \cdot 0,4052 \cdot 0,0568 = 0,00058$$

$$p(PRA = 1) = 1 - 0,00058 = 0,99942$$

$$c_c(IDQ = 0) = C_c(0,0,0,0,0) = 98.825,00$$

$$c_c(IDQ = 1) = \left(\sum_{(r_1, r_2, r_3, r_4, r_5, r_6)} C_c(r_1, r_2, r_3, r_4, r_5, r_6) p(r_1, r_2, r_3, r_4, r_5, r_6) - C_c(0,0,0,0,0,0) p(0,0,0,0,0,0) \right) \frac{1}{1 - p(PRA = 0)}$$

$r_i \in \{0,1\}$, for $i = 1,2,3,4,5$.

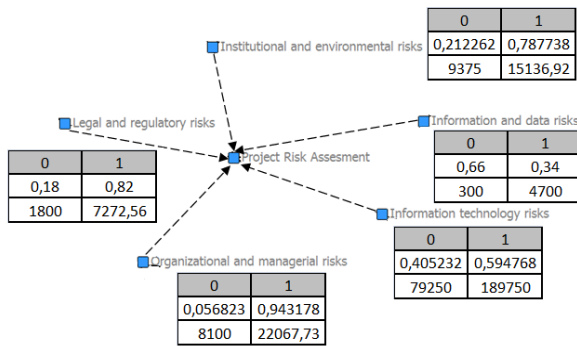


Figure 3. Integration of risk categories

Results of calculations are shortly presented in Table 3 (as we already mentioned, calculations are based on hypothetical inputs.) According to conducted calculation, there is very low probability (0,00058) that project has minor problems in all categories. Average cumulative costs in that case are equal 98.825,00. We defined that project has major problems if there is major problem (state=1) in any of five categories. Average cost in that case is 188.295,24 (Table 3). Further, from conducted calculation is evident that the greatest impact on overall project risk have organizational and managerial risks – the probability that this kind of risk will appear is 0,94318. Based on these results, organizational and managerial risks represent the greatest threat for the e-government project.

Project Risk Assessment	
0	1
0,00058	0,99942
98.825,00	188.295,24

Table 3. a-posteriori probability table upgraded with costs

5.8 Calculation of costs for summary nodes in the network according to slightly modified Bayes' formula

In order to calculate costs of *risk categories* and final node *Project Risk* in e-government project, not only summary nodes need special treatment. As you can see in Figure 1, there are some risks branching which means that one risk influences on two or more other risks. To avoid double costs calculation it is necessary to pay additional attention to these cases.

For example, Technology complexity risk influences on Security issues risk and Usability risk. Calculating average cumulative costs for Usability risk in usual way (using average cumulative costs for

all nodes influencing the node) would include cost of Technology complexity risk twice (similar situation with Technical skills and experience). To avoid that in calculation we used specific cost of Security issues risk (and not cumulative). Other similar situations (e.g. Autonomy of agencies) are solved in a similar way.

6 Conclusion

The methodology presented in this article, explains how to apply Bayesian networks in order to assess risk and costs in e-government projects. Typical Bayesian network based on causal probabilities was upgraded with costs. Further, used Bayes formula was slightly modified in order to apply it on combination of probability occurrence and costs.

Final results of applied methodology are costs and probability occurrence of every risk category and overall risk on the project. From final results it is easy to see which risk category is the greatest threat to the project so strategies for risk handling can be defined. In the end, it is important to mention limitations of presented methodology. For easy understanding of the methodology and calculation of probabilities and costs (1) only two risk states (occurrence and not occurrence) were considered and (2) only dependencies between risks of the same category were considered. For further research the methodology will be upgraded in a way that widely used scale (low-medium-high) in risk assessment will be used and dependencies between all possible risks will be considered. Finally, based on results from conducted methodology sensitivity analysis will be also conducted.

References

- [1] ADAE: MAREVA methodology guide: Analysis of the value of ADELE projects, http://ec.europa.eu/information_society/activities/egovernment/docs/pps/2_mareva_methodology_guide.ppt, Accessed: 12th April 2011.
- [2] Aven, T.: Risk Analysis: Assessing Uncertainties Beyond Expected Values and Probabilities, Wiley & Sons, 2008
- [3] Ben-Gal, I.: Bayesian Networks, in Ruggeri F., Faltin F. & Kenett R., Encyclopedia of Statistics in Quality & Reliability, Wiley & Sons, 2007.
- [4] Bonafede, C.E., Giudici, P.: Bayesian Networks for enterprise risk assessment, Physica A:

- Statistical Mechanics and its Applications, Volume 382, Issue 1, August 2007, pp 22-28
- [5] Booz Allen Hamilton: Building a methodology for measuring the value of e-service, http://www.idmanagement.gov/smartcard/information/measuring_finalreport.pdf, Accessed: 10th April 2011.
- [6] Charles River Analytics: About Bayesian Belief Networks, www.cra.com, Accessed: 12th April 2011.
- [7] Cheng, J., Greiner, R: Learning Bayesian Belief Network Classifiers: Algorithms and System, Lecture Notes in Computer Science, 2001, Volume 2056, 2001., pp 141-151
- [8] CIO Council, Best Practices Committee: Value Measuring Methodology, How To Guide, http://www.cio.gov/documents/ValueMeasuring_Methodology_HowToGuide_Oct_2002.pdf, Accessed: 10th April 2011.
- [9] Evangelidis, A.: FRAMES – A Risk Assessment Framework for e-Services
- [10] Eitel, J.M., Duchessi, L., Duchessi, P.J.: A methodology for developing Bayesian networks: An application to information technology (IT) implementation, European Journal of Operational Research, Volume 179, Issue 1, 2007., pp 234-252
- [11] Fan, C.F., Yu, Y.C.: BBN-based software project risk management, Journal of Systems and Software, Volume 73, Issue 2, October 2004, pp 193-203
- [12] Gil-Garci'a, J.R., Pardo, T.A.: E-government success factors: Mapping practical tools to theoretical foundations, Government Information Quarterly 22, 2005., pp 187 – 21
- [13] Kerzner, H.: Project Management: A Systems Approach to Planning, Scheduling, and Controlling, WILEY, 2009.
- [14] Lee, E., Park, Y., Shin, Y.G.: Large engineering project risk management using a Bayesian belief network, Expert Systems with Applications – An International Journal, Volume 36, Issue 3, 2009, Pages 5880-5887
- [15] Newell, M.W.: Preparing for the project management professional, Certification Exam, Second Edition, 2005.
- [16] PMI: A guide to the project management body of knowledge: PMBOK guide 3rd ed., USA, 2004.
- [17] Rotchanakitumnuai, S.: Measuring e-government service value with the E-GOVQUAL-RISK model, Business Process Management Journal, Vol. 14 Iss: 5, 2008., pp 724 – 737
- [18] Straub, D., Der Kiureghian, A.: Bayesian network enhanced with structural reliability methods: Methodology, Journal of Engineering Mechanics 136 (10), art. no. 007010QEM, 2010, pp 1248-1258
- [19] Straub, D., Der Kiureghian, A.: Bayesian network enhanced with structural reliability methods: Application, Journal of Engineering Mechanics 136 (10), art. no. 007010QEM, 2010., pp 1259-1270
- [20] WiBe 4.0 Recommendations on Economic Efficiency Assessment in the German Federal Administration, in Particular with Regard to the Use of Information Technology, <http://www.eu.wibe.de>, Accessed: 10th April 2011.