Stanisław Wrycza (Ed.)

Information Systems: Development, Learning, Security

LNBIP 161

6th SIGSAND/PLAIS EuroSymposium 2013
Gdańsk, Poland, September 2013
Proceedings

Springer

Lecture Notes
in Business Information Processing 161

Series Editors

Stanisław Wrycza (Ed.)

Information Systems: Development, Learning, Security

6th SIGSAND/PLAIS EuroSymposium 2013
Gdańsk, Poland, September 26, 2013
Proceedings

Springer

Volume Editor

Stanisław Wrycza
University of Gdańsk
Faculty of Management
Department of Business Informatics
Sopot, Poland
E-mail: swrycza@univ.gda.pl

ISSN 1865-1348 e-ISSN 1865-1356
ISBN 978-3-642-40854-0 e-ISBN 978-3-642-40855-7
DOI 10.1007/978-3-642-40855-7
Springer Heidelberg New York Dordrecht London

Library of Congress Control Number: 2013947537

Typesetting: Camera-ready by author, data conversion by Scientific Publishing Services, Chennai, India

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Preface

Systems analysis and design (SAND) has been the central field of research and education in the area of management information systems (MIS) or, as it is called more frequently in Europe, Business Informatics, almost from its origins. SAND continuously attracts the attention of both academia and business. The rapid progress of information and communications technology naturally generates the requirements for new generations of SAND methods, techniques, and tools. Therefore, international thematic conferences and symposia have become widely accepted forums for exchanging concepts, solutions, and experiences in SAND. In particular, the Association for Information Systems (AIS) is undertaking the initiative towards SAND's international development.

The objective of the EuroSymposium on Systems Analysis and Design is to promote and develop high quality research on all issues related to SAND. It provides a forum for SAND researchers and practitioners in Europe and beyond to interact, collaborate, and develop their field. The EuroSymposia were initiated by Prof. Keng Siau as the SIGSAND - Europe Initiative. Previous EuroSymposia were held at:

- University of Galway, Ireland – 2006
- University of Gdansk, Poland – 2007
- University of Marburg, Germany – 2008
- University of Gdansk, Poland – 2011
- University of Gdansk, Poland – 2012

The accepted submissions of EuroSymposia were published as:

- EuroSymposium 2007: A. Bajaj, S. Wrycza (eds), Systems Analysis and Design for Advanced Modeling Methods: Best Practises, Information Science Reference, IGI Global, Hershey, New York, 2009,
- EuroSymposium 2011: S. Wrycza (ed) 2011, Research in Systems Analysis and Design: Models and Methods, series: LNBIP 93, Springer, Berlin 2011,
- EuroSymposium 2012: Joint Working Conferences EMMSAD/ EuroSymposium in: I. Bider, T. Halpin, J. Krogstie, S. Nurcan, E. Proper, R. Schmidt, P. Soffer, S. Wrycza (eds.) 2012, Enterprise, Business-Process and Information Systems Modeling, series: LNBIP 113, Springer, Berlin 2012.

Three organizers of the 4th EuroSymposium on Systems Analysis and Design are as follows

- SIGSAND – AIS Special Interest Group on Systems Analysis and Design
- PLAIS – the Polish Chapter of AIS
- The Department of Business Informatics of University of Gdansk, Poland

SIGSAND is one of the most active AIS SIGs with a substantial record of contributions to AIS. It provides services such as the annual North American and European SAND Symposia, research and teaching tracks at major IS conferences, a listserver and special issues in journals.

The Polish Chapter of Association for Information Systems (PLAIS) was established in 2006 as the joint initiative of Prof. Claudia Loebbecke, former President of AIS and Prof. Stanislaw Wrycza, University of Gdansk, Poland. PLAIS co-organizes international and domestic IS conferences.

The Department of Business Informatics of the University of Gdansk is conducting intensive teaching and research activities. Some of its academic manuals are bestsellers in Poland. The department is also active internationally, organizing conferences including the 10th European Conference on Information Systems (ECIS 2002), The 7th International Conference on Perspectives in Business Informatics Research (BIR 2008), The 8th International Conference on European Distance and E-learning Network (EDEN 2009) and 24th Conference on Advance Information Systems Engineering (CAiSE 2012). The Department is the partner of the European Research Center for Information Systems consortium.

EuroSymposium 2013 had an acceptance rate of 40%, with submissions divided into the following three groups:

- Information Systems Development
- Information Systems Security
- Information Systems Learning

The accepted papers reflect the current trends in the field of systems analysis and design.

During EuroSymposium'2013 the following keynote speech was given:

Shmueli O., Fink L., Pliskin N., (Ben-Gurion University of the Negev; Israel), Explaining the Phenomenon of Over-Requirement in Software Development:Three Experiments Investigating Behavioral Effects

I would like to express my thanks to all authors, reviewers, advisory board, International Programme Committee and Organizational Committee members for their support, efforts and time. They have made possible another successful Systems Analysis and Design EuroSymposium.

August 2013 Stanislaw Wrycza

Organization

General Chair

Stanislaw Wrycza	University of Gdansk, Poland

Organizers

- AIS SIGSAND - Special Interest Group on Systems Analysis and Design of Association for Information Systems
- PLAIS - Polish Chapter of Association for Information Systems
- Department of Business Informatics, Faculty of Management of University of Gdansk

Advisory Board

Gove N. Allen	Brigham Young University, USA
David Avison	ESSEC Business School, France
Joerg Becker	University of Muenster, Germany
Dimitris Karagiannis	University of Vienna, Austria
Claudia Loebbecke	University of Cologne, Germany
Matti Rossi	Aalto University School of Business, Finland
Keng Siau	Missouri University of Science and Technology, USA

International Programme Committee

Eduard Babkin	Higher School of Economics, Russia
Akhilesh Bajaj	University of Tulsa, USA
Dinesh Batra	Florida International University, USA
Jan vom Brocke	University of Liechtenstein, Liechtenstein
Sven Carlsson	Lund University, Sweden
Vladislav V. Fomin	Vytautas Magnus University, Lithuania
Markus Helfert	Dublin City University, Ireland
Vijay Khatri	Indiana University Bloomington, USA
Marite Kirikova	Riga Technical University, Latvia
Andrzej Kobylinski	Warsaw School of Economics, Poland
Michał Kuciapski	University of Gdańsk, Poland
Miroslawa Lasek	University of Warsaw, Poland
Bogdan Lent	University of Applied Sciences, Switzerland

Leszek Maciaszek	Macquarie University Sydney, Australia and Wrocław University of Economics, Poland
Rolande Marciniak	Paris West University Nanterre La Defense, France
Bartosz Marcinkowski	University of Gdańsk, Poland
Saulius Maskeliunas	Vilnius University Institute of Mathematics and Informatics, Lithuania
Jacek Maślankowski	University of Gdańsk, Poland
Bjoern Erik Munkvold	University of Agder, Norway
Nava Pliskin	Ben-Gurion University of the Negev, Israel
Narcyz Roztocki	State University of New York at New Paltz, USA
Thomas Schuster	Forschungszentrum Informatik, Germany
Piotr Soja	Cracow University of Economics, Poland
Angelos Stefanidis	University of Glamorgan, UK
Reima Suomi	University of Turku, Finland
Iryna Zolotaryova	Kharkiv National University of Economics, Ukraine
Joze Zupancic	University of Maribor, Slovenia

Organizing Committee

Stanisław Wrycza (chairman)	University of Gdańsk, Poland
Łukasz Malon	University of Gdańsk, Poland
Anna Szynaka	University of Gdańsk, Poland
Bartosz Marcinkowski	University of Gdańsk, Poland
Jacek Maślankowski	University of Gdańsk, Poland
Michał Kuciapski	University of Gdańsk, Poland

Sigsand 2013 Topics

- Systems Theory and Principles
- Ontological Foundations of Systems Analysis and Design
- Philosophical Issues in Systems Analysis and Design
- Design Theory
- Requirements Engineering
- Business Process Modeling
- Conceptual Modeling
- Database Analysis and Design
- Information Systems Development Methods and Techniques
- Soft Systems Methodologies
- Object Oriented Techniques and Methodologies, Transitions from and to Related Approaches
- UML/SysML/BPMN

- Agility and MDA
- Rapid Systems Development
- Human-Computer Interaction and Prototyping
- Software Engineering
- Open Source Software (OSS) Solutions
- Ethical, Human and Organizational Aspects of IS Development
- ERP Systems
- Workflow Management
- Teaching Systems Analysis and Design
- Curriculum Design and Implementation Issues
- E-learning in Systems Analysis and Design
- Empirical Studies of Systems Analysis and Design Methods and Techniques
- SOA and Cloud Computing

Table of Contents

Chapter I - Information Systems Development

Rule Based Approach for Ensuring Consistency in Different UML Models 1
Diana Kalibatiene, Olegas Vasilecas, and Ruta Dubauskaite

Development of Requirements on the Role Business Analyst 17
Petr Doucek, Milos Maryska, and Lea Nedomova

ERP in Project-Driven Organizations: A Case-Study from IT Industry in Poland 27
Jan Trąbka and Piotr Soja

Chapter II - Information Systems Security

Specifying Security at the Systems Analysis Level: Operational, Public-Image and Legal Aspects 39
Firoozeh Rahimian and Akhilesh Bajaj

Identifying Factors of an Information Security Management System of Local Self-government Bodies 50
Przemyslaw Jatkiewicz

The Integration of Web-Based Information and the Structured Data in Data Warehousing 66
Jacek Maślankowski

Chapter III - Information Systems Learning

Matching Process Requirements to Learning Objectives 76
Christoph Becker and Thomas Schuster

Considerations in Introducing a Blended Learning Environment Dedicated to Training Computer Network Administrators 85
Bartosz Marcinkowski and Piotr Ostrowski

Author Index 95

Rule Based Approach for Ensuring Consistency in Different UML Models

Diana Kalibatiene, Olegas Vasilecas, and Ruta Dubauskaite

Vilnius Gediminas Technical University, Sauletekio al. 11, LT-10223 Vilnius, Lithuania
{diana.kalibatiene,olegas.vasilecas,ruta.dubauskaite}@vgtu.lt

Abstract. Unified Modeling Language (UML) presents us a collection of diagrams to model different aspects of a system, like behavior, structure, functionality, etc. Refinement of models and the evolving nature of software, may lead to inconsistencies between different diagrams of the same model. Inconsistencies in the IS model specification might result in the development of an incoherent and conflicting system. Current tools lack of support for maintaining the consistency between diagrams. This paper shows that the proposed methods are insufficient for consistency checking in UML models. Therefore, authors of this paper propose a rule based method for consistency checking in UML models. The proposed method is implemented as a software prototype for MagicDraw UML tool.

Keywords: consistency, UML, MagicDraw, model, rule, constraint.

1 Introduction

Unified Modeling Language[1] (UML) is Object Management Group's (OMG)[2] most-used specification and the way of modeling not only system structure, behavior, and architecture, but also business process and data structure. UML model may consist of different diagrams, which describe different aspects or views of an information system (IS). This situation opens to consistency problem where two or more overlapping elements of the different diagrams are not jointly satisfiable. It is very crucial to ensure consistency of a UML models as it is one of the attributes in measuring the quality of a model and a quality of a designing system.

Although, there are increasing researches on consistency ensuring, there are still lack researches of ensuring uniform consistency of a model. None of the analysed methods has been accepted as a standard yet. Since, a number of methods analyse consistency among a particular UML diagrams, like in [1] authors analyse consistency between use case-sequence diagram and use case-class diagram, in [2] authors focus mainly at checking the consistency between the UML statechart and the sequence chart, in [3] author specify consistency between use case, activity, sequence and statechart diagram using Colored Petri Net (CPN), in [4] authors define

[1] http://www.uml.org/
[2] http://www.omg.org/

S. Wrycza (Ed.): EuroSymposium 2013, LNBIP 161, pp. 1–16, 2013.

consistency between use case and class diagram using graphs, in [5] authors describe consistency between use case, activity, sequence, class and statechart diagram, in [6, 7] authors define consistency between use case, activity, sequence and class diagrams using Context Free Grammar (CFG) and etc. Not all researches deals with the quality of consistency rules. Moreover, not conforming to metamodel of modelling language, sometimes meaningless consistency rules reduce the reusability and practical applicability of the proposed methods. Therefore, it is relevant to propose a method for consistency checking of UML diagrams using rules and paying special attention to the requirements for consistency rules.

The last part of the paper is structured as follows. Section 2 presents related works on consistency checking of UML models. Section 3 introduces the rule-based method for consistency checking in UML models. Section 4 concludes the paper.

2 Related Works

Our research gives attention to *consistency of UML models*. Therefore, the related works that analyze conformance of different aspects models (expressed by consistency rules) are selected for a more detailed analysis. As presented in [8], there are several trends for consistency checking in UML diagrams: meta-model based methods [9], graph-based methods [10, 11], scenario-based methods, constraint-based methods (are the most popular) and knowledge-based methods (like [12, 13]). We are concentrated on meta-model and constraint based methods.

For the detailed study 50 consistency rules were elicited from 8 related researches (see Table 1) and examined in order to:

1. evaluate consistency rules, excluding redundant rules;
2. find out whether the provided rules may be understood unambiguously;
3. determine whether they conform to specification of a model – UML metamodel;
4. find out whether they are meaningful, i.e. whether they really show a conflict of consistency.

The count of consistency rules associating UML models of specific aspects provided in the specific research is presented in Table 1, Part “Associated different aspects models”. The four last columns in Table 1 indicate whether the rule expressed in a natural language or/and a formal language, is the proposed approach related with OMG IML metamodel, and is the proposed approach implemented. A plus sign (+) indicates that all the rules provided in the paper have specific expression; otherwise, a number shows the count of rules expressed in a natural language, containing metaelements from OMG UML specification or having a formal expression. The analysis shows that all the analysed rules are expressed in a natural language, and most rules have a formal expression. However, not all approaches are associated with OMG UML metamodel and implemented.

Table 1. Results of consistency rules analysis of 2007-2013 years

Consistency rules Author, Year [reference]	**Associated different aspects models**																			**Expressed in a natural language**	**Expressed in a formal language**	**Related with OMG UML metamodel**	**Implementation**
	Class – State	**Class – Sequence**	**Class – Activity**	**Class – Use Case**	**Sequence – State**	**Sequence – Activity**	**Sequence – Use case**	**Activity – State**	**Activity – Use case**	**Object – Sequence**	**Object - Activity**	**Object – Collaboration**	**State – Activity – Class**	**State**	**Use case – activity – sequence**	**Use case – class – activity – sequence**	**Class – activity – sequence**	**Class – use case – activity**	**Total**				
Egyed, 2007 [14]		1			1														2	+			UML/Analizer Tool
Sapna and Mohanty, 2007 [5]	2	2	1	1	1	2	1	1	1										12	+	1	+	OCL
Chanda et al., 2009 [6]			3						1										4	+	+		
Kotulski, 2007 [15]				1															1	+			
Ibrahim et al., 2011 [16]									3										3	+	+		-
Ha and Kang, 2008 [8]										3	1	3							7	+	+	+	OCL
Khai et al. 2011 [17]		6																	6		+	+	Prolog
Borba and Silva 2010 [18]	1	1	1	1			2		2				1	1	1	1	2	1	15		+	+	OCL
Total:	3	10	5	3	2	2	3	1	7	3	1	3	1	1	1	1	2	1	50				

Table 2 gives a summary of the analysed NoMagic MagicDraw, Sybase PowerDesigner, Gentleware Poseidon for UML, IBM Rational System Architect and Microsoft Visio tools.

Table 2. Comparison of the design tools Magic Draw 17.0, Power Designer 16.1, Poseidon for UML 8.0, Rational Software Architect 11.3.1 and Visio 2010

Tools / Criteria	Magic Draw 17.0	Power Designer 16.1	Poseidon for UML 8.0	Rational Software Architect 11.3.1	Visio 2010
1. MCM	+	+	partially	+	partially
2. CorC	+	+	-	+	-
3. ConC	partially	-	-	partially	-
4. Lang	OCL, Java	Visual Basic	Java	Visual Basic	Visual Basic (for Macros), .NET (for plugin)
5. Tech	module, plug-in	plug-in	plug-in	macros	macros, plug-in

The criteria for comparison are described as follows:

1. *MCM – Model in conformity with a metamodel.* Possible values are "+" (almost conform) and "partially" conform. If a model is in conformity with UML metamodel, it is checked according to one rule: *name of a class has to be unique*. If the tool does not allow creating a class with the same name in the model, then it is assumed that the model almost conforms to the metamodel. It is said "almost" because it is not checked whether all constraints defined in a metamodel are implemented in the tool. If the tool allows creating two classes with the same name, it is assumed that the model partially conforms to the metamodel. It is said "partially" because a tool does not implement all the constraints defined in the specification of UML; however, it provides metaelements defined in a metamodel.
2. *CorC – Correctness checking* – constraints are defined at the metamodel level for one aspect model, e.g. for a class diagram.
3. *ConC – Consistency checking* – constraints are among 2 and more different aspect models at the metamodel level. Value "partially" means that there are only several rules that constrain two different aspect models, e.g. Class and Sequence.
4. *Lang* – Language for expressing/ implementation rules.
5. *Tech – Technique of tools extension with new rules.* Examples are developing a module or plug-in, or macros or using other techniques for the extension of the tool with new rules.

Despite the existence of many tools, it is not easy to develop models that conform to UML metamodel. Moreover, not all available tools have a facility to check consistency of IS models, and almost all defined constraints are for one aspect models.

3 The Rule-Based Method for Consistency Checking in UML Models

In this paper we present extension of the proposal [19]. According to the results obtained during the analysis of the related works, 12 rules for consistency checking is defined and the rule-based method for consistency checking in UML models is proposed. The defined consistency rules were expressed in OCL and MagicDraw UML tool was chosen as a tool for UML models development. The main idea of the experiment is presented in Fig. 1.

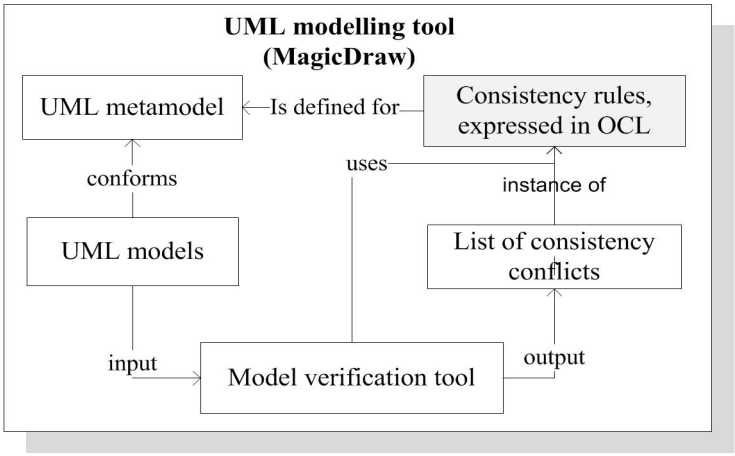

Fig. 1. Checking models, expressed in unified modelling language, using consistency rules

OCL was selected for expressing consistency rules because:

- consistency rules would be interpreted unambiguously (OCL is a formal language); and
- consistency rules would be compatible with well-formedness rules of UML metamodel that are also expressed in OCL.

3.1 Consistency Rules Defined

In this section as an example we present description of 7 rules (Table 3–9).

Table 3. Definition of the consistency rule 1

Consistency rule ID		R1	
Rule at a metamodel independent level		Transition from one class state to another state can be caused by calling operation of the class.	
Rule at the metamodel specific level	**Rule**	Protocol transition of protocol state model has to be defined by operation of the class model.	
	Associated metaelements	Operation of Class model	ProtocolTransition of Protocol State Model
Enforcement level		Medium	
OCL expression		`context ProtocolTransition inv ProtocolTransition_notDefined: self.refered->notEmpty()`	

Table 4. Description of consistency rule 2

Consistency rule ID		R2	
Rule at a metamodel independent level		The class which states are modelled has to be known in the Protocol states model.	
Rule at the metamodel specific level	**Rule**	Context of protocol states has to be defined by the class.	
	Associated metaelements	Context of Protocol State Machine	Classifier of Class model
Enforcement level		High	
OCL expression		`context ProtocolStateMachine inv protocolStates_without_context: self.oclAsType(StateMachine).region.context->notEmpty()`	

Table 5. Description of consistency rule 3

Consistency rule ID		R3	
Rule at a metamodel independent level		Type of the object used in Sequence model has to be defined in Class model.	
Rule at the metamodel specific level	**Rule**	The type of lifeline should be specified.	
	Associated metaelements	Class of class model	Type of Lifeline
Enforcement level		Medium	
OCL expression		`context Lifeline inv lifeline_without_type: not self.represents.type.oclIsUndefined()`	

Table 6. Description of consistency rule 4

Consistency rule ID		R4	
Rule at a metamodel independent level		Abstract class cannot be instantiated in a sequence model.	
Rule at the metamodel specific level	**Rule**	Type of Lifeline in Sequence/Communication model cannot be a class that is Abstract.	
	Associated metaelements	Type of Lifeline of Sequence/Communication model	Class of Class model
Enforcement level		High	
OCL expression		`context Lifeline inv abstract_class_instantiated: not self.represents.type.isAbstract = #true`	

Table 7. Description of consistency rule 5

Consistency rule ID		R5	
Rule at a metamodel independent level		Call Message of an object has to be associated with Operation of a class.	
Rule at the metamodel specific level	**Rule**	Message, which is caused by CallEvent, should have an operation assigned.	
	Associated metaelements	Operation of Class model	Message of Interaction model.
Enforcement level		High	
OCL expression		`context Message inv callmessage_without_operation: ((not receiveEvent.oclAsType (MessageOccurrenceSpecification).event.oclIsUndefined()) and (receiveEvent.oclAsType (MessageOccurrenceSpecification.event.oclIsTypeOf(CallEvent))) implies not receiveEvent.oclAsType (MessageOccurrenceSpecification).event.oclAsType(CallEvent).operation.oclIsUndefined()`	

Table 8. Description of consistency rule 6

Consistency rule ID		R6	
Rule at a metamodel independent level		Visibility of operation, which is assigned to Call Message, has to be not private.	
Rule at the metamodel specific level	**Rule**	Visibility of operation, which is assigned to Message that is caused by Call Event, has to be not private.	
	Associated metaelements	Visibility of Operation of Class model	Message of Interaction model
Enforcement level		High	
OCL expression		`context Message inv private_operation_call: let sourceLifeline:Lifeline = self.sendEvent.oclAsType(MessageOccurrenceSpecification).covered->any(true) in let targetLifeline:Lifeline=self.receiveEvent .oclAsType(MessageOccurrenceSpecification).covered ->any(true) in (sourceLifeline <> targetLifeline) implies (let sourceType:Type = sourceLifeline.represents.type in let targetOperation:Operation =self.receiveEvent.oclAsType(MessageOccurrenceSpecification).event.oclAsType (CallEvent).operation in (not sourceType.oclIsUndefined() and not targetOperation.oclIsUndefined()) implies ((targetOperation.visibility <> VisibilityKind::private) or (targetOperation.UMLClass = sourceType)))`	

3.2 ConsistencyConstraints4UML Module Prototype

A prototype of the ***ConsistencyConstraints4UML*** software module prototype for MagicDraw UML 16.0 tool is developed to carry out the experiment of automatic checking consistency of UML models. MagicDraw UML provides a validation mechanism (called Validation Profile) for extending tool with new rules, because UML is a method-independent language. Therefore, the UML modelling systems do not enforce those method-dependent rules, which are not standardized and vary among many different methods [20].

It is necessary to mention validation and verification concepts. In computer science *verification* means that a product, a service or a system meets standard or specification requirements; whereas *validation* refers to meeting the needs of the intended end-user or a customer [21, 22]. Hence verification term is more suitable for checking a model according to the defined rules. However, a validation term is used here because it is used in user interface and documentation of MagicDraw UML, which is a tool of our experiment.

Table 9. Description of consistency rule 7

Consistency rule ID		R7	
Rule at a metamodel independent level		Calling direction of a message has to conform to the direction of association.	
Rule at the metamodel specific level	**Rule**	Operation, which is assigned to Message that is caused by Call Event, should be navigableOwnedEnd.	
	Associated metaelements	navigableOwnedEnd of Association in Class model	Message of Interaction model
Enforcement level		High	
OCL expression		`context Message inv not_navigable_operation_call Property::isNavigable(): Boolean isNavigable = not classifier->isEmpty() or association.owningAssociation.navigableOwnedEnd->includes(self) Let openend: AssociationEnd=self.endData->selected(ed|ed.value->size()=0)->asSequence()->first().end in opened.isNavigable()`	

Developing a software prototype, we use a lightweight technique for tool extension – module. The Validation profile, whose structure is presented in Fig. 2 is used for extending tool with a new suite of rules for checking consistency of models. Severity level matches enforcement level in our proposed approach, and enforcement level has three values: high (interpreted as an error), medium (interpreted as a warning) and low (interpreted as information). Therefore, three attributes of 'SeverityKind' are used in our developed module.

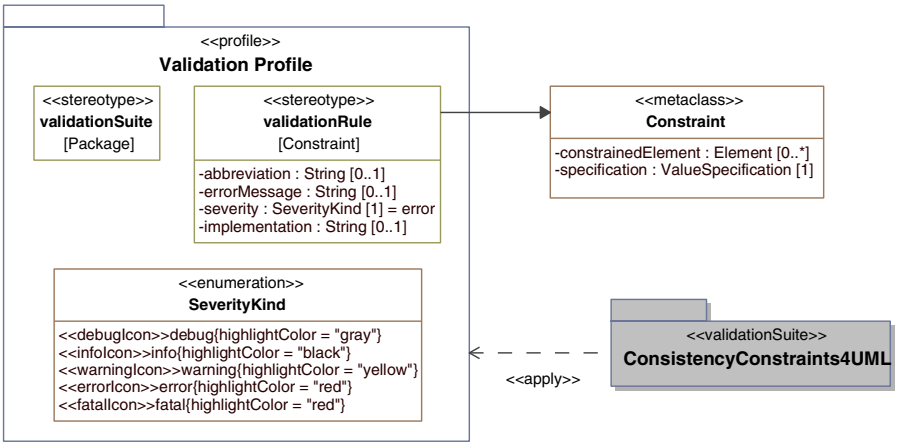

Fig. 2. Extension profile for validation rules

First of all, a package ConsistencyConstraints4UML is created and stereotype « validationSuite » is applied to it. Our developed suite of consistency rules (which is called ConsistencyConstraints4UML) is in grey in Fig. 2. The package groups related consistency rules and they can be invoked together.

The example of rule implementation is defined in Fig. 3.

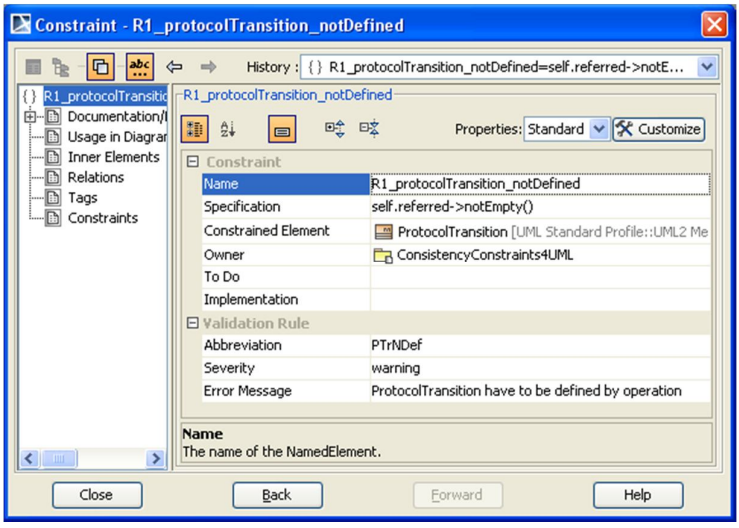

Fig. 3. Implementation of a consistency rule 1

The detailed process of ConsistencyConstraints4UML module implementation is defined in Fig. 4.

The main files in the project are:

- ConsistencyModuleImplementation.mdzip – the file where ConsistencyConstraints4UML rule set is created and tested; and
- ConsistencyConstraints4UML.mdzip – module prototype for consistency checking of UML models.

After the implementation of consistency rules (ConsistencyModuleImplementation.mdzip) they are exported to module ConsistencyConstraints4UML model and tested with specific UML models. The result of export is the file ConsistencyConstraints4UML.mdzip, which can be reused in other UML modelling projects for checking consistency of UML models. The developed module can also be extended with new rules. If the knowledge engineer wants to add new consistency rules using MagicDraw UML tool, first of all he/she has to select package ConsistencyConstraints4UML and then to repeat 2–10 steps according to Fig. 4.

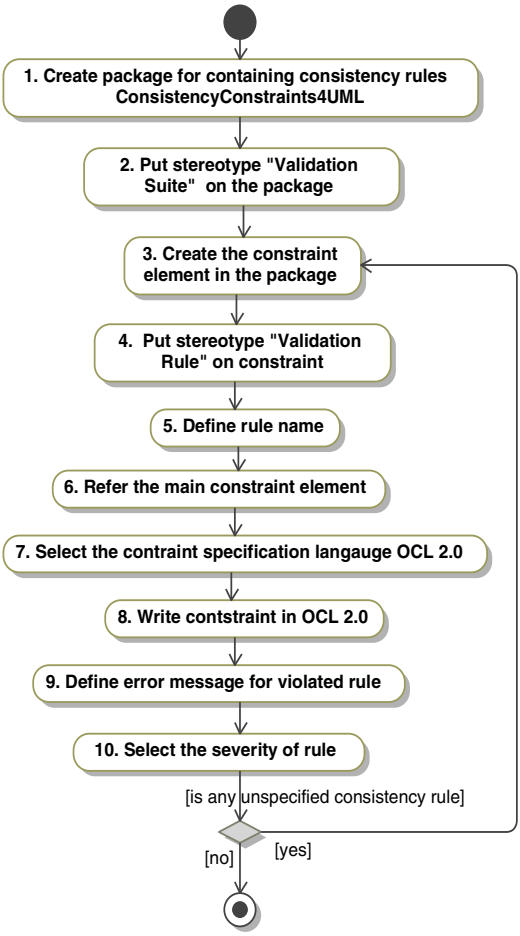

Fig. 4. Implementation of consistency rules

The prototype of ConsistencyConstraints4UML module can be reused in a test or real project by importing the developed consistency module to standard MagicDraw UML tool (Fig. 4). The detailed process of the usage of ConsistencyConstraints4UML is presented in Fig. 5.

The usage of ConsistencyConstraints4UML for checking consistency of the real UML model is presented as follows.

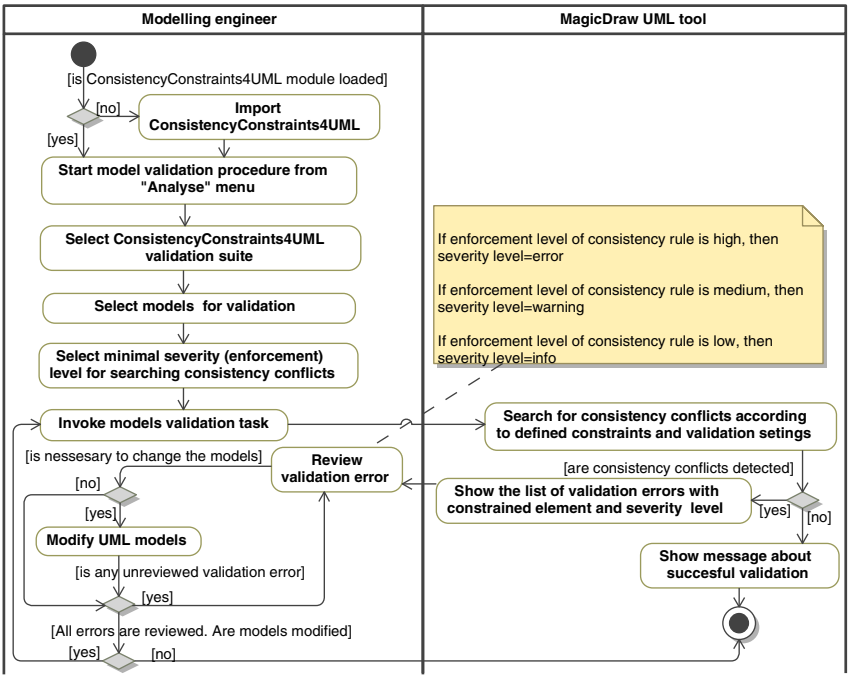

Fig. 5. Usage of ***ConsistencyConstraints4UML***

3.3 Checking Consistency of a Library UML Model

For the consistency checking of a UML model a Library UML model was created according to the OOAD method [23]. The library system was modelled from class, states and sequences aspects. UML class diagram is used for representing classes of Customer, Book, BookCategory, Category, Loan, Penalty, Administrator, their properties and association. States of Book and Loan are represented using UML protocol states machines. Processes of search book by category, search book by keywords, login, make reservation, register loan, cancel reservation, register return, pay penalty, register customer and remove customer are modelled in UML sequence diagrams. Part of a Library UML model is shown in Fig. 6.

The process of checking UML models is executed automatically. The input of the models validation function is a Library UML model and the output is a list of errors.

The example of the detected consistency conflict is shown at the bottom of the right column in validation results section of Fig. 7 (see red ellipse).

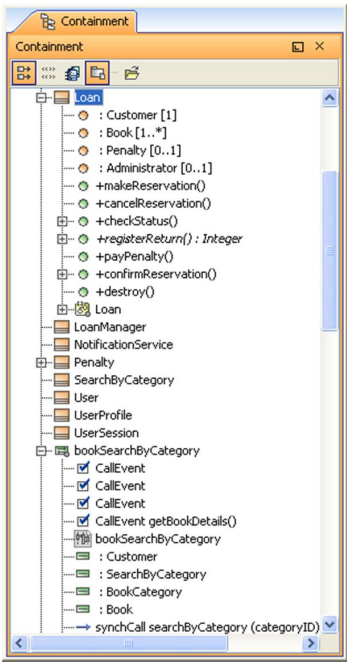

Fig. 6. Part of input models to the consistency checking process

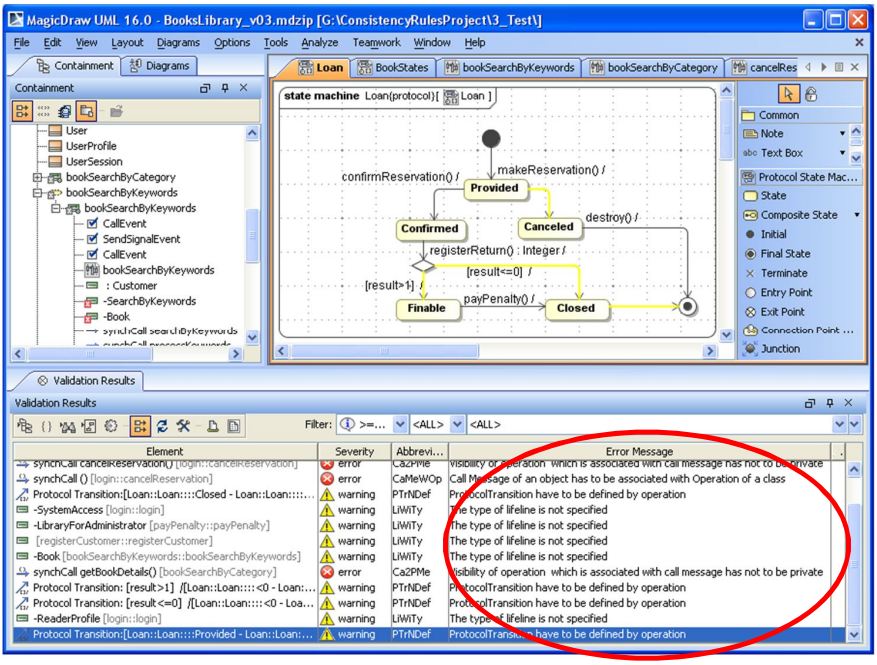

Fig. 7. Checking models of Library UML model using a developed software prototype

In this case a tool can be assistant that helps to detect violations of consistency. 'Error' means that some defined consistency rule is violated and diagram should be improved. 'Warning' meets a enforcement level 'Medium' and means if a rule is violated, then 'sometimes' it is necessary to perform changes of the model. An example of such a situation when there is no need to change a model according to the detected consistency conflict is transition from the choice point to states *Finable* and *Closed* in Fig. 7. Transition depends on the result of the previous operation (*registerReturn()*); therefore these transitions do not have to be associated directly with the operation and 2 violations of the rule can be ignored.

4 Conclusions and Future Work

Analysis of approaches for consistency checking in UML models shows, that there is luck of a complete rule set for consistency checking in UML models. A number of authors present their own rule sets for consistency checking in UML models. In some cases these rule sets overlap. Most rules are expressed in natural and formal languages. Rules expressed in a natural language may be interpreted ambiguously. Formal rules usually use their own description of UML models. Therefore, it remains unclear what elements of OMG UML metamodel they conform to. Moreover, some consistency rules do not conform to OMG UML metamodel, and their practical necessity is doubtful. Not all approaches of consistency checking in UML models are implemented.

The analysis of UML design tools demonstrates that most of them allow developing models that do not conform to UML metamodel. It means that consistency rules have to associate metaelements from different aspects models despite the fact that they are directly associated in metamodel.

The rule-based method for consistency checking in UML models is created and implemented as *ConsistencyConstraints4UML* prototype for MagicDraw UML. The experiment of checking consistency of Library UML model shows that the usage of the proposed method allows detecting inconsistencies of different aspect models. An enforcement level of constraints helps to decide whether it is necessary to react to violation of the rule and to change the UML model.

Our future work is developing full consistency rule set and extending prototype.

References

1. Ibrahim, N., Ibrahim, R., Saringat, M., Mansor, R., Herawan, T.: Use case driven based rules in ensuring consistency of UML model. AWERProcedia Information Technology and Computer Science 1, 1485–1491 (2012)
2. Du, J., Jiang, G.H.: Consistency Check Between UML State Chart and Sequence Chart Based on Model Checking. Electronic Science and Technology 25(2), 100–104 (2012)
3. Shinkawa, Y.: Inter-Model Consistency in UML Based on CPN Formalism. In: Jalote, P., Kumar, R. (eds.) Proc. of the 13th Asia Pacific Software Engineering Conference (APSEC 2006), pp. 414–418. IEEE Computer Society, Washington (2006)

4. Fryz, L., Kotulski, L.: Assurance of System Consistency During Independent Creation of UML Diagrams. In: Zamojsky, V., Mazurkiewicz, J., Sugier, J., Walkowiak, T. (eds.) Proc. of the 2nd International Conference on Dependability of Computer Systems (DepCoS-RELCOMEX 2007), pp. 51–58. IEEE Press, Washington (2007)
5. Sapna, P.G., Mohanty, H.: Ensuring Consistency in Relational Repository of UML Models. In: Patra, P., Panda, P.R. (eds.) Proc. of the 10th International Conference on Information Technology (ICIT 2007), pp. 217–222. IEEE Computer Society, Washington (2007)
6. Chanda, J., Kanjilal, A., Sengupta, S., Bhattacharya, S.: Traceability of Requirements and Consistency Verification of UML UseCase, Activity and Class diagram: A Formal Approach. In: Proc. of the International Conference on Methods and Models in Computer Science 2009 (ICM2CS), pp. 1–4. IEEE Press, Washington (2009)
7. Chanda, J., Kanjilal, A., Sengupta, S.: UML-Compiler: A Framework for Syntactic and Semantic Verification of UML Diagrams. In: Janowski, T., Mohanty, H. (eds.) ICDCIT 2010. LNCS, vol. 5966, pp. 194–205. Springer, Heidelberg (2010)
8. Ha, I., Kang, B.: Cross Checking Rules to Improve Consistency between UML Static Diagram and Dynamic Diagram. In: Fyfe, C., Kim, D., Lee, S.-Y., Yin, H. (eds.) IDEAL 2008. LNCS, vol. 5326, pp. 436–443. Springer, Heidelberg (2008)
9. Paige, R.F., Brooke, P.J., Ostroff, J.S.: Metamodel-based model conformance and multiview consistency checking. Transactions on Software Engineering and Methodology 16(3), 11 (2007)
10. Taentzer, G.: AGG: A Graph Transformation Environment for Modeling and Validation of Software. In: Pfaltz, J.L., Nagl, M., Böhlen, B. (eds.) AGTIVE 2003. LNCS, vol. 3062, pp. 446–453. Springer, Heidelberg (2004)
11. Shuzhen, Y., Shatz, S.M.: Consistency Checking of UML Dynamic Models Based on Petri Net Techniques. In: Gelbukh, A., Guerra, S.S. (eds.) Proc. of the 15th International Conference on Computing (CIC 2006), pp. 289–297. IEEE Computer Society, Washington (2006)
12. Wang, Z., He, H., Chen, L., Zhang, Y.: Ontology based semantics checking for UML activity model. Information Technology Journal 11(3), 301–306 (2012)
13. Wang, S., Jin, L., Jin, C.: Ontology Definition Metamodel based Consistency Checking of UML Models. In: 10th International Conference on Computer Supported Cooperative Work in Design (CSCWD 2006), pp. 1–5. IEEE Computer Society, Washington (2006)
14. Egyed, A.: Fixing inconsistencies in UML design models. In: Proc. of the 29th International Conference on Software Engineering (ICSE 2007), pp. 292–301. IEEE Computer Society, New York (2007)
15. Kotulski, F.L.: Assurance of system consistency during independent creation of UML diagrams. In: Zamojsky, V., Mazurkiewicz, J., Sugier, J., Walkowiak, T. (eds.) Proc. of the International Conference on Dependability of Computer Systems (DepCoS-RELCOMEX 2007), pp. 51–58. IEEE Press, Washington (2007)
16. Ibrahim, N., Ibrahim, R., Saringat, M.Z., Mansor, D., Herawan, T.: Consistency Rules between UML Use Case and Activity Diagrams Using Logical Approach. International Journal of Software Engineering and Its Applications 5(3), 119–134 (2011)
17. Khai, Z., Nadeem, A., Lee, G.: A Prolog Based Approach to Consistency Checking of UML Class and Sequence Diagrams. In: Kim, T.-H., et al. (eds.) ASEA/DRBC/EL 2011. CCIS, vol. 257, pp. 85–96. Springer, Heidelberg (2011)
18. Borba, C.F., da Silva, A.E.A.: Knowledge-Based System for the Maintenance Registration and Consistency among UML Diagrams. In: da Rocha Costa, A.C., Vicari, R.M., Tonidandel, F. (eds.) SBIA 2010. LNCS, vol. 6404, pp. 51–61. Springer, Heidelberg (2010)

19. Dubauskaite, R., Vasilecas, O.: The approach of ensuring consistency of UML model based on rules. In: Proc. of the 11th International Conference on Computer Systems and Technologies (CompSysTech 2010), vol. 471. ACM Press (2010)
20. Vitiutinas, R., Silingas, D., Telksnys, L.: Model-driven plug-in development for UML based modelling systems. Information Technology and Control 40(3), 191–201 (2011)
21. IEEE, IEEE Standard Glossary of Software Engineering Terminology. IEEE Std. 610.12 (1990)
22. Sommerville, I.: Software Engineering, 8th edn. Edinburgh (2007)
23. Bennet, S., McRobb, S., Farmer, R.: Object-Oriented Systems Analysis and Design Using UML, 4th edn., London (2010)

Development of Requirements on the Role Business Analyst

Petr Doucek, Milos Maryska, and Lea Nedomova

University of Economics, W. Churchill sq. 4, Prague, Faculty of Informatics and Statistics, Czech Republic
{doucek,maryska,nedomova}@vse.cz

Abstract. The article analyzes the role Business Analyst in Czech companies in the context of current requirements of economic entities. The requirements represent the level and structure of knowledge and experience that the person entering the role should possess. We analyze them in time according to selection surveys carried out between 2006 and 2010. The article also includes a brief presentation of the methodology used for the survey including the method for measuring the knowledge in the different knowledge domains, selection of statistical sample for the selection survey among the companies and definition of how the data obtained was evaluated. In the end, the identified requirements of this role are stated as well as their dependence on the investigated aspects and the impacts of those requirements on the options for training this type of experts in the universities.

Keywords: Business Analyst, knowledge, requirement, corporate sphere, ICT role.

1 Introduction

The dynamic boom of information and communication technologies (ICT) has significant influence on the changes to the professional requirements of all roles [1] that are found in the corporate information technology. However, besides these changes, the corporate practice is also characterized by changes to the role of ICT professionals [2]. In view of this fact, it is therefore not only necessary, but also very important to answer the questions related to changes in competencies and qualifications of professional ICT personnel in connection with changing technologies and their deployment. Competencies and qualification can be analyzed from many different angles. One of these is, for instance, the factors impacting the fluctuation of workforce in companies [3], or the total concept of competencies and qualifications of Business Analysts [4].

The role Business Analyst is one of the traditional roles for ICT professionals in companies and its importance is proved by other researches [19 – 21]. It has a wide range of tasks and acts as a "bridge" transforming the requirements of the customer (requirements originator) to a language intelligible to developers and often communicates, for instance, with the role Enterprise Architect, who sets the direction for the development of the whole corporate information system [5].

Business Analyst [6] works as a liaison among stakeholders to elicit, analyze, communicate and validate requirements for changes to business processes, policies, and

S. Wrycza (Ed.): EuroSymposium 2013, LNBIP 161, pp. 17–26, 2013.

information and information systems. He understands business problems and opportunities in the context of the requirements. This definition is quiet complex. On contrary easier definition of the role Business Analyst is provided by ASRI [7] that define Business Analyst as role that identifies and communicates with users to formulate and produce a requirements specification to create system and software solutions. The last approach and view on Business Analyst role is provided by CEPIS [4]. Based on the CEPIS [4] the Business Analyst is expected to be very effective in understanding business cases, eliciting requirements, modeling business processes and identifying the appropriate type of ICT solutions. For this role, a high level professional attitude and the ability to communicate are as vital as a wide and thorough ICT competence.

The importance of the role Business Analyst in companies is supported by various research works and analyses [8 – 10].

In order to track the assessment of the different ICT roles in practice, we looked for a suitable basis. The concept offered in the paper [3] allows to understand viewing of ICT professionals on the labor market and their behavior as an incremental structure, shown in Fig. 1.

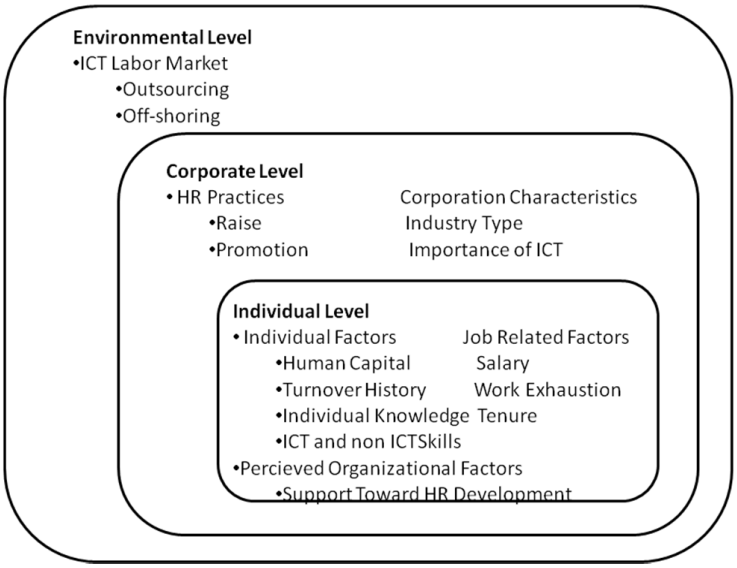

Fig. 1. Research Framework for Turnover of ICT Professionals Source: [3]

Within our analyses, we identified that one of the main current problems is the structure of knowledge expected by companies from the people entering the role Business Analyst. Their identification and exact description is important for the academy as educational establishments can adopt their training programs to the requirements of practice in training professionals for the role BA, thus increasing the employment opportunities of their graduates. Description of required knowledge and level of required knowledge is mentioned for example in [16, 22]. The problem was identified through discussion with professional both practice and academics are. They were usually closely oriented on their area and they did not see anything else.

We have taken and assessed their requirements and we have prepared structure of knowledge presented in part 2.3 and for example in publications [16, 22].

The objective of this article is to present the results of the survey for the role Business Analyst, i.e., to demonstrate the demands on the knowledge and competencies for this role. The role is analyzed by the company size, by the importance of ICT for the company and also by the development of companies' requirements on knowledge competencies between 2006 and 2010.

2 Data Collection and Methodology

The survey prefers classification of specific roles in ICT (rather than particular professions which are in this context too detailed) underlining the competitive ability of graduates based on their knowledge potential. ICT specialist in this context is educated and qualified to use his/her knowledge potential mainly in the design, implementation and operation of ICT and their application

2.1 Role Definition

We introduced the principal definition of the role Business Analyst at the outset of this article. We introduced detailed definition of the role Business Analyst that was analyzed in our survey in Table 1. It is clear from the Table that we defined the role primarily through the key requirements of knowledge and the main activities the person in the role must perform.

Table 1. Definition of the Role Enterprise Architect Source: Authors

BUSINESS ANALYST **Professions:** *analyst,* *corporate process designer,* *business consultant,* *standard software implementer,* *Knowledge engineer,* *Information broker,*	**Key competences:** • Management of relations between business – informatics, • Conception of process organized corporate system based on linking strategic goals with ICT, • Modeling and re-engineering of corporate processes, organization structure in order to optimize them, • Management of knowledge and competences of employees in order to develop system of knowledge management. **Activities:** • analysis, design, standardization and optimation of business processes and corporate organization, • analysis and design of business effects realized by ICT, • analysis and design of knowledge management in corporation, • risk analysis if ICT, business continuity management, • design of ICT services supporting core business processes, • implementation/customization of standard software, • business content management for corporate management support.

2.2 Categories and Levels of Knowledge

The ICT roles can be characterized by knowledge and skills the incumbent should possess. Many authors address this field, for instance Tvaronavičiene [11], who advocates teaching by practical examples, or Melnikas [12], who stresses the need of developing teamwork and communication skills.

Besides "obligatory" knowledge, such as creativity, command of English and capability of teamwork, we defined so-called optional knowledge – specific for the different roles, which we grouped in sixteen knowledge domains (the knowledge domains that are, in our opinion, critical for the role Business Analyst are emphasized by boldface): **MS01 Process modeling**, MS02 Functionality and customization, MS03 IS/ICT management and definition of ICT services and operating options, **MS04 Analysis and design (of the corporate information system as a whole and its parts),** MS05 Software Engineering (techniques and procedures for software development), MS06 Data and information engineering, **MS07 IS/ICT knowledge,** MS08 Operational excellence, **MS09 Team leadership skills,** MS10 ICT market knowledge, MS11 Organizational management methods, MS12 Enterprise finance and economics, MS13 Sales and marketing, MS14 Mathematics, MS15 Law, **MS16 Knowledge in business sectors.**

We described each of these knowledge domains so that it is understandable for the survey respondents. Their description is provided, for instance, in [13 – 16].

The companies determined the level of knowledge using a six-grade nonlinear scale, where**: 0 – No knowledge, 1 – General idea of the field** (corresponds to approximately 1–2 days of training), **2 – Basic orientation in the field and its terminology** (corresponds to approximately 3–5 days of training), **3 – Solid knowledge of the field and basic practical skills** (corresponds to approximately 6–20 days of training), **4 – Solid knowledge of the field and solid practical skills** (corresponds to approximately 21–40 days of training), **5 – Superior quality of knowledge – deep current knowledge and advanced practical skills** (corresponds to 41+ days of training).

2.3 Survey among Businesses

According to data of the Czech Statistical Office, there were 1,266,336 subjects of various size and main economic activity. Sector of main economic activity: There are 17 main sectors in Industrial Classification of Economic Activities of the Czech Statistical Office. According to the requirements on information technologies, these sectors were coded into 3 categories: sectors with the lowest requirements (LIT), sectors with the middle requirements (MIT) and sectors with the highest requirements (HIT). We have coded companies also according to the number of employees into three categories (less than 49 employees, 50-249 employees and more than 249 employees). The structure of companies selected for the survey is described by the Table 2.

The survey was realized by the private research company among clusters of companies that respect structure of companies in the economy of the Czech Republic. The interviewing methods used were CAWI (Computer Aided Web Interviewing) and CATI (Computer-assisted Personal Interviewing).

Table 2. Structure of the Observed Sample 2010 Source: Authors

	0 – 49	50 – 249	250 +	Total
LIT	45	46	44	135
MIT	57	474	98	629
HIT	66	142	39	247
Total	168	662	181	1,011

2.4 Methodology and Assessment

We processed the data on two levels. On the first level, we used mainly statistical analysis of the responses from the survey participants, which is based on the statistical method of median. In order to compare the differences on the different professional roles and the differences between the knowledge categories, we used the technique of box plots (for sampling distributions) and scatter plots (for the median) [17, 18].

The results presented below are based on the replies from companies the structure of which is shown in the following Table 3. The total number of replies from the entities does not correspond to the figures shown in the previous Table 2, because replies were not received from all the respondents.

Table 3. Numbers of Companies Responding the Role Business Analyst Source: Authors

Dependence on ICT			Number of Employees		
	2006	2010		2006	2010
LIT	13	14	**<50**	13	13
MIT	48	62	**50-249**	66	66
HIT	90	189	**250+**	38	40
Total	**151**	**265**	**Total**	**117**	**119**

2.5 Similar Research

There were several researches in the world during last 5 years. Based on our opinion we think that one of the best researches in the world are presented in [19] and [21]. The first one was realized in Saudi Arabia and the second one was realized in Slovenia.

These researches find out similar conclusions as we have. Our conclusions are presented in part 4.

3 Results

This chapter provides detailed findings about required ICT knowledge and skills on Business Analyst profession on perspectives in following two dimensions:

- ICT intensity of the branch (LIT, MIT, HIT sectors),
- Number of employees (small and medium enterprises).

3.1 Requirements on the Role Business Analyst by ICT Intensity of the Branch

The following Tables describe a detailed analysis of the knowledge domains that we regard as essential for Business Analysts.

The following Table implies that the average values tend to decline over time for all types of companies according to their dependence on ICT (LIT, MIT, HIT). We identified decline in requirements with more than 80 % of knowledge domains in the companies in group LIT, in almost 44 % of knowledge domains in MIT companies and more than 60 % of knowledge domains for HIT companies. These figures confirm decline in companies' expectations. With respect to the field of our analysis, it is very interesting that the decline of expected knowledge was identified also for ICT-related knowledge domains (i.e., the group of knowledge domains MS01-MS08). For LIT companies, expectations fell in almost all ICT knowledge domains. For MIT and HIT companies, the decline was approximately in 40 % of ICT knowledge domains (3 of 8).

Table 4. Requirements on ICT Business Analyst by Dependence of the Company on ICT in 2006 and 2010 Source: Authors

Number of Employees	LIT		MIT		HIT	
Year/Domain	2006 n=13	2010 n=14	2006 n=48	2010 n=62	2006 n=90	2010 n=189
MS01 Process modeling	3.15	2.93	3.50	3.06	3.49	2.08
MS04 Analysis and design	**3.15**	**3.43**	**3.00**	**3.23**	3.30	2.77
MS07 IS/ICT knowledge	3.46	3.14	3.44	3.29	3.62	3.52
MS09 Team leadership skills	3.77	3.00	3.25	2.94	3.34	2.47
MS16 Knowledge in business sectors	**2.62**	**3.57**	3.02	2.82	2.89	2.62

The identified declines indicate “sobering up” of companies from their expectations that were identified in the first survey. These requirements significantly exceeded the knowledge possessed by the graduates (Doucek et al. 2012).

The decline in the domain **MS01 Process modeling** can be explained by the fact that a person entering the role Business Analyst does not need significant knowledge in order to carry out process analysis, as this primarily involves the logical comprehension of the problem area and its graphic representation. What is substantially more important is the knowledge in the subsequent domain MS04. Besides, there are increased requirements of the domain **MS16 Knowledge in business sectors**, which explains the requirements of knowing the company's core business. **MS04 Analysis and design**, which requires the knowledge and capabilities of conceptual work, and as is implied by the Table, it is obvious that growth of knowledge requirements is identified in this domain, except companies in the category HIT.

MS07 IS/ICT knowledge and **MS16 Knowledge in business sectors** are knowledge domains which are relatively specific for a concrete field in which the employee is active. Despite this relative focus of a certain field, and thus wide variability, this is a domain, in particular for MS07, which can be, in our opinion, relatively easily

learned and missing knowledge and skills can be acquired over a short period of time. As far as MS16 is concerned, learning is always dependent on the concrete field of the employee's focus. It can be expected that, for instance, knowledge of manufacturing management in automotive industry will be easier and faster to acquire than knowledge of financial services (banks, insurance companies), which are a very complex field.

MS09 Team leadership skills need not be possessed by all employees in this role. In view of the company's growth, only part of the staff must possess the leadership capability, those who are subsequently often moved onto the roles of Project Managers or Enterprise Architects.

3.2 Requirements on the Role Business Analyst by Company Size

According to this parameter, we obtained a total of 117 responses in 2006 and a total of 119 responses in 2010. A detailed breakdown of the responses by ICT requirements of the company is shown in Table 5.

The following Table provides an analysis of the survey results among companies in 2006. Also in this case, the average value of required knowledge in the different domains declines, with an exception of domain MS04, also in the breakdown by company size.

Decline of expected knowledge, but only for 55 % of knowledge domains. An interesting finding is the fact that compared to the analysis of dependence of companies on ICT, the largest decline in the number of knowledge domains with declining requirements took place in medium-sized (50-249 employees) and large companies (250+ employees). Expressed numerically, we identified decline in 75 % of ICT knowledge domains in medium-sized companies and in 50 % of large companies. On the other hand, for small companies up to 50 employees, the decline of expected knowledge was only identified for 25 % of knowledge domains. We believe that this is caused by very high requirements identified for the role Business Analyst in 2006 and its division between Business Analyst and Enterprise Architect. Thanks to this, the companies could define lower requirements of the role Business Analyst and thus eventually increase the number of potential candidates of the role.

By analysis of the figures in the Table we find out that the results are similar as for the breakdown by ICT requirements, which was analyzed in the preceding Section.

Table 5. Requirements on ICT Business Analyst by Company Size in 2006 and 2010 Source: Authors

Number of Employees	10 – 49		50-249		250+	
Year/Domain	2006 n=13	2010 n=13	2006 n=66	2010 n=66	2006 n=38	2010 n=40
MS01 Process modeling	3.41	3.08	3.59	3.21	3.63	3.21
MS04 Analysis and design	**3.30**	**3.92**	**3.22**	**3.30**	**3.29**	**3.30**
MS07 IS/ICT knowledge	3.61	3.58	3.49	3.39	3.74	3.39
MS09 Team leadership skills	3.11	2.77	3.54	3.15	3.55	3.15
MS16 Knowledge in business sectors	**2.59**	**3.62**	3.35	2.98	3.11	2.98

In this case, it is worthwhile to notice the increase of requirements for specialized knowledge in the domain MS04 Analysis and design. For all results of the companies we identified increase of requirements – that is higher requirements are put on the role BA in professionalism in time. In general it can be said that increased requirements of the role BA is obvious in SMEs (in particular the domains MS04 and MS09). This means that teamwork is supported in these companies, as opposed to previous, rather individual projects, and role of BAs in these companies is growing as a bridge between the business and ICT.

4 Conclusion

The results gave rise to several interesting questions and conclusions, at which we arrived from the respondents' comments on the results of the surveys.

Although the ICT role Business Analyst definitely cannot be considered new or newly emerging on the labor market, it must be emphasized that this is a role subject to significant changes in the field of corporate ICT. An example can be new methodologies for the development of information systems, optimization of processes, process management, security management or new legislation. A consequence of this dynamic development can include the commencing specialization of analysts to certain limited areas of the ICT. A typical example may be specialization on management of processes according to ITIL or according to ISO. The importance of this role and its training and development is related mainly to the following:

- **he/she must be capable of communicating with the programmers, to whom he/she presents his/her analyses and explains his/her intentions developed according to communication with the customer's representatives** (therefore, practice requires knowledge in domains MS01 Process modeling, MS04 Analysis and design and MS07 IS/ICT knowledge, at level 3-4 in 2010),
- **he/she must be capable of communicating with the representatives of the customer side, who speak in the business language, and of transforming their requirements to the formal language understood by the programmers developing the solution** (therefore, practice requires knowledge in domains MS04 Analysis and design and MS16 Knowledge in business sectors at level 3-4),
- **within detailed analyses, he/she identifies concrete risks related to ICT development and maintenance** (therefore, practice requires knowledge in domains MS04 Analysis and design and MS07 IS/ICT knowledge at level 3-4).

An important move within our surveys was the finding that between 2006 and 2010, there was a change in what level of knowledge and skills the companies expected. The change involved lowering such expectations. From the history of our project, we assume that an important factor that influenced the demand side was our first survey in 2006, the results of which were widely publicized. According to these, economic entities found out that their requirements were unrealistic, given the present state of training corporate Business Analysts, and tried to replace widely educated BA-type professionals with a number of more specialized experts.

The results of the survey were used for motivating educational establishment to create new majors and courses. These courses were included in the plan of public educational establishments as well as the sector of commercial education.

References

1. Doucek, P., Maryska, M., Novotny, O.: Requirements on the competence of ICT managers and their coverage by the educational system – experience in the Czech Republic. Journal of Business Economics and Management (2012), http://www.tandfonline.com/doi/pdf/10.3846/16111699.2012.658436
2. Kunstova, R.: Changes in position of the IT department in organizations. In: IDIMT 2011. Trauner Verlag Universitat, Linz (2011)
3. Joseph, D., Ng, K.-Y., Koh, C., Ang, S.: Turnover of information technology professionals: a narrative review, meta-analytic structural equation modeling, and model development. MIS Quarterly 31 (2007)
4. CEPIS: Professional e-Competence in Europe. Council of European Professional Informatics Societies, http://www.cepis.org/media/
5. Gala, L., Jandos, J.: After Graduate Education for Enterprise Architects. In: IDIMT 2011. Trauner Verlag Universitat, Linz (2011)
6. Babok: Business Analyst Body of Knowledge (2013), http://babokonline.org/
7. ASRI: ICT Business Analyst, http://www.immi.gov.au/asri/occupations/i/ict-business-analyst.htm
8. Barret, K.: Business Analysis: The Evolution of a Profession. International Institut of Business Analysis (2012)
9. EUCIP: Business Analyst. Professional Profile Specification
10. McVey, M., Bridges, G.: If You Think You Can Do Without a Business Analyst Think Again!, International Institute for Learning, Inc., http://www.irmuk.co.uk, http://wenku.baidu.com/view/cf1a674bf7ec4afe04a1dfe3.html
11. Tvaronaviciene, M., Ginevicius, R.: Application of case teaching technique in educational process. Journal of Business Economics and Management (2003), http://www.tandfonline.com/doi/pdf/10.1080/16111699.2003.9636058
12. Melnikas, B.: Management specialists in the knowledge based society: Life-long learning oriented human resourse development. Journal of Business Economics and Management, http://dx.doi.org/10.1080/16111699.2005.9636104
13. Maryska, M., Novotny, O., Doucek, P.: ICT Knowledge Analysis of University Graduates. In: IDIMT 2010 Information Technology – HumanValues, Innovation and Economy. Trauner Verlag Universitat, Linz (2010)
14. Maryska, M.: Requirements of Small and Medium Companies on ICT Professionals' Knowledge. In: IDIMT 2012. Trauner Verlag Universitat, Linz (2012)
15. Doucek, P., Novotny, O., Pecakova, I., Vorisek, J.: Lidské zdroje v ICT. Professional Publishing, Praha (2007)
16. Doucek, P., Novotny, O., Maryska, M., Nedomova, L.: Competitiveness of Czech ICT graduates. In: Charles, M., Sohail, C. (eds.) Advances in Enterprise Information Systems II. CRC Press, Leiden (2012)
17. Pecakova, I.: Statistika v terénních průzkumech. Professional Publishing, Praha (2010)

18. Vojáček, O., Pecáková, I.: Comparison of Discrete Choice Models for Economic Environmental Research. In: Prague Economic Papers (2010)
19. Al-Jabri, I., Fraihat, M.: Professionals in Saudi Arabia, http://faculty.kfupm.edu.sa/MISAC/imjabri/pub/IIMA1.pdf
20. OECD: ICT Skills and Employment, http://www.oecd.org/sti/ICT-employment
21. Varga, M., Stiffler, C., Luzar-Stiffler, V.: Evaluating IT Knowledge Requirements for Business Professionals, http://bib.irb.hr/datoteka/149808.Evaluating_IT_Knowledge_Requirements_for_Business_Professionals_ITI2004.pdf
22. Maryska, M., Novotny, O., Doucek, P., Pecakova, I., Skarlandtova, E., Vorisek, J., Zid, N.: Lidské zdroje v ICT. Professional Publishing, Praha (2012)

ERP in Project-Driven Organizations: A Case-Study from IT Industry in Poland

Jan Trąbka and Piotr Soja

Cracow University of Economics, Department of Computer Science, Kraków, Poland
{Jan.Trabka,Piotr.Soja}@uek.krakow.pl

Abstract. The analysis of contemporary approaches to management illustrates the growing significance of the concept of project and project-driven organization. The goal of this study is to investigate what requirements should an ERP (Enterprise Resource Planning) system meet to support effectively management of a project-driven company. In doing so, this research builds on a case study of an ERP system implementation in Asseco System S.A., one of the Polish largest IT system integrators. Drawing from the experience of Asseco System S.A., this study discusses information problems, goals, considerations, and stages typical of an implementation project conducted in a project-driven organization. The analysis concludes with the evaluation of the implementation project results and formulation of critical requirements for an ERP system designed for project-driven organizations.

Keywords: ERP, project-driven organization, implementation, Poland.

1 Introduction

While analyzing modern approaches to management used successfully in practice and widely discussed in literature over the recent years, we may notice a significant role of two ideas: process and project. The focus of this study is project and a project-driven organization, i.e. a company whose activities are connected with running various projects [6]. In today's economy many companies use a project-based approach in conducting their activities; they deliver a unique product within a predefined time and financial budget. The project-based approach is popular among various industries such as IT, consulting, advertising, construction, manufacturing etc.

The project-based approach consists in creating dynamic organizational structures around hierarchical functional organization. Project structures have decision-making competence and resources needed to deliver a unique product to a customer. In order to simultaneously manage various projects within a project-driven organization, managers need a tool allowing them to allocate resources to projects and evaluate their profitability. These needs can be satisfied by an integrated enterprise system having additional "project" capabilities.

Enterprise resource planning (ERP) systems are at the moment the most popular software tools used for planning and controlling of company's resources [9]. However, a question arises whether ERP systems are useful in project-driven organizations.

S. Wrycza (Ed.): EuroSymposium 2013, LNBIP 161, pp. 27–38, 2013.

This study hypothesizes that project-based organizations pose a new challenge to ERP systems. In particular, the analyzed case study will reveal what kinds of improvements in a typical ERP system are needed in order to satisfy the needs of a project-based organization.

2 Project-Based Organizations and Their Information Needs

2.1 Project-Based Organizations

A project can be defined as a unique undertaking consisting of a complex set of precedence-related activities that have to be executed using diverse and mostly limited company resources [6]. In a slightly more detailed view, a project is understood as a unique set of coordinated activities, with definite starting and finishing points, undertaken by an individual or organization to meet specific objectives within defined schedule, cost and performance parameters [3].

A project-driven organization has a number of characteristics which cause numerous challenges that an organization of such a type has to face [4]. Project-driven organizations have to use "matrix organization" structures that provide a focus for the management of projects. Company's senior executives should be committed to the use of projects in the design and execution of enterprise strategies and culture of project management and teamwork should be evident and fostered in the organization. Stakeholder management should be a crucial task of the project team's undertakings. Project management is a core business process in the project-driven company and its role should be reinforced as the means for dealing with product, service, and process changes in the enterprise.

2.2 Need for Integration and Knowledge Management

In project-driven organizations there is a need to integrate data not only from multiple organizational units but also from external client organizations. This is also complicated by the fact that the project business requires that the organization is highly flexible due to the complexity and uniqueness of many projects. Additionally, the functional instability might be increased because the needs that projects pose for the organization of the project-driven firm are subject to change. They change not only across projects, but even from one project phase to another [19].

Project-driven organizations face some unique challenges connected with knowledge management. In particular, due to their project-intensive activities, project-oriented companies have to deal with three kinds of knowledge: knowledge connected with desired business value, knowledge of the organizational solution, and knowledge connected with the technical solution [11]. In project-driven organizations an increasing part of useful knowledge is held within involved groups and communities. Therefore, matching between the functioning of autonomous communities is risky and the hierarchical system profoundly depends on the context of interactions within communities [1]. In fact, knowledge transfer might be impeded by a number of communication problems which might occur due to unique and temporary characteristics of

projects, the distance between projects, and the lack or weakness of formal links across projects [20].

2.3 Software Solutions for Project-Driven Organizations

A study conducted by IFS in September 2012 among 200 manufacturing companies in the USA revealed a number of issues faced by project-driven organizations [7]. The IFS study showed that the main challenges faced by project-driven organizations refer to integration between project management systems and enterprise systems. Although project-based activities were crucial for 85% of researched organizations, almost 40% did not have any kind of integration among the systems. The study also revealed that organizations suffered from fragmented data generated by various loosely coupled desktop applications.

The IFS study demonstrates that integration fosters better control of the company's processes. Nonetheless, control mechanisms used by the researched companies boil down to first and foremost budget and schedule monitoring. Other important issues, such as human resources, machinery, and risk, were controlled to a much lesser extent. Nevertheless, we have to bear in mind that project-driven organizations which rely on strict project control and evaluation methods and organize their operations around the project control systems may face problems with fostering innovations [8]. This is connected with the fact that the project-driven firms prioritize efficient management of projects and tolerate the use of slack resources only when absolutely necessary, thus hindering innovations.

In order to face the abovementioned challenges, project-driven companies need comprehensive software solutions to manage their projects. Apart from integration- and knowledge-related issues, such solutions should support an integrated document management system and should provide a complete and customizable set of technical variables defining characteristics of a project [21]. The data structure employed by the software solution has to support the fact that projects can have many modifications throughout their lifecycle. Finally, project-driven organizations need tools to handle communication among project members, who can belong to one or several groups linked to each project. Summing up, it appears that the most convenient software solution able to meet the abovementioned requirements is an ERP system which has a flexible and open architecture and has integrated project-management functionality.

3 Description of the Case Company and the Implementation Environment

3.1 Company Description

Asseco Systems S.A. is the project-driven organization used in this case study. The company operates in IT industry and specializes in system integration. After merger

with its mother-company Asseco Poland S.A., the organization became a business unit. The company's business activity might be divided into three areas: system integration, ICT infrastructure, and IT service and outsourcing. Such complex activities involve a variety of processes in many domains and require extensive project management, embracing projects of different scale and characteristics. The most complex projects involve those conducted in the area of service and IT outsourcing. Service Department offers comprehensive services including the maintenance of the whole information and communication infrastructure of the client companies. This is performed on a both standard and outsourced basis.

The company has 18 branches scattered over the whole country that allow for a delivery of time-critical services. Asseco Systems offers its services to first and foremost companies and institutions operating in public administration, banking and financial sector, telecommunication, manufacturing, trade, and service. In 2009 the company employed 400 people and was the Polish fifth largest system integrator in the ranking performed by the Computerworld magazine [5].

3.2 Implementation Project Considerations: Information Problems and Goals

Asseco Systems is a project-driven organization and its activities are divided into projects conducted for the company's clients and governed by the signed contracts. The projects are restricted in time, have predefined budget and involve company's resources. The key information required by the company's management refers to the project profitability. During the project realization various organizational units are involved and a range of products and services might be delivered. Therefore, as compared with traditional companies, the cost and profit accounting must be more granular and multidimensional. For the purpose of this study we will call various dimensions of cost and profit accounting (i.e. projects, organizational units, product types etc.) controlling cross-sections. The main information problem is that all source data gathered within an organization (payroll, purchase and sale invoices, etc.) have to be classified (recorded) simultaneously into all controlling cross-sections that are important from the perspective of managerial analysis.

The project is naturally the basic cross-section; however, other important perspectives include organizational units, product and service types, or location. The employed set of controlling cross-sections is decided by the company's management and in practice their number ranges from 6 to 11. This requirement is difficult to satisfy when an organization has a number of independent systems governing operational processes and autonomous analytical tools for controlling analysis. In such a situation, performing complex project analyses is very time consuming, error prone (largely due to lack of common dictionaries) and, most importantly, subject to miss the deadline. A diagram illustrating data flow between the ERP areas (main processes) and the project-related, analytical data structure is presented in Figure 1.

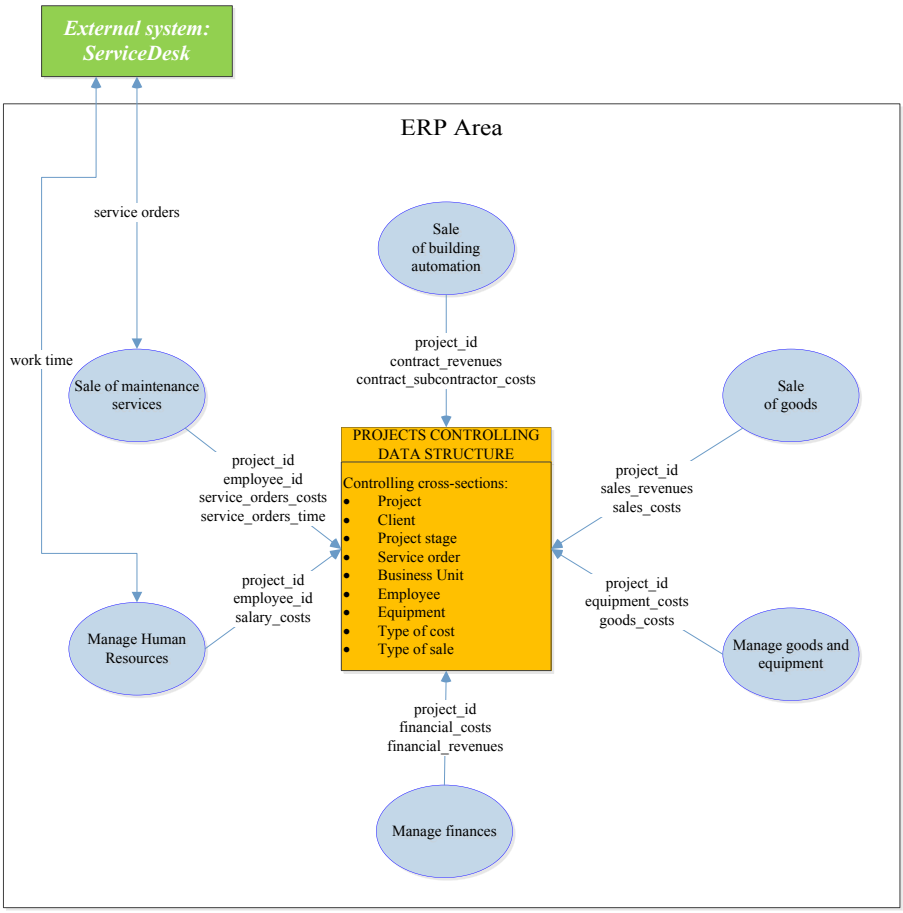

Fig. 1. Data flow between the main processes and the project-controlling data structure

The case organization, before making a decision on the implementation of a new software solution, was using an ERP system which was not integrated with project management system at the operational level and did not have any analytical tool satisfying the requirements of a project-driven approach. The analyses were performed in the external, independent tools. The used ERP system did not have any dedicated module for project management; therefore, the projects repository was maintained in an external application. In consequence, improvement and updating project reporting became the first goal of the new system implementation. The second goal included the optimization and integration of business processes at the project level within the whole organization. The most difficult area involved services with a special focus on so called service logistics involving management of parts which might have been owned by the company or held on deposit. Processes in this area are unusually complex and multidimensional, and the result of related projects hinges upon their

performance. The main parameter defining quality of services (SLA) is the guaranteed time of disaster recovery which varies from a dozen to even two hours.

The case organization employed a dedicated portal – service desk – to manage service requests. Service desk can be accessed by both customers and company's employees from various departments such as service, logistics, and call center. The service portal was not integrated with the ERP system which resulted in the necessity of manual entering data connected with service requests into the ERP system's logistics module. Next, feedback on the related process status had to be keyed in the service portal. Lack of integration resulted in process discontinuity and excessive workload required in order to meet service deadlines.

4 ERP System Choice

The company came to a conclusion that, in order to realize the defined goals, the new system has to possess, apart from the basic features of an ERP system, a repository of projects and ability to record project-related transactions in all system modules – integration at the project level. The needed software tool had to be flexible and configurable in order to capture complex and changing logistic-service processes. Due to a required online data exchange with the service system, during the selection process attention has been paid to the system openness.

The system Softlab SQL, provided by Softlab Trade Sp. z o.o. (at the moment Asseco Business Solutions S.A.), has been selected. It is worth mentioning here that there are several suppliers (both local and global companies) on the Polish ERP systems market who provide integrated solutions dedicated to project-driven organizations. These providers include:

- IFS (IFS Applications 8, Oracle database),
- SAP (mySAP, MS SQL Server/Oracle database),
- BPSC (Impuls 5, Oracle database), emphasizing database as a critical driver for integration with a company's legacy systems.

The chosen solution, Softlab SQL, is a multi-module integrated enterprise system based on a central database operated by MS SQL Server. The chosen system revealed all characteristics of a modern enterprise system and appeared also to satisfy specific requirements defined by Asseco Systems S.A.

4.1 Flexibility

One of the most important characteristics of the Softlab SQL system is flexibility, which is achieved by means of separating business logic, user interface, and data. Business logic is programmed in database procedures, while definitions of user interface, such as dialog windows or reports, are stored in external configuration files. Such a construction allows for changing system layout, behavior of functions and processes, and business rules without the intervention of programmers and without changing application's core functionality. A separate module "Solutions Creator" was

designed for developers in order to create and modify business logic, user interface, and database structure. Using this module in the implementation process allows the adopters to achieve good fit between the system and client organization within reasonably short time and low costs, without time consuming modification of source code.

4.2 Support for Projects

The integral module for project management, "Softlab Projekty", was one of the main features determining the choice of the Softlab SQL system. This module allows the company to record, plan, control, and account for conducted projects. The main elements of the module are registers of projects, contracts, schedules, human resources, working time, purchase, and other costs and incomes. The module is integrated with financial, accounting, and controlling areas. All financial documents gathered in the project module are automatically entered in the ledger. Another very important feature of the system is connected with simultaneous, together with accounting records, recording of controlling data on the basis of controlling cross-sections defined in the source documents.

4.3 Openness

Openness is supported by Softlab SQL mainly due to the fact that the system is based on the same database platform (i.e. MS SQL Server) as the service system. Such a configuration allows the company to use the native drivers and mechanisms of data exchange between databases. This should result in a greater stability and error resistance of the whole solution.

5 Implementation Process and Adopted Solutions

5.1 Pre-implementation Analysis

The realization of the implementation project started with a detailed pre-implementation analysis which involved business process modeling and definition of data structure and inter-system interfaces. A project team has been set up and various roles and responsibilities have been allocated. A member of the Asseco Systems board has been appointed as a Chief of the Steering Committee and also served as the project Sponsor. The most experienced employees from all company's areas have been selected as Key Users.

The analysis started with a definition of the company's main processes and assigning Key Users and provider's analysts to them. Workgroups defined in this way performed processes specification and decomposition at lower levels. As a result of the performed analysis, seven main and eighty five elementary processes have been identified. On the basis of elementary processes, the project products have been defined, together with related prototype iterations, and all individual elements have been aligned and ordered in the logical whole. Also, methods of communication and

project tasks performance monitoring have been defined. The analysis stage lasted for three months and was concluded by the acceptance of the final version of the document. The last element of this stage included the definition of a detailed project schedule.

5.2 Implementation of the Solution

The implementation has been conducted simultaneously in all involved areas; however, it was systematized in order to start with the most complex functionalities or those common for several areas (e.g. contractors dictionary, materials dictionary, projects dictionary). The transactional interface with the service system was the most complicated and the riskiest element of the implemented configuration. The transactional character of this solution consisted in the fact that an action performed in the service system, such as entering a service request, automatically generated a related demand in the logistic module of the ERP system. On the other hand, changes of demand status in the ERP system generated feedback and resulted in changes in the service system.

Project management and project controlling were the implementation areas that caused the greatest workload. They were largely developed from scratch on the basis of the client's requirements. The implementation of multidimensional controlling analyses required, as a precondition, the proper configuration of all modules being implemented as these modules generate source information used in the controlling analyses. An additional configuration has been performed and involved activating the possibility of breakup into controlling cross-sections as regards documents connected with sales, purchasing, payroll, depreciation etc. A multidimensional data warehouse based on OLAP technology has been created for managerial purposes. Other implemented modules included CRM, sales and logistics, finance and accounting, human resources, and fixed assets.

In the implementation of other processes, an iterative approach using prototypes has been employed (a prototype-iterative approach). The implementation consultants delivered prototypes of working elementary processes to the Key Users. Next, the Key Users on a regular basis verified the created solutions and not only indicated small errors, but also, more importantly, conceptual mistakes whose correction during the next stages of the project would have been extremely costly. The prototype-iterative implementation approach is illustrated in Figure 2. The implementation stage lasted for 6 months and its final tasks were performed together with the operational system rollout.

5.3 System Rollout and Stabilization

A month-long period preceding the system rollout has been devoted to activities related to data conversion and group trainings of users on the new system. The group trainings were conducted by joint teams in which the Key Users were often trainers and external consultants acted as technical assistance. The adopted approach to productive start assumed that the new system rollout will take place together with simultaneous work in legacy systems. Abandoning legacy systems took place gradually over the period of two months.

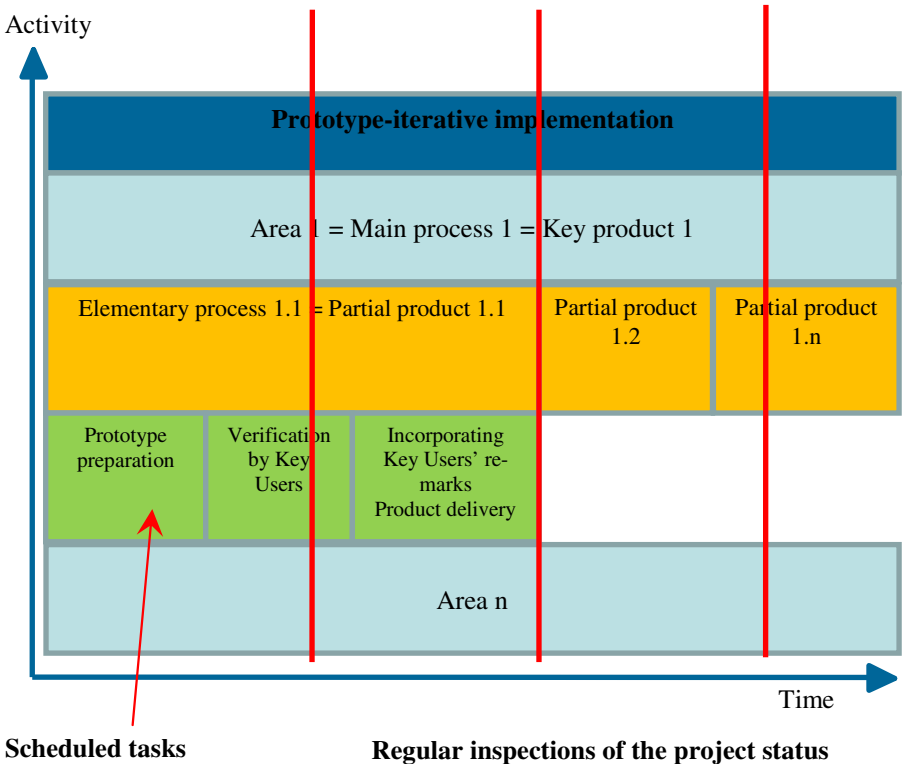

Fig. 2. Prototype-iterative implementation (based on materials of Asseco Business Solutions S.A.)

6 Lessons Learned

The case study allows us to draw several conclusions that should be useful for organizations whose operations rely heavily on projects and who face the challenge of a new enterprise system implementation. The recommendations, grouped according to the related implementation cycle stage, are presented below.

- **System choice:** Project-driven organizations need a special type of system modules integration – integration at the project level. Such functionality consists in common dictionaries of products, materials, contractors, and employees. Lack of project-level integration results in a time-consuming cost and profit accounting while taking into account the perspective of individual projects.
- **Pre-implementation analysis:** Within a project-driven organization project teams are multidisciplinary and inter-departmental structures, and internal processes within projects are more complex and prone to inter-departmental disturbances.

Therefore, analytical teams working on pre-implementation analyzes should include a broad group of Key Users representing all involved areas of a company.

- **Implementation:** The analyzed case study illustrates that if an organization has specific, non-standard processes, the verification of new solutions should be started by users as soon as possible. In such a case it is advised to avoid a waterfall implementation approach. These requirements are satisfied by the prototype-iterative implementation approach presented in this study.

The main limitation of this study's findings is connected with the scope of research results. This is due to the researched companies who represent organizations from Poland. In consequence, we should generalize the results for other countries or regions with caution. In doing so, we should be aware that Poland is a transition economy, i.e. an economy in transition from communist style central planning to a free market system [12] and extant research suggests that ERP adoptions in transition economies may differ from those conduced in highly developed countries.

In particular, prior literature reveals that ERP adoptions in transition economies are characterized by lack of strategic approach [10, 15, 17], greater problems with planning activities and resources [13], greater impact of people and their roles [16, 18], and expected higher levels of external support [2, 14].

As a result, we might hypothesize that the scope of this study's findings should cover first and foremost organizations from Poland and other transition economies from Central and Eastern Europe, i.e. countries which belonged to the Communist Bloc after the World War II but recently joined the European Union and are now undergoing economic transition.

7 Conclusion

Summing up the results of the implementation project in Asseco Systems S.A. we should state that the company realized two main goals of the project. First, a solution has been implemented which satisfies the needs of a project-driven organization. The solution has a central repository of data about projects and allows managers to monitor all source operations connected with a given project that are realized in other modules of the ERP system. Thanks to such integration it was possible to create a multidimensional analytical mechanism which is able to supervise the performance of conducted projects in an online mode. The developed solution might be adopted as a reference model for IT industry and possibly other project-driven industries.

Second, optimization and integration of all business processes at the project level have been achieved, which is especially important in a very complex area of services. The transactional interface linking the ERP system with the service system guaranteed service processes continuity and, at the same time, eliminated redundant operations performed by the employees of logistics and service departments. Such a solution resulted in lower costs of service processes and, at the same time, guaranteed a desired speed of information processing. This is especially important because fast information processing is crucial from the point of view of the most significant parameter of service quality – disaster recovery time.

Upon analyzing the evolution of management of project-driven organizations and their specific information needs, we may expect a fast development of project-related functionality within ERP systems. Such a progress might lead to the development of a new family of enterprise systems: Project Resource Planning.

References

1. Bellini, E., Canonico, P.: Knowing communities in project driven organizations: Analysing the strategic impact of socially constructed HRM practices. International Journal of Project Management 26, 44–50 (2008)
2. Bernroider, E.W.N., Sudzina, F., Pucihar, A.: Contrasting ERP absorption between transition and developed economies from central and eastern Europe (CEE). Information Systems Management 28(3), 240–257 (2011)
3. British Standards Institute, BS 6079-1: 2000 Project Management: Part 1: Guide to Project Management, BSI Standards (2001)
4. Cleland, D.L., Ireland, L.R.: PROJECT MANAGEMENT: Strategic Design and Implementation. McGraw-Hill Osborne Media (2006)
5. Computerworld TOP 200 Report 2008, 2009, 2010. Ranking firm IT. IDG, http://www.computerworld.pl/top200/
6. Hans, E.W., Herroelen, W., Leus, R., Wullink, G.: A hierarchical approach to multi-project planning under uncertainty. Omega 35, 563–577 (2007)
7. IFS, How Well Does ERP Facilitate Project Based Business Models?, http://download.ifsworld.com/home/if1/page_877/how_well_does_erp_facilitate_project_based_busines.html (accessed April 19, 2013)
8. Keegan, A., Turner, J.R.: The Management of Innovation in Project-Based Firms. Long Range Planning 35, 367–388 (2002)
9. Klaus, H., Rosemann, M., Gable, G.: What Is ERP? Information Systems Frontiers 2(2), 141–162 (2000)
10. Ketikidis, P.H., Koh, S.C.L., Dimitriadis, N., Gunasekaran, A., Kehajova, M.: The use of information systems for logistics and supply chain management in South East Europe: Current status and future direction. Omega 36(4), 592–599 (2008)
11. Reich, B.H., Gemino, A., Sauer, C.: Knowledge management and project-based knowledge in it projects: A model and preliminary empirical results. International Journal of Project Management 30, 663–674 (2012)
12. Roztocki, N., Weistroffer, H.R.: From the special issue editors: Information technology in transition economies. Information Systems Management 28(3), 188–191 (2011)
13. Soja, P.: Difficulties in enterprise system implementation in emerging economies: insights from an exploratory study in Poland. Information Technology for Development 14(1), 31–51 (2008)
14. Soja, P.: Understanding determinants of enterprise system adoption success: lessons learned from full-scope projects in manufacturing companies. Production Planning & Control 21(8), 736–750 (2010)
15. Soja, P.: Examining determinants of enterprise system adoptions in transition economies: insights from Polish adopters. Information Systems Management 28(3), 192–210 (2011)
16. Soja, P., Paliwoda–Pękosz, G.: What are real problems in enterprise system adoption? Industrial Management & Data Systems 109(5), 610–627 (2009)

17. Soja, P., Paliwoda–Pękosz, G.: Comparing Benefits from Enterprise System Adoption in Transition and Developed Economies: An Ontology-based Approach. Information Systems Management 30(3), 198–217 (2013)
18. Themistocleous, M., Soja, P., Cunha, P.R.: The same, but different: enterprise systems adoption lifecycles in transition economies. Information Systems Management 28(3), 223–239 (2011)
19. Turkulainen, V., Kujala, J., Artto, K., Levitt, R.E.: Organizing in the context of global project-based firm—The case of sales–operations interface. Industrial Marketing Management 42, 223–233 (2013)
20. Wiewiora, A., Trigunarsyah, B., Murphy, G., Gable, G., Chen, L.: The Impact of Unique Characteristics of Projects and Project-Based Organisations on Knowledge Transfer. In: Proceedings of the European Conference on Knowledge Management, p. 888 (2009)
21. WorkPLAN, Project Based Companies need Specialized ERP solutions, `http://www.workplan.com/fileadmin/pdf/myworkplan/WhitePaper_ERP-JobManag.pdf` (accessed April 19, 2013)

Specifying Security at the Systems Analysis Level: Operational, Public-Image and Legal Aspects

Firoozeh Rahimian[1] and Akhilesh Bajaj[2]

[1] Tandy School of Computer Science
The University of Tulsa
Tulsa, OK USA
Firoozeh-rahimian@utulsa.edu
[2] Collins College of Business
The University of Tulsa
Tulsa, OK USA
Akhilesh-bajaj@utulsa.edu

Abstract. Current software security approaches involving software and information assurance, involve security activities such as threat modeling, misuse cases, and rigorous testing during the development, implementation and maintenance phases of the software lifecycle. With OPL (operational, public-image, legal) model, we propose that security requirements should be elicited at the data field level from end-users during the requirements modeling phase of the lifecycle. The elicited classification can then be used to drive the process of identifying critical processes of software, which leads to more effective threat modeling and testing regimens downstream.

Keywords: Systems development lifecycle, security requirements, data security, software trust, software assurance, information assurance.

1 Introduction

In recent years, security has emerged as a critical issue not only when developing new software but also when maintaining existing software. In an effort to protect and secure their data and software from attacks, organizations have focused on protecting their operating environments from the outside perimeter (network protection, authorization/authentication, malware protection) [2]. Despite these efforts, security breaches and software exploitation continue to increase daily.

Implementation of policies and procedures to achieve and maintain high trust levels in software has become a necessary requirement of the software development process. Building software in today's environment is complex and expensive, with many new and conflicting requirements. Eliciting adequate requirements, especially those related to software security, remains one of the most challenging aspects of the software development life cycle (SDLC).

In most applied environments, software processes perform read or write actions on one or more data fields. Without understanding the security impact of each data field,

S. Wrycza (Ed.): EuroSymposium 2013, LNBIP 161, pp. 39–49, 2013.

critical processes of the software cannot be identified. Identification of critical processes is essential to performing effective threat analysis, which allows development of appropriate testing regimens, downstream in the SDLC. It is well known that incomplete testing leads to undetected defects and bugs, which in turn can lead to exploitable vulnerabilities in software [1, 3].

Software security approaches are grouped into two areas: information assurance and software assurance. Information assurance is the measure of confidence that software features meet the requirements as defined by the overall security policy [4]. Software assurance is the "confidence that software, hardware and services are free from intentional and unintentional vulnerabilities and that the software functions as intended" [5]. Risk management involves balancing the security needs and security costs tradeoffs for different software in an organization [6, 7].

It has been recognized that security specifications at the initial stages of the software lifecycle can lead to significant cost savings and better security downstream, both in the construction and implementation phases as well as the long term maintenance phase [8, 1]. In this work in progress, we provide initial proposals of a parsimonious yet comprehensive model that allows users to specify security properties at the granular level of an individual data field. The model termed OPL (operational, public-image, legal) requires users to classify data fields along three orthogonal dimensions. By performing the classification during the requirements modeling phase, critical processes of software can be identified early in SDLC so that i) security measures can be built into software design, ii) threat analysis can be focused on the critical processes, and iii) testing strategies can be built for the critical data and processes. To have the business users assign the classification is another essential factor for the OPL model. Since the business users are the consumers of the data via software and are considered to be the owners of the data, they also serve as the initial information source for where security resources would need to be focused.

The rest of this paper is organized as follows. In section 2, we present a review of current practices for software security. In section 3, we develop the OPL model and its impact to requirements, analysis, and design phases of the software development process. The conclusion and directions for future research are presented in section 4.

2 Background

Currently, several activities related to software security are being performed to ensure that the software produced at the end of the SDLC is as trustworthy as possible. Every security activity performed during each phase of SDLC achieves one of two goals: i) identify the information that needs to be secured and the trust levels (information assurance); and ii) to show how those trust levels will be reached and maintained (software assurance). Through rigorous testing and the analysis of the test results, software features are measured to ensure compliance with the security policies are achieved [4]. Incorporation of various security activities at every phase of SDLC is used to create confidence that software will function as designed [5]. The fundamental principle of information and software assurance is that by securing the

development process and rigorous testing, the software produced will be trustworthy [4, 5, and 14].

There are four major security activities that are currently performed throughout the SDLC process toward meeting software and information assurance requirements. First, *misuse cases* are being used to capture security requirements. They document unacceptable behavior of software [9, 10, and 11]. Misuse cases can fail due to the complexity of software and the fact that a misuser cannot always be identified for a given threat, and not all actions that lead to the occurrence of a threat can be anticipated [11]. Extensive knowledge of software and the security measures that is required are other barriers to creating effective misuse cases [10].

Second, *threat modeling* has become a common methodology for identifying security risks to software [11, 12, and 13]. Threat modeling is the process of understanding the software and its components, and analyzing its access points to identify potential threats [12]. The threats are then reviewed for their occurrence likelihood and risk factor [12, 13]. Software complexity, dependencies on other software, constant changes to the software, and the high-level of expertise required are the major weakness of threat modeling. Since threat modeling is performed after software design is complete, major design flaws that are discovered are generally very costly to fix. Thread modeling can also be rendered ineffective because it is typically performed once during the SDLC whereas new threats emerge continually.

Third, organizations use the *data classification* process to identify the security requirements for their data. Military data classification schemes use unclassified, sensitive-but-classified, confidential, secret and top secret to classify the data [14]. In recent years, these schemes have been adapted for commercial use by classifying the data as private, sensitive, critical, or confidential [14]. Data classification for security purposes is also based on risk impact: low, medium, or high [15]. Data classification schemes adapted from the military do not translate well into the commercial environment. The classification categories and their definitions are too vague for the business users. Many data classification projects fail due to complexity of the scheme, the expense, or are too ambitious to achieve [14, 15]. All of these challenges lead to software security requirements that are incomplete or inadequate.

Fourth, organizations are conducting more *rigorous testing,* ex post. Testing includes positive testing (verifying software performs as it should) as well as negative testing (determining that software does not perform as expected) [1]. Penetration testing, dynamic analysis, fuzz testing, threat modeling, and surface reduction attacks are also being performed to detect bugs and defects. These tests are used to verify that access is restricted, the operating environment is secure, and the software will recover from failures in a secure manner [8]. Although, software testing has become more rigorous, software complexity and its many internal/external interfaces make it impossible to test for all vulnerabilities and threats. Figure 1 summarizes by phase the various security activities that are being performed as part of SDLC.

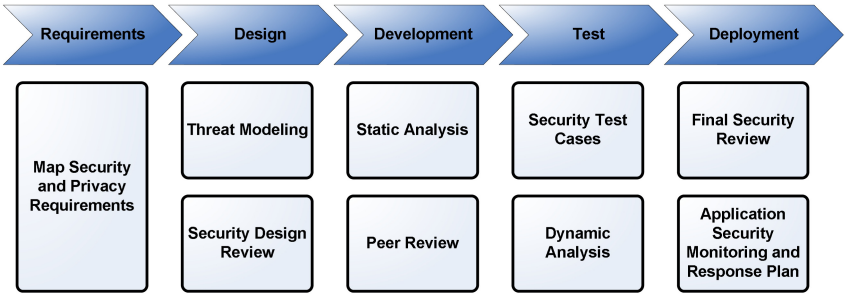

Fig. 1. Software Development Phases and Security Activities [11]

3 Operational Public-Image Legal Model

The basic premise of the OPL model is that the security impact of the data fields should drive the security impact of a process on the business it supports. Intuitively, this is how the critical processes of software should be determined. As the owners and consumers of the data fields, the end-users, along with appropriate legal and strategic level input from other organizational players, must determine the security impact of the data during the requirements and analysis phase so that security solutions can be built into the software design. Fixing software vulnerabilities and design flaws is much more expensive during the construction and testing phases. Figure 2 shows how the data security classification process at the data field level fits into the SDLC phases

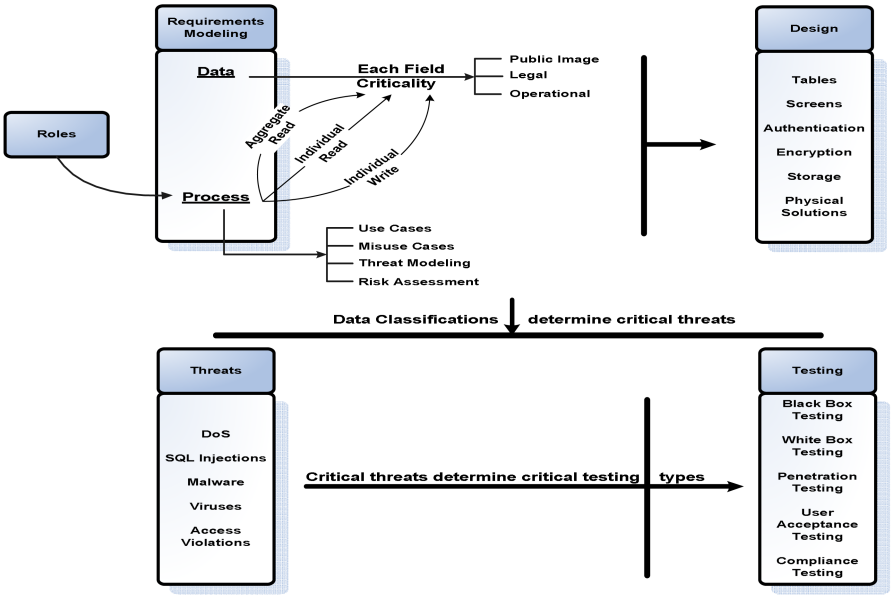

Fig. 2. Data and Process Relationship in Software

to identify threats. As the figure indicates, the OPL model requires specifications of each data field along three orthogonal dimensions. This classification is then used to determine the security classification of each process, which would in turn drive the resources downstream used in the design, construction, testing and maintenance phases of a typical lifecycle. Next, we describe the OPL specification in some detail.

3.1 Data Security Classification

The OPL classification consists of a tuple, with each element of the tuple comprised of two parts, the *security dimension* and *impact classification*. The security dimension identifies an area of impact. OPL considers three dimensions: public-image, legal, and operational.

- **Operational (O):** Organizations must protect their networks and software against attacks and exploitation. Customer and employee information must be secured against theft, and all internal/external facing interfaces must be monitored against attacks. Breakdown of security along this dimension will degrade the internal functioning and operational competitiveness the organization. Impacts on this dimension will be felt at the tactical and strategic levels of the organizations.
- **Public-image (P):** Organizations spend a great deal of time and resources to establish a public image that invokes trust with their customers in order to sell their products and services. Any damage to the public image will break the trust between a company and its customers, causing a negative financial impact [5].
- **Legal (L):** Lack of compliance with regulations and industry standards may lead to litigation or fines. Any damage to customers due to unauthorized use of company products and services may also be grounds for litigation [7].

These three dimensions are orthogonal to each other in that it is straightforward to conceptualize examples where the impact is only along one dimension and not the others. For example, customers' personal information and credit card numbers together would constitute risks along the public-image dimension and possibly along the legal dimension, but not along the operational dimension. Similarly, product drawings with high intellectual capital would have higher risk along the operational dimension but not along the P or L dimensions.

For each of these dimensions, we specify three potential levels of impact.

- **Low (L):** There is marginal or negligible impact [15]. Nothing can be gleaned from the data in the event of a security breach. Organizations are willing to assume the risk in this scenario.
- **Medium (M):** There is some level of impact [15]. There is potential risk that some sensitive data is exposed. There is a possibility that the network is attacked but not penetrated. Organization will perform risk assessment and implement contingency plans or solutions to prevent future attacks. An organization's business is not interrupted in the event of a security breach.

- **High (H):** An organization is not able to do business in case of a security breach [15]. Some examples of high impact security breaches are DoS and SQL injection attacks to steal sensitive data.

While it is possible to specify more fine-grained levels, this leads to greater cognitive burden on the end-user. The goal is to balance parsimony of the OPL model with sufficient detail in the specification. Three levels is a reasonably accepted level of granularity when specifying impact levels [15].

In the OPL model, each data field gets assigned a classification tuple. Thus, for example, every field dealing with, say, employees, such as *employee_id, employee_first_name*, and so on, gets its own classification. Each element of the classification tuple is a pair, (OPL dimension, impact) from each of the dimensions listed below. So, for example, *employee_id* may be classified as *(P,L), (O,H), (L,H)* meaning that if information pertaining to that field were compromised, the public –image risk would be low, but the operational and legal risks would be high. This classification is done based on a subjective assessment by end-users of the data aswell as strategic and legal experts in the organization.

- Public (P): (P,L), (P,M), (P,H)
- Legal (L): (L,L), (L,M), (L,H)
- Operational (L): (O,L), (O,M), (O,H)

The formal Backus Naur form of the OPL grammar is shown in figure 3. Next, we discuss application of the OPL model in a case study.

3.2 ACME Case Study

ACME a realistic description of a typical small manufacturing enterprise's sales and marketing, and engineering and manufacturing departments. This has already been used in a large scale requirements modeling exercise documented in [17]. ACME was chosen to show the application of the data security classification since it is a representation of a typical organization with products being sold to customers as well as internal processes for producing the products. The ACME data model involves over 60 tables of data and several columns in each table. The DSD (Data Screen Design) tool [18, 19] was used to model the data fields for the sales, marketing, and operation departments. The data security classification was then performed to assign the data security classification tuple to all the fields of all the tables. All data fields that were referenced by other tables inherited their data security classification tuple.

As an example snippet, Table 1 depicts the data security classification applied to the EMPLOYEE table for ACME. In the EMPLOYEES table, the *employee_titles_id* field was classified as part of another table, EMPLOYEE_TITLES. It retained its classification from the EMPLOYEE_TITLES table.

```
[ ]        Content within the brackets is optional

{ }        Content within the braces can be repeat 0 or more times

D          Represents the data field
           Data field name consists of <table name>.<field name>

T          Represents the terminals of the grammar

V          Represents the non-terminals of the grammar

S          Represents the start

P          Represents the productions for the grammar



T =   {   ",", "(", ")", ".","0" , "1", "2", "3", "4", "5', "6", "7", "8", "9",
          "a", "b", "c", "d", "e", "f", "g", "h", "i", "j", "k", "l", "m", "n",
          "o", "p", "q", "r", "s", "t", "u", "v", "w", "x", "y", "z", "A",
          "B", "C", "D", "E", "F', "G", "H", "I", "J", "K", "L", "M", "N",
          "O", "P", "Q", "R", "S", "T", "U", "V", "W", "X", "Y", "Z",
          "(P,L)", "(P,M)", "(P,H)", "(L,L)", "(L,M)", "(L,H)", "(O,L)",
          "(O,M)", "(O,H)"
      }

V =   {   D, name, security_dimension, digit, letter, number    }

P =   {
          S                  =  model
          model              =  "(" D ")" | model
          D                  =  name "," security_dimension
          security_dimension =  "(P,L)"|"(P,M)"|"(P,H)" ","
                                "(L,L)"|"(L,M)"|"(L,H)" ","
                                "(O,L)"|"(O,M)"|"(O,H)"
          name               =  letter {letter|number}.letter {letter|number}
          number             =  digit {digit}
          digit              =  "0"|"1"|"2"|"3"|"4"|"5'|"6"|"7"|"8"|"9"
          letter             =  "a"|"b"|"c"|"d"|"e"|"f"|"g"|"h"|"i"|"j"|"k"|"l"|"m"|"n"|
                                "o"|"p"|"q"|"r"|"s"|"t"|"u"|"v"|"w"|"x"|"y"|"z"|"A"|
                                "B"|"C"|"D"|"E"|"F'|"G"|"H"|"I"|"J"|"K"|"L"|"M"|"N"|
                                "O"|"P"|"Q"|"R"|"S"|"T"|"U"|"V"|"W"|"X"|"Y"|"Z"
      }
```

Fig. 3. Grammar for OPL Model

The application of the OPL model in the overall ACME case, along with conversations with real-world security analysts, reaffirmed that the security dimensions (operational, public-image and legal) are based on concepts that apply to any organization. They represent the reasons why data is important to an organization

based on their business use or purpose. The security impact levels (low, medium, high) are comprehensive but not too granular that the users are not able to apply to the security dimensions.

Table 1. ACME: EMPLOYEES Table

Data Field	Data Security Classification Tuple
employees_cell_phone	(P,H), (O,H), (L,H)
employees_email_addr	(P,M), (O,M), (L,H)
employees_firstname	(P,H), (O,H), (L,H)
employees_home_addr	(P,H), (O,H), (L,H)
employees_home_phone	(P,H) (O,H), (L,H)
employees_id	(P,L), (O,H), (L,H)
employees_lastname	(P,H), (O,H), (L,H)
employees_office_location	(P,L), (O,H), (L,M)
employees_office_phone	(P,L), (O,M), (L,M)
employees_ssn	(P,H), (O,H), (L,H)
employees_supervisor_id	(P,L), (O,H), (L,L)
employees_zip	(P,L), (O,L), (L,L)
employee_titles_id	(P,L), (O,M), (L,L)

The data security classification process is a very human intensive process, which means it must be easy to understand and simple to apply, but granular enough to capture serviceable information. The data security classification will remain with the data field once it has been assigned throughout the data's life cycle – from the time data is created through to its destruction. The data security classification can also be utilized to determine the decomposition requirements of data.

4 Discussion

Successful implementation of the OPL model relies heavily on the end-users' understanding of each dimension. Hence it is important to use a parsimonious and orthogonal set of dimensions. The OPL dimensions representing the operational, legal and public-image aspects of impacts of security compromised of a data field fulfill these requirements. While the end-users of each data field provided the levels for each of the three dimensions, it is appropriate to get input from different experts such as legal and strategic managers, when evaluating the impact of security breakdowns for each data field.

Once the security classification for each data field has been specified, derivation of the security impact of each process becomes possible. This is because software processes handle multiple data fields from various tables. One possibility of deriving the classification of each process is by looking at the highest level of classification of all the data handled by that process. For example, a single high impact data field from one table may cause a process to be flagged as critical even though all other data fields may not be high risk. Once critical processes have been identified, then adequate security resources can be allocated for each process. Threat modeling can be concentrated on critical processes, which can result in more effective attack identification. There is also the potential to conduct multiple threat modeling drills throughout SDLC. Testing strategies and methodologies can be tailored for the potential attacks, as well as more rigorous testing for the critical processes.

In this work, we have discussed only classification of each data field individually. Another avenue for analysis is the derivation of the security impact of sets of attributes (or fields) from the classification of each field in the set. This is part of our ongoing research.

While we tested out the OPL model for a single case, real world systems consist of much larger numbers of data fields, accessed by multiple users. The process of getting users to jointly specify security levels for each data field that they use along each dimension is non-trivial. A large body of work on group decision making exists, and several avenues are possible. For example, it may be possible to get all end-users to independently specify the levels for each data field along each dimension, and then along each dimension, pick the highest level or the most common level specified for each data field.

5 Conclusion

In this research in progress, we provided initial specifications for a systems analysis level model that allows end-users to specify the importance of each data field they use, from a security standpoint. Users can specify the importance using three levels (high, medium or low) along three dimensions (operational, public-image and legal). In order to reduce cognitive load on the end-user, the OPL model uses a parsimonious and orthogonal set of dimensions. Testing of the OPL model using an illustrative case study indicated that it is practicable in real world settings.

Until now, security testing and hardening has been largely ex-post, after the system has been designed, constructed or even implemented. The OPL model attempts to provide a foundation for security specification that emanates from the context of the end-users, before the system is designed, during the analysis phase itself. Specification of security at the data field level by end-users provides the bedrock for derivation of security classification of business processes that handle data. This in turn allows a more meaningful threat analysis and management, so that limited resources can be allocated more optimally for processes that have higher threat ratings derived from the data fields they touch.

The OPL model specification proposed here is part of a larger project that aims at providing a systems analysis foundation for the security specification of database systems in organizations. Our future work involves providing more complex derivations of process security classifications using data field specifications as building blocks. Second, we aim to address a more precise mapping between the analysis level specification and the downstream security activities in the design, construction, implementation, and maintenance phases of the information system.

References

1. Owens, D.: Integrating Software Security into The Software Development Lifecycle, IMPACT, http://www.impact-alliance.org (accessed 2013)
2. Dynamics, General. Venturing Beyond the Castle Walls – The Need for Data-Centric Security Models in Cloud Computing Environments, General Dynamics Information Technology (2012), https://meritalk.com/uploads_resources/000081_4435.pdf
3. Pfleeger, C.P., Pfleeger, S.L.: Security in Computing, 4th edn. Prentice Hall, Indianapolis (2006)
4. Rauch, M.: What is Information Assurance, Articlesbase (2009), http://www.articlesbase.com/security-articles/what-is-information-assurance-1142179.html
5. SAFECode, Software Assurance: An Overview of Current Industry Best Practices, Wakefield, Massachusetts, USA (2008), http://www.safecode.org
6. Dash, R.: Risk Assessment Techniques for Software Development. European Journal of Scientific Research 42(4), 629–636 (2010), http://www.eurojournals.com/ejsr.htm
7. Le Grand, C.H.: Software Security Assurance: A framework for Software Vulnerability Management and Audit. CHL Global Associates (2005), http://www.ouncelabs.com
8. Microsoft, Microsoft Security Development Lifecycle, Simplified Implementation of the Microsoft SDL (2010), http://www.microsoft.com/sdl
9. Williams, L.: Misuse (or Abuse) Cases, North Carolina State University, North Carolina, United States, http://www.cigital.com/justiceleague/wp-content/uploads/2007/07/touchpoints.gif (accessed 2013)
10. Steven, J.: Defining Misuse within the Development Process. IEEE Security and Privacy, United States (2006)
11. Sindre, G., Opdahl, A.L.: Eliciting security requirements with misuse cases. Requirements Eng. 10, 34–44 (2004)
12. Myagmar, S., Lee, A.J., Yurcik, W.: Threat Modeling as a Basis for Security Requirements. In: Symposium on Requirements Engineering for Information Security (SREIS), United States (2005)
13. Howard, M.: Demystifying the Threat-Modeling Process. IEEE Security and Privacy, United States (2005)
14. Etges, R., McNeil, K.: Understanding Data Classification Based on Business and Security Requirements. Journal Online 5 (2006)

15. Heiser, J.: Data classification best practices: Techniques, methods and projects, http://www.SearchSecurity.com (accessed 2013)
16. Verizon, Data-Centric Vulnerability Management (2012), http://www.verizonenterprise.com/resources/whitepaper/wp_data-centric-vulnerability-management_en_xg.pdf
17. Bajaj, A.: Large Scale Requirements Modeling: an Industry Analysis, A Model and a Teaching Case. Journal of Information Systems Education, United States (2006)
18. Bajaj, A. , Large Scale Requirements Modeling, University of Tulsa, United States (2008).
19. Bajaj, A.: The Effect of Abstraction of Constructs in Data Models on Modeling Performance: An Exploratory Empirical Study. In: Americas Conference on Information Systems (ACMIS), United States (2010)

Identifying Factors of an Information Security Management System of Local Self-government Bodies

Przemyslaw Jatkiewicz

University of Gdansk, Department of Business Informatics
Piaskowa 9, 81-864 Sopot, Poland
jatkiewicz@ip-info.pl

Abstract. This article presents synthetically expressed results of studies on information security management system in Polish local self-government bodies. Standardized surveys, internal network scans and penetration tests of web portals were carried out in five institutions. These were municipal offices and local budgetary companies. Research surveys were conducted among all employees of the aforementioned organizations.

The research allowed identifying key factors and proposing methods to solve the most common problems faced by local self-government bodies. Moreover, the paper presents a model of a system supporting information security management developed based on the identified factors.

Keywords: Security information, local self-government, security management.

1 Introduction

Performance of task by local self-government units, as well as commercial enterprises, requires processing of related data sets. Whether there is a legal duty to protect them or not, the process of collecting requires involvement of an organization's resources with a measurable value. Losses that may result from information security violation are difficult to estimate. They involve not only the loss of the organization's operation flow and a need for extra resources in order to remove their effects, but also the loss of an intellectual property, damages and legal consequences. A well-known software company engaged in computer system security software, McAfee, based on studies conducted by CERIAS (Center for Education and Research in Information Assurance and Security) among more than 800 heads of IT in the U.S., UK, Germany, Japan, China, India, Brazil and Dubai, estimates that on a global scale, the amount of companies' loss due to security violation has exceeded one trillion U.S. dollars[1]. The assessment does not include such an important factor as reducing future benefits as a result of confidence loss.

Information systems are increasingly exposed to unwanted interference with the increasing awareness of market information value and its influence on a competitive advantage. The 'Personal Internet Security report'[2] presented in the House of Lords of Great Britain in 2007, mentions a price of 100 pounds per record containing

S. Wrycza (Ed.): EuroSymposium 2013, LNBIP 161, pp. 50–65, 2013.

personal information. Although the price seems fairly high it gives an idea how great the temptation of stealing personal data is.

State and local government organizations process the most comprehensive range of information on all spheres of life. To satisfy the still growing computer society, more and more numerous and complex systems are formed.

Given the growing importance of information systems in public institutions and the diversity of government and business purposes, it seems appropriate to undertake research on information security system management that is specific to local self-government bodies (the largest group of public bodies).

It is assumed that the variety and availability of information technology is sufficient to build a system with a high degree of security. Moreover, its development is appropriate in relation to the needs of local self-government organizations. However, the belief in the effectiveness of technical measures puts a shadow on an important factor - human resources and their work organization methods[3].

2 Methodology of Research

The aim of the presented research was to identify factors of information security system management of local self-government bodies to develop a model of the system. Multiple case studies were used[4], including:

- Documentation analysis,
- Standardized in-depth survey,
- Questionnaires,
- Penetrative surveys of information portals,
- Internal network scans.

Documentation analysis is obligatory to identify the principles to follow by an organization when managing information security. Interviews and research surveys provide information about their knowledge and application. Tests and scans are used to verify them. They limit the influence of employees' natural tendency to show the represented organization in a better light.

Modeling techniques based on notation of such languages as SysML[5] and UML[6] were used to prepare an original model to support information security system management. Use Case Diagrams (UCD) and Requirements Diagrams (RD) were used.

The adopted methodology is similar to that presented by Shao and Jensen when studying American state institutions' websites[7]. However, it has been extended to include surveys for computer systems users. Work on the security model have also been conducted by Karokola, Kowalski, and Yngstrom. They are based on surveys on a very limited number (18) of respondents who are employees of six government institutions in Tanzania[8]. Richer research techniques with a wider range were used in the presented research studies.

Multiple case studies included 5 bodies of various sizes and located at different levels of local self-government - two Local Self-Government Budget Bodies,

3 Municipal Offices, including one in a city with county rights. All the institutions were located on the territory of Poland. Due to the security of the analyzed systems, the names of entities using them have been replaced by letters A to E. The object of the study was to determine the quality of an information security management system, identify formal and informal processes carried out as part of the system and define its conditions.

3 Course of Research

The research conducted in the years 2010-2012 began with in-depth interviews with respondents. Questions formulated by the author related to protections contained in Annex A to ISO 27001 standard which is recommended for information systems used by public entities to carry out public activities. The degree of implementation of its guidelines provides a clear view of both the degree of system security and meeting regulatory requirements.

The surveys were conducted with the owners of the information security management process. This function was usually fulfilled by administrators, IT managers or Information Security Administrators. They were preceded by an analysis of such documents as security policy and operational procedures. It helped narrowing down the amount of questions and focusing on implementation of provisions contained in the document. In addition to questions associated with the aforementioned standard, the author tried to obtain information on the size of a body, a separate organizational structure for managing information security and an information system size. The survey included 13 thematic areas listed in Table 1.

Each area consisted of main and specific questions formulated based on ISO 27002 standard. Not every question was asked during the survey because the results of document analysis or answers to the main questions provided the necessary information. Therefore, the wide range of descriptions of the surveys for each organization varied.

The next stage of research was polling officials. The questionnaires were distributed to all employees, i.e. 893 people. They explained the anonymous nature of the study and the purpose for which they were conducted. Moreover, the author's telephone number and e-mail were provided in order to clarify any doubts. A comprehensive manual explaining how to answer to particular questions was attached to the survey. The survey form included questions about:

- age,
- gender,
- education,
- work position,
- knowledge of the Security Policy
- assessment of security of the operated information system,
- assessment of impediments at work related to elements of the security system,
- probability of occurrence of specified threats
- number of logins and passwords used,
- authentication methods and techniques that facilitate work,
- sharing logins and passwords with other employees.

Table 1. Thematic areas of the in-depth survey (*Source: own study)*

Thematic area	**Description**
Risk analysis	A method for conducting a risk analysis and making resource estimates
Security policy	Content of documents related to information security and methods for their updating
Organization of information security	An organizational structure connected with information and task security, responsibility for means of information processing.
Assets management	Responsibility for ICT assets and classification of information
Human resource security	Rules for dealing with information system users – training, responsibility, control, authorization
Environmental and physical security	Methods of protecting equipment, facilities and installations against environmental and physical factors
System and network management	Rules for using information systems
Access control	Granting, accepting and controlling access to means of information processing
Obtaining, developing and maintaining information systems	Requirements connected with security of obtaining, changing and maintaining elements of information systems
Managing incidents connected with information security	Rules for collecting information on incidents and methods of dealing with them
Managing the continuity of activity	Methods for elaborating and controlling plans of continuity of activity
Conformity	Conformity to law and internal regulations

527 questionnaires were returned which is over 59%. This low return rate can be explained by the fact that the respondents did not pay attention to the survey despite being assured that it was anonymous. They were afraid to express their own opinion.

Due to the relatively low return rate and numerous errors in the surveys, the overall studies contain a high number of errors that can be classified as errors of completeness, a subgroup of errors resulting from failing to obtain data and errors of content, a subgroup of response errors[9].

The conducted penetrative tests of web portals were supported by free software Netsparker 1.5. Despite differences of opinion on effectiveness of vulnerability detection by automated computer tools, the use of a uniform method based on them is sufficient to make a comparative assessment of website security of the chosen institutions[10]. Detected threats were divided due to their importance, according to a classification set out in Table 2. In addition to website vulnerability, the research studies included tests of technical execution correctness by performing parsing with the use of the W3C Validator tool and checking reference links up to 3 levels deep[11].

Table 2. Classification of threats (*Source: own study)*

Threat class	Threat	Description
Critical	SQL injection	This threat allows the attacker to carry out his own SQL commands and obtain a possibility of unauthorized access to data, including its modification or deletion.
Important	Code display	A possibility of a source view giving an insight into the techniques used and in extreme cases information about establishing a connection with the database.
	File transfer	A poorly controlled possibility to upload files to a server.
	Sending passwords through an unsecured channel	Sending passwords using an unsecured protocol.
	Server status view	Access to the web server status containing crucial information.
	Referred XSS	A variant of XSS. Injecting a malicious code that is stored in the database.
	Out-of-date software having holes	The software found has commonly known security holes.
Average	XSS	Injecting a malicious code into a website
	Out-of-date software having holes	The software found has commonly known security holes.
Low	Cookie files not marked as HTTPOnly	A possibility of an unauthorized download of cookie files.
	Remembering passwords in forms	Access authorization to a portal is strictly dependent on access control to an operating system. Ineffective when using equipment that is available to the public.
	View of a server version and its modules.	A possibility of obtaining information about security holes in specific implementations.
	Non-recommended http methods	Methods other than POST and GET that are crucial during normal page browsing.
	IP addresses of an internal network	Gaining information on the internal network's architecture.
	Backup files	Copies of files that can contain potentially dangerous information.
	A possibility to trigger an error message	Bringing instability of the system and obtaining information about mechanisms used.
Information	Unsecured e-mail addresses	A possibility to send spam.
	Browsing catalogue content	The portal was made carelessly.

Internal networks scans of each body were also made. They aimed at detecting the weakness of their protection against attacks from within the organization. They can be carried out by dishonest employees, contractors performing work on the premises of a company or other people who have gained direct access to the technical infrastructure of the system due to organizational deficiencies. They were carried out using a free version of a network scanner - GFI LanGuard[12]. It can detect more than 15.000

threats on Windows, Linux and Mac OS, including virtual machines. It downloads updated information on vulnerabilities from such databases as SANS Top 20[13] developed by SANS Institute.

4 Research Results and Conclusions

The conducted research provides a basis for expressing an opinion that local self-government bodies protect the data they process to an inadequate degree. It should also be noted that modern, and therefore costly, technological solutions are rarely used by less affluent institutions. However, it is difficult to judge the validity of such savings as it can be easily seen that organizations typically do not implement security systems.

The planning stage that includes setting goals, identifying problems, finding solutions and evaluating consequences, is nothing but making a risk analysis[14]. Out of the five organizations studied, only organization A had developed such a document, vital to information security. However, even in organization A, risk analysis was made based on own knowledge which does not guarantee its objectivity and reliability.

Therefore, a comparison of advantages and disadvantages of risk analysis was made by own staff and external entities. The result is shown in figure 1.

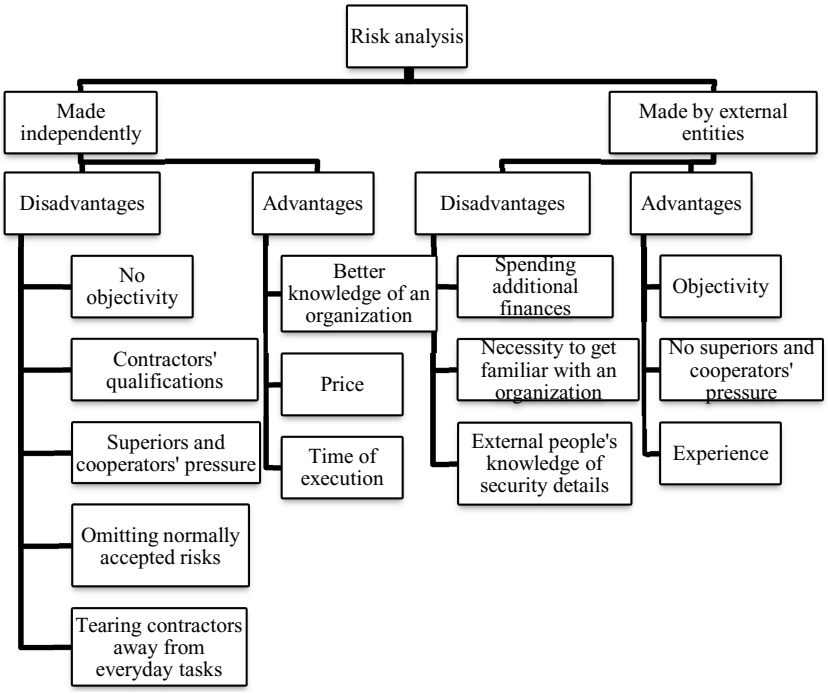

Fig. 1. Advantages and disadvantages of risk analysis methods (*Source: own study)*

The completed research studies were original. Despite difficulties in obtaining permission each time when trying to conduct them, in-depth surveys were carried out and their results were confronted with appropriate tests. They were similar to audits. However, the main difference was the investigator's objectivity and impartiality ensuring that the results will not result in legal or business consequences. A quiet conversation or sharing own insights and experiences made the respondents more open and less likely to disclose information that will not be made public. Results of surveys are presented in Table 3 in a synthetic manner. The table presents the implementation degree of security contained in Annex A to ISO 27001 standard.

Table 3. Synthesis of in-depth surveys *(Source: own study)*

Organization	% of implemented security from ISO 27001 standard
A	79,02
B	53,15
C	10,49
D	32,17
E	32,17

A natural tendency to overestimate the importance of implemented security has been verified over penetrative tests of technical components of an information technology system.

A total of 463 network nodes were scanned and 2 787 threats were found. This gives an average of about 6 per single node. Detailed quantitative data on nodes and threats in every tested organization are shown in Table 4. Organization A was not included, which refused permission for scanning due to the type of security used in the institution. The scanning procedure could interfere with the institution's information system.

Table 4. Network nodes and their threats in the tested organization *(Source: own study)*

	Org. B	Org. C	Org. D	Org. E
Number of nodes	315	60	48	40
Number of threats	2 198	183	358	48
Average number of threats per node	7	3	7	<1
Median	10	1	10	0
Maximum	66	66	11	20

Analyzing the data from the table, it is easy to see a big difference between the median and the average number of threats per single network node. The only exception to this observation is the result of organization D. A more detailed monitoring that was conducted helped to establish that in organizations B, C and E, all the most vulnerable network nodes are computers used by IT engineers. Paradoxically, therefore, in terms of information security, IT teams are the most vulnerable group.

Another type of research presented in the paper concerns security tests of internet portals. Occasionally, they detected critical threats. It would seem that the absence of

such threats confirms the safe nature of a portal. Such an approach ignores the importance of other threats. Therefore, it was decided to impart an appropriate weight to them. Assuming that each threat class is two times more important than the previous one, their weight was shown in values in Table 5.

Table 5. Significance of threats of particular classes *(Source: own study)*

Class index [j]	Threat class	Weight [Wg]
1	Critical	16
2	Important	8
3	Average	4
4	Low	2
5	Information	1

Degree of threat Sz is calculated with the following formula:

$$Sz = \sum_{j=1}^{k} W_j \cdot n_j \tag{1}$$

Where:
Sz – degree of threat
W_j – weight of 'j' threat class,
n_j – number of threats of 'j' threat class,
k – number of threat classes. In the above example k=5.

The degree of threat calculated this way for the tested portals is shown in values in table 6.

Table 6. Degree of threat of tested web portals *(Source: own study)*

Organization	Sz
A	1 576
B	87
C	60 319
D	353
E	159

All the portals, except for the portal of organization B, were designed by third parties. Both the number of threats detected and the calculated degree of threat indicates that no-control outsourcing is not a good solution.

Despite significant differences between the tested organizations, many common features that have a clear influence on the security of information systems used and portals have been identified. Even the condition of buildings occupied by the organizations limits the use of new technologies. All the studied institutions occupy old buildings which are hard to modernize and very difficult to expand (if possible). The growing demand for office space is met by adapting previously unused rooms. ICT and support systems are connected to the nearest point possible. As a result, network and power architectures that are optimal in terms of activity and servicing are not

applied. None of the organizations maintained a star topology network and a centralized backup power. Low-cost, replacement equipment was used.

The respondents interviewed, who were usually those most involved in the management process of information security pointed not only to the lack of financial resources but also the lack of management board's support. Organizational bodies, responsible for information security in organizations, employed few workers.

The low involvement of middle-ranking management in information security management results in insufficient supervision over employees' activities and limits access to data on incidents. In all the studied organizations, IT managers are burdened with the highest number of activities aimed at maintaining security of operated systems.

Reducing the number of employees who affect the information security management process translates into a relatively low awareness of its importance. The relatively high sense of system security despite the disturbing, actual state of affairs is also of great significance as illustrated in the following table.

Table 7. System security assessment and knowledge of the Security Policy, as well as following its assumptions by employees *(Source: own study)*

Organization	Does not know the policy	Does not follow the rules of the policy	Security assessment median (1÷10)
A	16%	42%	8
B	73%	46%	7
C	54%	51%	8
D	67%	33%	8
E	74%	24%	8

The further part of the analysis aimed at obtaining answers whether features such as knowledge of the Security Policy, compliance with its provisions and system security assessment depend on factors such as gender, education, age and a position held in an organization. Self-government organizations are usually feminized institutions employing high-educated, at least middle-aged people. The statistical methods used such as cross tables, Chi^2 Pearson tests, V-Cramer and Friedman rank tests did not allow to determine correlations that are common for all the studied bodies. This would enable to determine a group to focus on when undertaking activities to improve the aspects related to information security[15].

It should be noted that the data presented relate only to employees of the studied bodies. However, they are not the only users of information technology systems. Apart from organization A, the security policy of other subjects did not include other persons providing services for these organizations. Only marginally they related to sub-contractors or contractors. This translated into the quality of services provided by contractors as evidenced by the technical quality and degree of threat of portals.

The above-mentioned poor quality of the technical components of the system is partly caused by the provisions of the Public Procurement Law, which favors contractors that offer products or services at the lowest prices. It is extremely difficult to make a public procurement that would guarantee high quality and timeliness of its

implementation. The mentioned provisions do not allow determining the brand and often the technology of an ordered product or service.

IT departments of the studied bodies have a very wide range of duties. Owning a relatively small staff group, they are responsible for:

- making plans for the development of computer systems,
- making and settling for procurements associated with computer systems,
- training,
- supporting end users,
- supervising the implementation of third-party services,
- administration and maintenance of computer systems,
- creating new system elements,
- repairing system components,
- managing computer system security.

Therefore, one department consisting of a few employees is focused on both executive and supervisory functions. They require different competencies, which are hard to develop when offered wages and financial means for training purposes in budget units are so sparse. The role of IT engineers in institutions is quite difficult to define. It is not known whether they are officials or technical service employees. Competence- and technological terminology-oriented employees usually have problems in communicating with senior management and justifying the need to implement necessary solutions in an effective way. Great examples are the failed attempts to implement authority limitations in network traffic. Reluctance, and consequently, pressure exerted by workers prevented their implementation.

Surfeit, complexity and diversity of information services' responsibilities often results in poor execution. No test environment is created for newly implemented solutions. Updates are installed automatically without any supervision of their influence on the whole system. IT engineers learn about errors by accident during repairs or when providing users with support. In most cases, documentation of work conducted on computer systems, both regular and occasional, is not kept.

The tested institutions have virtually no control over the use of removable storage media. Therefore, there are not aware of its content and where it is kept. Therefore, it is not known what protection methods are used and applied to storage media. Used hard drives of computers and servers are protected, though some of the organizations do not have a proper idea how to destruct and dispose them. During the studies, no established rules of conduct of storage media installed in other devices, which can be sources of important information for each institution, were found. Disposal of these devices without removing storage media can end up in loosing important information processed by an organization. Such information can pose a threat to network security, institutions and even the state.

Most of the tested institutions do not keep cyclical maintenance of ICT equipment. Only organization E recognizes hardware failures as a highly likely threat. As shown by a global study conducted by Kroll Ontrack, as many as 29% of data loss is caused by failures[16]. Although failures are difficult to predict, regular maintenance significantly reduce the frequency and probability of their occurrence.

The above mentioned problems do not constitute a complete list. However, they are common to most bodies. It is very probable that they occur in other institutions as well. Solving such problems will definitely have a significant effect on improving security of information systems at local self-government bodies.

As a result of the conducted analyses of literature, normative acts, own surveys and case studies, the following main factors of information security system management in local self-government bodies were identified:

- ensuring compliance with standards,
- asset management,
- threat and incident management.

The developed model of a system supporting information security management consists of three subsystems containing 12 thematic areas that result from ISO 27001 standard[17], which are shown in Table 8.

Table 8. Proposed areas of information security management *(Source: own study)*

Factors	**Areas resulting from Annex A to ISO 27001 standard**
Compliance with standards	Risk analysis
	Security policy
	Information security organization
	Compliance
Asset management	Asset management
	HR security
	Environmental and physical security
	System and network management
	Access control
	Obtaining, developing and maintaining computer systems
	Activity flow management
Threat and incident management	Managing incidents connected with information security

The subsystem that supports maintaining compliance with provisions covers the range of regulations and internal documentation of organization, as well as agreements with contractors that contain issues related to information security. Therefore, legal applications, explain the issue in a narrow sense. On the other hand, most of their provisions do not apply to information security. Thus, their complementation with such documents is unjustified. The following figure (Figure 3) presents a subsystem requirement diagram that supports maintaining compliance with regulations.
The main functional requirements of the proposed subsystem are:

- inserting new documents
- amending or updating regulations,
- presentation,
- providing information.

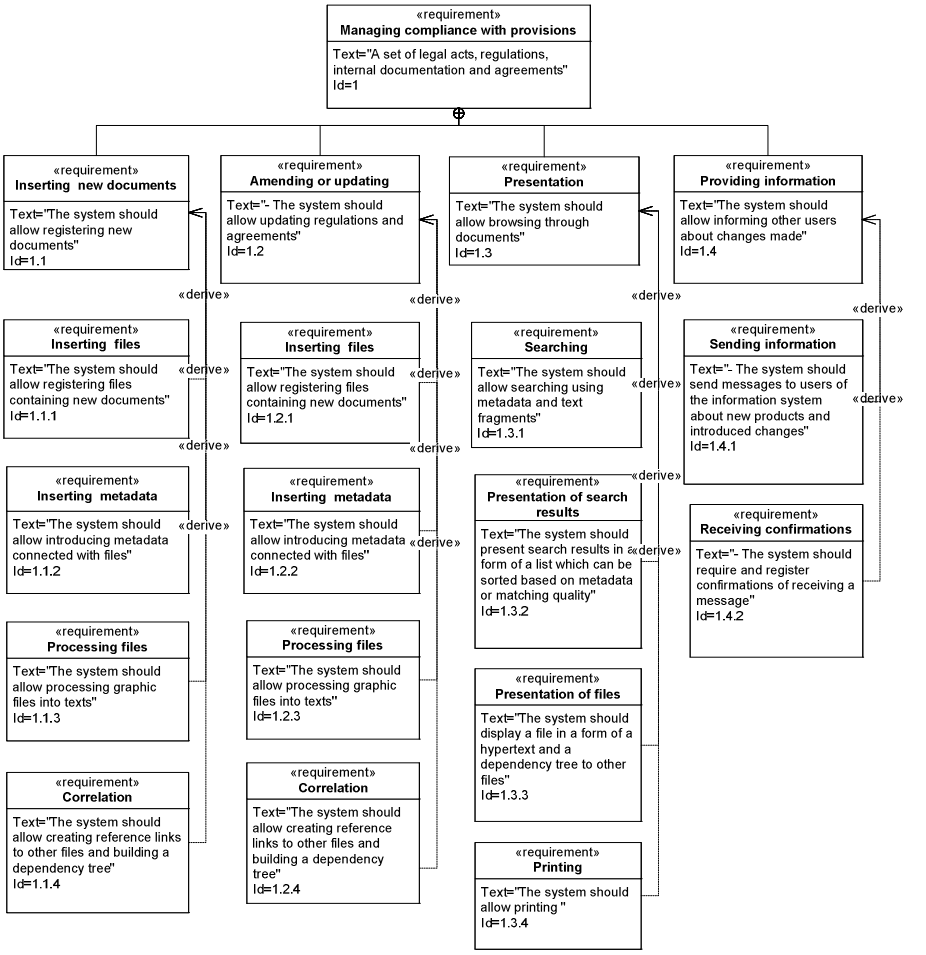

Fig. 2. Subsystem requirement diagram that supports maintaining compliance with regulations *(Source: own study)*

The most important functional requirement is the applied presentation of documents contained in the database. It must include a possibility of browsing through documents by using titles, thematic areas, keywords, dates of application or parts of texts. Results should be presented in a form of a source or a single text as a hypertext with relevant references. It is recommended to allow the user to navigate through the tree structure of references to find all related materials.

For each resource of an organization, the following stages of its operation can be distinguished. They become the basis for defining functional requirements of a subsystem that supports asset management:

- acquisition,
- insertion,

- maintenance,
- withdrawal.

There are different methods used at various stages and types of assets. Due to the diversity of assets, they have been divided into ICT, support, and personal assets. ICT assets constitute equipment such as computers, printers, active network elements, communications equipment, etc. as well as software and databases. The second type of assets - support assets - are buildings, rooms, offices, server rooms, zones in which computer systems are operated, and installations such as power, cooling, alarm and extinguishing systems. Personal assets are employees, contractors and other human resources who provide an organization with services.

The most extensive subsystem that supports asset management is presented in Figure 3 in a form of a SysML language diagram. Its most important part should record execution of all activities expected to be conducted within a procedure associated with a given asset and inform appropriate users about the need for a cyclical performance of these activities. Implementation of procedures in a subsystem, acting as a 'wizard' which offers a tutorial for the user, facilitates work to a great extent.

The subsystem that supports incident and threat management, whose requirement diagram is shown in Figure 4, has six functional requirements:

1. collecting information about vulnerabilities,
1. registration of incidents,
2. registration of reactions to vulnerabilities and incidents
3. collecting evidence materials,
4. concluding,
5. providing information.

Collecting information about vulnerabilities to threats for operated elements of a computer system is based on user observations, hardware and software manufacturers' publications, newsletters of security organizations dealing with information security, news, and more.

Incident reporting cannot be limited to computer system users. Even a completely independent third party can observe an event related to information security. A valuable source is technical components of the system such as antivirus software, firewalls, intrusion detection systems, etc. It is important to store and categorize information about incidents in one place.The described subsystem allows for a registration of reaction to vulnerabilities and incidents. It forces users indirectly to undertake activities related to removing effects of incidents and drawing conclusions for future. The subsystem also provides data on their effects and stores evidence materials.

The requirement marked in figure 4 - Concluding - allows for measuring and monitoring the size and cost of reactions to vulnerabilities and incidents. This function is extremely important for local self-governments bodies which plan their expenditures ahead.

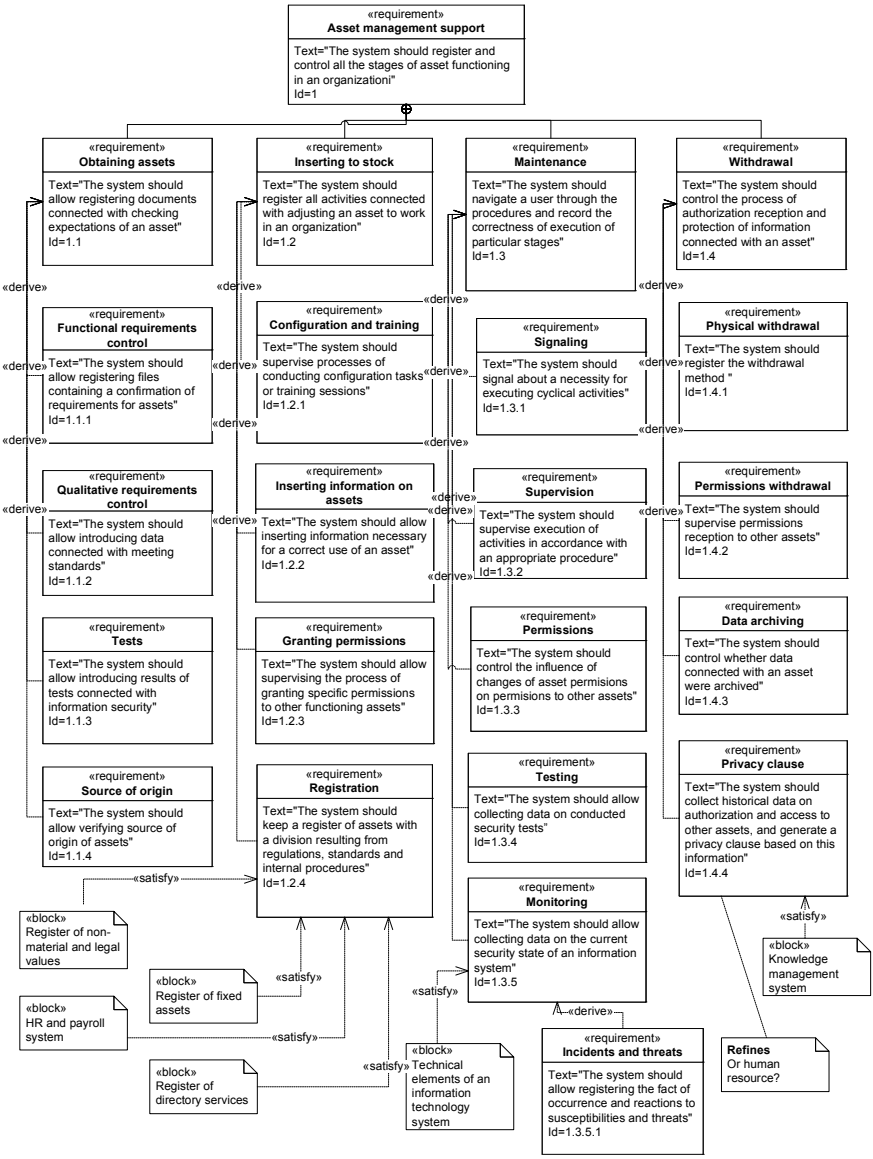

Fig. 3. Subsystem requirement diagram that supports asset management *(Source: own study)*

Another requirement (Fig. 4) - Providing information - allows for a significant increase of the security level. The system must immediately transfer information about a disclosed vulnerability to all system users who are at risk. Reception of the message must be confirmed. People responsible for the state of information security should receive a message about an incident equally fast. The message should contain accurate data that would allow for an adequate, effective reaction.

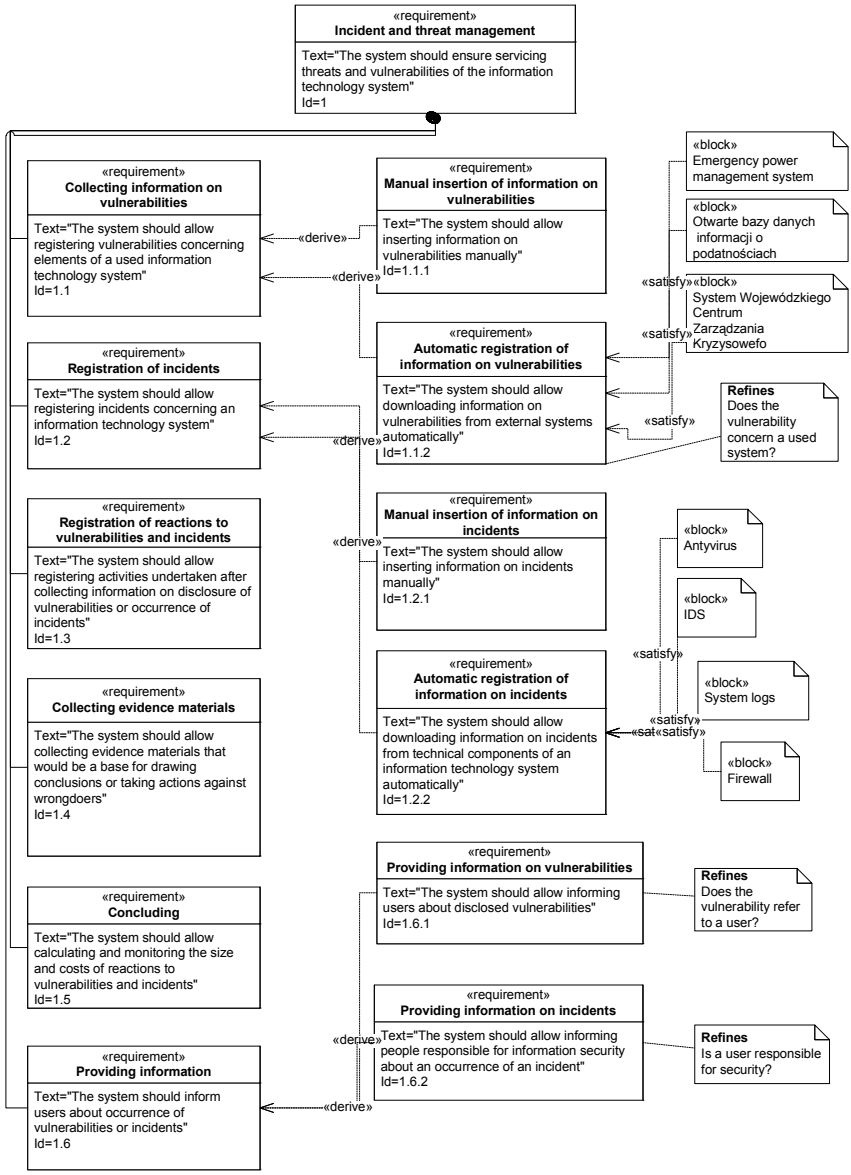

Fig. 4. Subsystem requirement diagram for vulnerability and incident management *(Source: own study)*

The communicative function of the subsystem allows for a rapid flow of information between users who are involved in various processes. This is done via email, SMS, instant messenger or pop-up message.

The developed, original model of a system that supports information security management, illustrated with numerous requirement diagrams, should become a factor

that significantly increases the security level of systems operated by local self-government bodies.

Unlike other models, it includes the whole of human and technical resources without making explicit differentiations between them. It presents a broad aspect of compliance with regulations which is usually considered to be marginal. It also presents how to deal with vulnerabilities and incidents.

References

1. De Walt, D.: Unsecured Economies: Protecting Vital Information. McAfee Inc. (2009)
2. House of Lords: Personal Internet Security. In: Science and Technology Committee 5th Report of Session (2006)
3. Frangopoulos, E.D., Eloff, M.M., Venter, L.M.: Psychosocial risks: Can their effects on the security of information systems really be ignored? Information Management & Computer Security 21(1), 53–65 (2013)
4. Yin, R. K.: Case study research design and methods. Sage Publications Inc. (2003)
5. Friedenthal, S., Moore, A., Steiner, R.: A practical guide to SysML: the systems modeling language. Morgan Kaufmann, San Francisco (2011)
6. Weilkiens, T.: Systems engineering with SysML/UML: modeling, analysis, design. Morgan Kaufmann, San Francisco (2011)
7. Jensen, J., Sherry, Y.: Opportunities and threats: A security assessment of state e-government websites. Government Information Quarterly (2010)
8. Karokola, G.R., Kowalski, S., Yngström, L.: Towards An Information Security Maturity Model for Secure e-Government Services: A Stakeholders View. In: Proceedings of the 5th International Symposium on Human Aspects of Information Security & Assurance, HAISA 2011, pp. 58–73 (2011)
9. Groves, R.M., Fowler, F.J., Couper, M.P., Lepkowski, J.M., Singer, E., Tourangeau, R.: Survey Methodology. John Wiley & Sons Inc. (2009)
10. Bau, J., Bursztein, E., Gupta, D., Mitchell, J.: State of the art: Automated black-box web application vulnerability testing. In: IEEE Symposium on Security and Privacy (SP) (2010)
11. Ofuonye, E., Beatty, P., Dick, S., Miller, J.: Prevalence and classification of web page defects. Online Information Review 34(1) (2010)
12. Posey, B.: GFI network security and PCI compliance power tools. Syngress (2011)
13. Lowis, L., Accorsi, R.: On a classification approach for SOA vulnerabilities. In: 33rd Annual IEEE International Computer Software and Applications Conference, COMPSAC 2009, vol. 2, pp. 439–444 (2009)
14. Feng, N., Wang, H.J., Li, M.: A Security Risk Analysis Model for Information Systems: Causal Relationships of Risk Factors and Vulnerability Propagation Analysis. In: Sixth International Conference on Information Assurance and Security (IAS), pp. 143–148 (2010)
15. Fienberg, S.E.: The analysis of cross-classified categorical data. Springer (2007)
16. Technology users believe human errors is the leading cause of data loss. Kroll Ontrack Releases Results of Global Data Loss Causes Survey, Revealing Gaps Between Data Loss Cause Perceptions and Realities (July 26, 2010), http://www.krollontrack.com/company/news-releases/?getpressrelease=61462
17. Calder, A.: Implementing Information Security Based on ISO 27001/ISO 17799: A Management Guide. Van Haren Publishing (2006)

The Integration of Web-Based Information and the Structured Data in Data Warehousing

Jacek Maślankowski

Department of Business Informatics, University of Gdańsk, Poland
jacek@ug.edu.pl

Abstract. The article presents the concept of the solution for feeding the data warehouse from website forums including opinions about selected products. The key of the solution is to add a new data warehouse dimension called Variable that allows identifying both structured and unstructured data. In suggested solution the results of websites analysis will be stored in the same repository as the data from traditional corporate systems, such as CRM or ERP. The concept was presented regarding Internet shops that offered a selected kind of products.

Keywords: data warehouse, big data, unstructured data, data analysis.

1 Introduction

In recent years it has been observed that the value of unstructured data is increasing. This is the result of the fact that a large amount of valuable information is available on the Internet. For instance consumer preferences can be checked by looking deeper into Internet forums. To search through the websites to seek for consumers' behaviours and habits, Big Data technology can be used. It involves semantic algorithms to search through websites trying to find matching text. Those websites may include Twitter and Facebook as valuable sources for retrieving users' opinions on various topics [Bartram, 2013, pp. 28]. But the data just from the website can not be enough for the company to make the right decision.

The goal is to integrate Big Data approach with the existing databases in the company. Nowadays the typical analytical database used in companies is the data warehouse [Reddy et al., 2010, pp. 1186]. The data warehouse with OLAP tools has been treated as the most significant technology in business data processing since it was developed [Thomas & Datta, 2001, pp. 83].

The thesis of the article is to show that the data can be more valuable when it is loaded from structured and unstructured sources. The aim is to propose a concept of the framework to feed data warehouse directly by using Big Data tools. The way the goal has been formulated indicates that the data warehouse should be a repository to store the analytical data from several sources, including structured and unstructured information. The suggested data warehouse environment should be used in businesses selling different kind of products, such as Internet shops.

S. Wrycza (Ed.): EuroSymposium 2013, LNBIP 161, pp. 66–75, 2013.

The article was divided into six parts. The first part is introduction. In the second part there is a theoretical background of the possibilities of unstructured data analysis. The third part shows proposal for the alternative method to gather the result of analysis based on unstructured data sources in the data warehouse. In the fourth part there is a short comparison between proposed solution and acknowledged methods of unstructured data integration. The last part shows conclusions.

2 Possibility of Unstructured Information Analysis

2.1 Using Big Data to Filter Websites

Decisions made in organizations are usually based on the Data-Driven Decision-Making Process in which the decision process is based on information systems as well as internal and external factors. The result is the course of action [Picciano, 2012, pp. 12]. Those external factors are not only derived as structured data, but also as unstructured information.

Although there is no consensus on the Big Data term, there are several definitions that show the idea of the Big Data technology. Big Data is a generic term that assumes that the information or database system used as the main storage facility is capable of storing large quantities of data longitudinally and down to very specific transactions [Picciano, 2012, pp. 12]. Big Data technology is necessary when the data are too big for traditional systems to handle it [Gobble, 2013, pp. 64]. From the business point of view, Big Data is usually defined as three V's: volume, velocity and variety [Chen et al., 2012, pp. 1182], but sometimes it is defined together with the four V – the fourth one is cited as veracity [Harris, 2013, pp. 29].

But one of the goals of the Big Data is to support an analysis of large amount of unstructured data. More than 80% of all potentially useful business information is unstructured data, including e-mails, social media, videos, images, sensor readings, console logs and others [Das & Kumar, 2013, pp. 153].

Useful information on consumer habits and preferences can be obtained by collecting and analysing the information about product returns, warranties and customer complaints. But this information should be used to potential [Bughin et al., 2011, pp. 104-105]. In literature we can find several concepts of frameworks to analyse social networks, such as blogs. In early stage of analysing the blogosphere, several techniques like data and text mining were used to extract information from blogs by analysing its content [Chau & Xu, 2012, pp. 1190]. Data and text mining used together are well known as the duo mining technology [Maślankowski, 2006, pp. 973].

On the other hand it is important to study the media effects in the electronic media environment, such as Internet sphere with several techniques, using for example social networks analysis [Guo, 2012, pp. 617]. Social network analysis and cluster analysis are wide used to analyse keywords in text documents [Khansa et al., 2012, pp. 20].

The most referred way to make an analysis using Big Data technology is developing the solution on the Apache Hadoop platform [Teplow, 2013, pp. 38], [Harris, 2013, pp. 29]. One of the results of the importance of the Big Data technology is

increasing number of Internet users. Based on the Eurostat data, the number of household having access to the Internet in EU27 countries has increased from 41% in 2005 to 76% in 2012 [Eurostat, 2013]. It can be presumed that this increase will correspond to the number of users using Web 2.0 tools.

2.2 Unstructured Data and Data Warehousing

Integration of unstructured and structured data in data warehousing was one of the issues taken into consideration during the data warehouse evolution. Well known proposal of the unstructured data inclusion in the data warehouse was suggested by W.H. Inmon in the fourth edition of the book Building the Data Warehouse. His approach was to match all the information from unstructured documents, such as e-mails, text files, spreadsheets and similar. The matching was mostly based on probabilistic match as well as themed match, to find a relationship between structured and unstructured documents. Two basic approaches were presented – first was to pull data over into the structured environment. The second was to create the two-tiered data warehouse – one tier for unstructured and another for structured data. In the same elaboration he also wrote about Granularity Manager that is used to collect the data from the web logs, mostly clickstream data. In this approach the extraneous data must be created in the single record, then incorrect data edited, after that the data must be converted, summarized and aggregated [Inmon, 2005, pp. 290-291, 305, 311-313, 320]. Therefore the main goal is to provide the mechanism to create a single record that will be processed by Granularity Manager, which is also responsible for passing the refined data into the data warehouse. Unstructured textual data were placed in the data warehouse repository in the approach called Data Warehouse 2.0 [Inmon et al., 2008, pp. 34-35].

However, the solution for filtering web data to feed the data warehouse had been published before the DW 2.0 concept was developed. The first important step into feeding the warehouse from web data was made by defining the web farming Data Warehouse. In this type of the warehouse business-relevant Web content is the input to the data warehouse [Hackathorn, 1997, pp. 43]. This approach was used mostly to collect clickstream logs in the whole user interaction process [Hu & Zhong, 2008, pp. 296-297]. Another approach was named “enhanced Data Warehouse” and the goal of this type of the warehouse was to use the Reader to filter business information from the Web to the data warehouse [Abramowicz et al., 2000, pp. 5, 105, 121-122]. Nowadays we can find the term Hadoop Data Warehouse that is referred to the data warehousing related to Big Data [McKenna, 2013, pp. 9-10].

3 The Concept of the Warehousing Environment

3.1 The Architecture of Data Warehousing

This chapter shows proposal of the schema of the infrastructure used to search through websites and put it into the data warehouse.

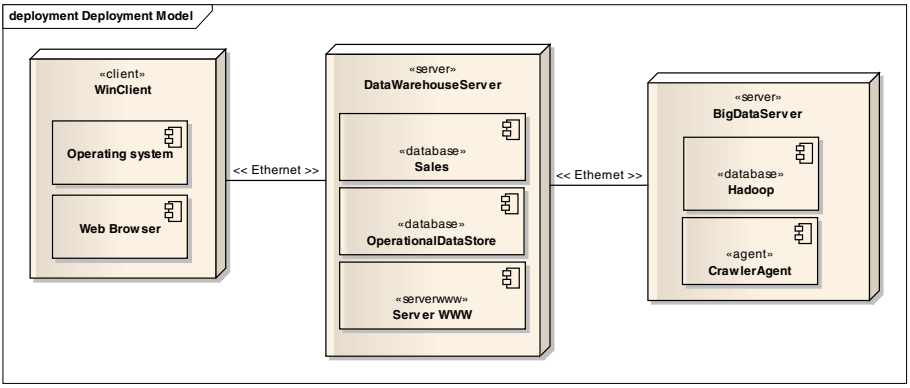

Fig. 1. The data warehouse environment with components to filter unstructured data (Source: own elaboration)

The proposed data warehouse environment is shown in figure 1. The concept of using component and deployment diagrams of UML framework to present the architecture of data warehousing is based on [Lujan-Mora & Trujillo, 2004, pp. 51-53].

The suggested data warehouse environment consists of the following subsystems that were presented as components in figure 1:

- Sales – a data warehouse which stores a data in a star schema as shown in figure 2.
- OperationalDataStore – a staging area in which the unstructured and structured data are gathered.
- Hadoop – tools used to support unstructured data analysis, by parallel processing large amount of data.
- CrawlerAgent – the agent used to gather information from websites.
- Server WWW – server used to host the result of analysis to clients.
- Operating system and Web Browser are typical components used to access the result of analysis hosted by Server WWW.

3.2 The Structure of the Data Warehouse

To store the results of unstructured data processing in the data warehouse, following rules concerning the data warehouse schema must be applied:

- there should be a dimension called Variable that will be able to indicate and describe the keywords that were analysed,
- other dimensions must be the same as in typical data warehouse.

The way the UML class diagram was used to present the star schema in the figure 2 is based on the Data Warehouse Conceptual Schema (DWCS) presented in [Lujan-Mora et al., 2004, pp. 193-194].

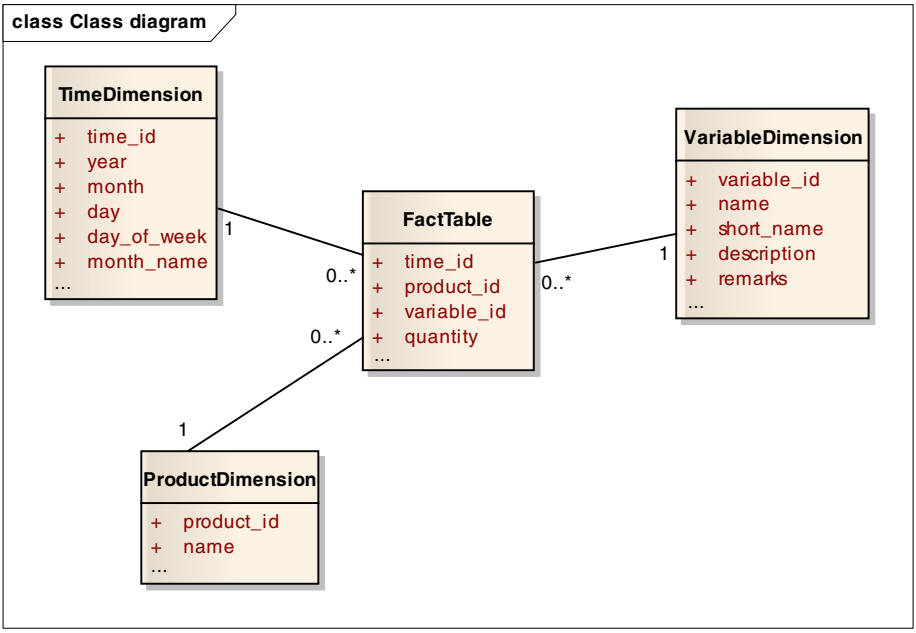

Fig. 2. The data warehouse structure (Source: own elaboration)

As mentioned above, the key of the data warehouse structure is the dimension called "Variable". This dimension is used to describe both structured and unstructured attributes that are used to define the data stored in the warehouse. Using the warehouse terminology, the "Variable" dimension can be treated as business metadata.

The dimension has the following structure, as shown in table 1.

Examples of the data stored in the Variable dimension are presented in table 2.

Table 1. The structure of the "Variable" dimension

Name of the attribute	Data type	Remarks
Variable_ID	Number(5)	ID number – created by using sequences
Name	Varchar2(100)	Used to name the variables
Short_Name	Varchar2(25)	Short names are used in BI tools to be included in header rows or columns
Description	Varchar2(500)	More detailed information about the variable
Remarks	Varchar2(500)	Used to explain a methodology of data gathering

Table 2. The data stored in the "Variable" dimension

Variable_ID	Name	Short_Name	Description	Remarks
1	Number of product sold	Sales	The total number of the particular product sold in the selected period of time.	Structured data source
2	Positive opinion on the product	Positive opinion	The variable is based on the opinion from selected websites forums.	The opinion is based on keywords matching.
3	Negative opinion on the product	Negative opinion	The variable is based on the opinion from selected websites forums.	The opinion is based on keywords matching.
…				

As shown in the table 2, the Variable dimension is used as the metadata to describe the data stored in the fact table of the data warehouse. This concerns both structured and unstructured data.

3.3 Selecting Sources of Unstructured Data

Implementing this solution is just the half of a success, as it is necessary to perform several steps that will ensure company that unstructured data retrieved from selected websites is reliable and can be used as the source system.

The key issue that must be regarded is a proper selection of unstructured data sources. In this article the focus is on assessing the products offered by a particular Internet shops. In this example there are several potential sources to feed the warehouse, which can be mostly adapted to almost every type of products offered by Internet shops:

- Internet shops forums that sell the same products as offered by the shop for which the data warehouse is implemented,
- subject-oriented Internet forums concerning the group of products offered,
- forums on thematic websites, such as eopinions.com,
- website of the shop selling the products,
- e-mails sent to the shop.

It is obvious that some Internet forums on shops websites are filtered and sometimes the opinion published on the website is limited only to the good one. Therefore the suggestion is to create a rank of the websites that can be included in implemented system. The list was prepared to select the right source, as shown in the table 3.

Table 3. The list of potential data sources

No.	Name of the source	Remarks	Reliability
1	eopinions.com	A lot of opinions about miscellaneous products.	high
2	amazon.com	Mostly about products offered by Amazon.	high
3	consumersearch.com	Lots of professional reviews.	high
…			

Based on the list of potential data sources, the proper as well as the most reliable data sources must be selected to be included in the data warehouse.

3.4 Matching the Keywords

The matching keyword algorithm used to filter Internet forums to find useful opinions about offered products is usually limited to one language. The effect of the globalization is that we can search through the Internet and find some useful opinions about products in several thematic websites from different countries. It means that the algorithms of matching keywords should be extended to be able to compare the keywords from different languages in one variable, such as pattern:

is good – ist gut – es bueno – est bon – é bom – är bra – on hyvä

must be equivalent to each other and should increase the value of the Variable "positive opinion". On the other hand the list of the following keywords means the same:

not so expensive – not expensive – no expensive – cheap – rather cheap

But in some cases there are some patterns that use the same words but are different:

so expensive – not so expensive

The last key issue that should be regarded before making algorithms to filter the websites is that the sentences published on the Internet forums include lots of mistakes and the following patterns should be regarded as the same:

expensive – expensve – expensiev (intentionally written with mistakes)

On the other hand there are some mistakes such as misspellings that can be ambiguously interpreted:

on expensive – no expensive

Therefore the solution proposed in this article will never be fully reliable. But looking deeper into psychological issues, Internet forums are full of subjective opinions from users, sometimes just because they are not satisfied from their shopping, not the product they bought.

In this article there is no recommendation on semantic algorithm used to matching the keywords included in different opinions, as it is not the goal of this paper.

4 The Comparison between Proposed Solution and Existing Ones

The goal of the article was to show a concept solution to include information from Internet forums in the data warehouse. The biggest advantage of the suggested

solution is that it doesn't affect the warehouse environment that is used in most of the companies.

In suggested solution, a traditional data warehouse used in a company can be enhanced by adding new dimension called Variable. This dimension will make the data warehouse more flexible to apply to changes in the environment, such as describing new phenomenon that will occur in the future. This phenomenon could be a new voting mechanism for products or any other social forum about products.

The proposed solution uses one repository for structured and unstructured data, and unstructured data are only available as a result of analysis included in the structured star schema. Looking into other proposals, such as DW 2.0 proposed by W.H. Inmon, there are two different repositories, separately for structured and unstructured data (compare the figure 2.2 in [Inmon et al., 2008, pp. 28]). However the DW 2.0 is much more sophisticated than proposed architecture in this article because it assumes including almost any type of unstructured document, while the solution described in this article focused on websites only.

Other proposals of integrating web feed opinions with the data warehouse assumes to use several fact tables and dimensions as well [Moya et al, 2011, pp. 22-23], which differentiate from the solution proposed in this paper that includes only one fact table and several dimension tables, depending on the need of including them in the warehouse schema.

The suggestion of implementing the solution is to develop it with Oracle database, as it offers Big Data connectors to Apache Hadoop, which can be used for parallel processing the websites. The data warehouse implemented with guidelines suggested in this article leads to the conclusion that the following information, presenting in table 4, can be received from the warehouse.

Table 4. Example results of the analysis

Product_name	Month	Year	Positive opinion	Negative opinion	Sales
A	January	2013	23	2	110
A	February	2013	42	3	159
A	March	2013	73	4	169
B	January	2013	34	5	210
B	February	2013	55	6	204
B	March	2013	62	4	190

Based on the table above which presents a snapshot of the data report from the suggested warehouse repository, we can easily compare the sale amount and the number of positive and negative opinions written on Internet forums. This is the value added to the data warehouse. Efficient use of the Big Data technology and unstructured data filtering will allow company to build their market strategy supporting by the list presented above.

5 Conclusions

The goal of the article was to show an alternative way to design and implement a solution to integrate both structured and unstructured data. The main advantage of suggested solution to integrate structured and unstructured data together is that it is rather simple and easy to implement.

As it was written in this article, unstructured data are one of the most relevant sources of information in business today. Rapid increase in the number of users on the Internet has a big impact on the way how company is perceived by society. In that sense, Internet is one of the sources for lots of people trying to get an opinion about particular products. This lead to the conclusion that companies today cannot lose their chances to rapidly react to changes in the market, based on opinions of different Internet forums.

To have a strong and fast reaction to changes on the market, it is necessary to implement a solution that will be able to gather and process users' opinions about different products. These opinions are freely available on various websites. The goal is to ensure that the data warehouse will provide rapidly and on time information to make changes in the rules of products offered. It means that the information must be processed continuously and automatically, what is one of the aims of the Big Data technology.

Although there is variety of information systems that allows filtering data from websites, the availability of Big Data technology, such as Apache Hadoop, will make it easier to implement and use real-time web filtering system.

Abbreviations

CRM – Customer Relationship Management
DW – Data Warehouse
EU – European Union
ERP – Enterprise Resource Planning
UML – Unified Modelling Language

References

1. Abramowicz, W., Kalczyński, P., Węcel, K.: Filtering the Web to Feed Data Warehouses, pp. 5, 105, 121–122. Springer (2002)
2. Bartram, P.: The value of data. In: Financial Management, p. 28 (March 2013)
3. Bughin, J., Livingston, J., Marwaha, S.: Seizing the potential of 'big data'. McKinsey Quarterly (4), 104–105 (2011)
4. Chau, M., Xu, J.: Business Intelligence in Blogs: Understanding Consumer Interactions and Communities. MIS Quarterly 36(4), 1190 (2012)
5. Chen, H., Chaing, R.H.L., Storey, V.C.: Business Intelligence and Analytics: From Big Data to Big Impact. MIS Quarterly 36(4), 1182 (2012)
6. Das, T., Kumar, P.M.: BIG Data Analytics: A Framework for Unstructured Data Analysis. International Journal of Engineering Science & Technology 5(1), 153 (2013)
7. Eurostat database, http://epp.eurostat.ec.europa.eu/

8. Gobble, M.M.: Big Data: The Next Big Thing in Innovation. Research Technology Management 56(1), 64 (2013)
9. Guo, L.: The Application of Social Network Analysis in Agenda Setting Research: A Methodological Exploration. Journal of Broadcasting & Electronic Media 56(4), 617 (2012)
10. Harris, C.: Dividing into Big Data. Canadian Underwriter 80(2), 29 (2013)
11. Hackathorn, R.: Farming the Web. Byte.com 22(10), 43 (1997)
12. Hu, J., Zhong, N.: Web farming with clickstream. International Journal of Information Technology & Decision Making 7(2), 296–297 (2008)
13. Inmon, W.H., Strauss, D., Neushloss, G.: DW 2.0. The Architecture for the Next Generation of Data Warehousing, pp. 28, 34–35. Elsevier Inc. (2008)
14. Inmon, W.H.: Building the Data Warehouse, 4th edn., pp. 290–291, 305, 311–313, 320. Wiley Publishing, Inc. (2005)
15. Khansa, L., Zobel, C., Goicochea, G.: Creating a Taxonomy for Mobile Commerce Innovations Using Social Network and Cluster Analyses. International Journal of Electronic Commerce 16(4), 20 (Summer 2012)
16. McKenna, B.: King.com gaming site unlocks big data to switch to Hadoop database. Computer Weekly, 9–10 (May 3, 2013)
17. Maślankowski, J.: Integration of Text- and Data-Mining Technologies for Use in Banking Applications. In: Nillson, et al. (eds.) Advances in Information Systems Development, p. 973. Springer Science, New York (2006)
18. Moya, L.G., Kudama, S., Aramburu Cabo, M.J., Berlanga Llavori, R.: Integrating web feed opinions into a corporate data warehouse. In: Proceedings of the 2nd International Workshop on Business Intelligence and the WEB. ACM, New York (2011)
19. Lujan-Mora, S., Trujillo, J.: Physical Modeling of Data Warehouses using UML. In: Proceedings of the 7th ACM International Workshop on Data Warehousing and OLAP, pp. 51–53 (2004)
20. Luján-Mora, S., Vassiliadis, P., Trujillo, J.: Data Mapping Diagrams for Data Warehouse Design with UML. In: Atzeni, P., Chu, W., Lu, H., Zhou, S., Ling, T.-W. (eds.) ER 2004. LNCS, vol. 3288, pp. 191–204. Springer, Heidelberg (2004)
21. Picciano, A.G.: The Evolution of Big Data and Learning Analytics in American Higher Education. Journal of Asynchronous Learning Networks 16(3), 12 (2012)
22. Reddy, G., Srinivasu, R., Rao, M., Rikkula, S.: Data Warehousing, Data Mining, OLAP and OLTP Technologies are Essential Elements to Support Decision-Making Process in Industries. International Journal on Computer Science & Engineering 2(9), 2866 (2010)
23. Teplow, D.: The emperor has no clothes. Business Intelligence Journal 18(1), 38 (2013)
24. Thomas, H., Datta, A.: A Conceptual Model and Algebra for On-Line Analytical Processing in Decision Support Databases. Information Systems Research 12(1), 83 (2001)

Matching Process Requirements to Learning Objectives

Christoph Becker and Thomas Schuster

FZI Forschungszentrum Informatik am Karlsruher Institut für Technologie
Haid-und-Neu-Str. 10-14, D-76131 Karlsruhe, Germany
{christoph.becker,thomas.schuster}@fzi.de

Abstract. Rapidly altering market requirements foster shortened evolution cycles of business processes. Regarding human resources these requirements can be integrated in business process models by usage of capability descriptions. Matching these descriptions with pre-existing profiles of human resources is reasonable. In this article an innovative approach for efficient management of further education in alignment to business process requirements will be introduced. Areas for improvement can be revealed by using enhanced modelling methods that allow the calculation of requirements for the execution of business process tasks. In order to enable the allocation of learning modules according to capability gaps this article presents a formalized representation for learning objectives of further education modules. Furthermore an algorithmic technique that matches learning objectives to capability descriptions to improve cost efficient selection of further education modules is proposed. In consequence this method enables an improved allocation of appropriate further education modules and forms the basis for automated decision support.

Keywords: Business Process Management, Resource Management, Further Education, Decision Support.

1 Introduction

Since companies have to fulfill changing market demands more quickly, business processes are subject to shorter evolution cycles. On an enterprise-level perspective this results in altered claims to business partners, on a more detailed level requirements to resources will change as well. As far as human resources are affected, these requirements can be integrated in business process models by competence and qualification descriptions. Matching these descriptions with those already existing in human resource departments is reasonable and can lead to improved resource management. In business process management nomenclature resources are entities (such as materials, machines or employees) that can be assigned to a task. During runtime these resources have to be coordinated in order to enable execution of tasks and hence process execution. Resource management must deal with resource assignment, problems, for instance, may arise if resources are temporally unavailable or unqualified. Thus reasoning about business processes and resources requires an understanding of requirements (given by process models) and properties of resources. Especially knowledge of similarities and differences among resources influences

S. Wrycza (Ed.): EuroSymposium 2013, LNBIP 161, pp. 76–84, 2013.

process execution and fosters further insights (such as identification of critical resources or bottlenecks).

This article will introduce an innovative approach for efficient human resource management. Areas for improvement will be revealed by modeling methods that allow the calculation of capability gaps. The suggested capability gap calculation will be extended by an algorithmic approach to allow decision support for choice of further education.

2 Resource Management

Resource management has to prevent or handle concurrent access to resources as well as access to (temporally) unavailable resources appropriately. Resources are entities (such as employees, data objects or materials) that can be assigned to carry out tasks [2]. In resource management, resources with similar properties are grouped to resource classes. This idea has evolved within last decades, while new class types (such as information) have been added according to industrial development.

In the context of resource management relevant resources have to be comprised in resource models, thereby facilitating their coordination by workflow management systems. Along with resources a lot of additional information (e.g. temporal availability, capabilities, and spatial information) has to be monitored and managed to allow their coordination alongside intercompany processes [3,10]. The next subsection will summarize approaches of resource management and modeling, while subsection 2.2 will introduce a distance measure to enable capability gap identification.

2.1 Resource Modelling

Resource models depict profiles of existing resources. In order to model resources several approaches (such as [12,16]) have been proposed so far. Most of these approaches are either domain specific or general description approaches. Allocation of resources can be enacted by models which comprise two types of information: a) runtime aspects b) resource properties that can be matched to requirements (such as [2,7,11]). Since many approaches neglect to model capability information [3,15] explicitly, they are not adequate to enact capability based allocation, neither are they usable to calculate capability gaps nor necessary further education of resources. In order to support this allocation principal and, if necessary suggest further education the Resource Modeling Language (RML, [7]) will be employed to model resources. RML is based on a MOF-compliant meta-model which allows precise definitions of concepts and relationships.

This article does not intend to introduce RML in detail. However, it is important to mention the most important aspects about RML. RML does allow typical resource modeling that distinguishes resource categories, regarding companies (organizations) and human resources. RML does allow modeling of organizational hierarchies alongside with a lot of resource details. Most important for this approach is, that capabilities (competences, skills and knowledge) can be associated to human

resources as well as to roles they carry out. The RML meta-model can be formalized by the definition of a relation HR that does express properties of given human resources. All properties of resources can be expressed as sets over which HR is defined – for this article we will only focus on human resources and capabilities. Let R be a set of resources, $K = 9^n$ be a set of knowledge instances including level description, $S = 9^n$ be a set of skills including level description and $C = 9^n$ be a set of competences including level description, then $HR \subseteq R \times K \times S \times C$. Properties of a single resource may be expressed as vectors displayed in **Table 1**. In short:

$$(Employee\ A, (K_{13}: 6), (S_2: 3), (S_{11}: 5), (S_{21}: 4), (C_0: 1), (C_1: 2), (C_3: 5), (C_5: 6), (C_7: 3))$$

Table 1. Properties of Member A

Domain	Profile of Employee A
K	(0,0,0,0,0,0,0,0,0,0,0,6,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0)
S	(0,0,3,0,0,0,0,0,0,0,0,5,0,0,0,0,0,0,0,0,0,4,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0)
C	(1,2,0,5,0,6,0,3)

2.2 Calculation of Requirements Fulfilment

In order to calculate adequacy of resource to execute tasks the distance vector $d_{ij} = \alpha_i \cdot v_{ij}$ that maps a distance of task i to resource j is calculated. α_i specifies task importance. In the remainder of this article α_i was set to one (all tasks are considered equally important). Thus v_{ij} is given by requirements and resource profiles as:

$$v_{ijk} = \begin{cases} 0 & K_{jk} > A_{ik} \\ A_{ik} - K_{jk} & K_{jk} < A_{ik} \\ 8 & K_{jk} = 0 \end{cases}$$

v_{ijk} is entry k of vector v_{ij}. A_{ik} does express the level of capability k required to execute task i. K_{jk} does reflect the actual level of capability k possessed by resource j. In case that a resource does not possess a capability at all ($K_{jk} = 0$), the distance entry is maximized. In this case 8 is the maximum value, because according to European Qualifications Framework for Lifelong Learning (EQF) eight capability levels are defined in RML. In a following step each distance vector is transformed to a length measurement $\Phi_{ij} = \|d_{ij}\|_2$ based on the Euclidean norm. Hence Φ_{ij} expresses the distance of resource capabilities to requirements. After calculating the distance Φ_{ij}, a threshold parameter δ can be used to derive recommendations for action. A resource with minor deviation to the capability requirements still qualifies for task execution whereas a resource with a deviation beyond the threshold requires further education or recruitment of new staff.

3 Learning Objects

Learning objects (LO [4,13]) represent one instrument that can be used in further education. As stated above adequate resource management enables the identification

of gaps between task requirements and capabilities of personal. In order to assure the execution of business processes in collaborative scenarios these gaps need to be eliminated by either recruitment of qualified staff or further education of existing staff. Increasingly available internet technologies foster the use of multimedia learning content for further education [13]. After introducing the concepts of learning objects, the extension of LOs with learning objectives is suggested in this section.

3.1 Learning Object Description

Currently several definitions of learning objects are existent. Even though there are slight differences amongst existing definitions, they are generally based on the concept of reusability of learning objects. Within this paper LOs are considered as digital or non-digital, self-contained entities of learning, training or assessment content that can be reused in multiple contexts [4,9,14]. LOs are modular elements that can be composed to larger more complex units, so called learning modules [1]. Depending on size and complexity of a LO it is considered to be an atomic entity (e.g. a picture or text), a reusable learning object (RLO), a course or a complete training [1]. However, there is no agreement on advisable size and complexity of LOs. Size and complexity of LOs strongly depend on the author who created them [5]. Digital learning objects can be executed in learning content management systems (LCMS). In order to reuse LOs in different LCMS several metadata standards such as IEEE Learning Object Metadata (LOM) have been developed. LOM is a widespread metadata standard used to describe LO characteristics and facilitate the automated identification, selection and application of LOs [4]. Due to a missing formal model to describe learning objectives, LOM does not appropriately support an automated allocation of Learning Objects based on learning objectives and capability models.

3.2 Modelling Learning Objectives

We propose to model learning objectives for LOs. This approach will reuse the capability model introduced in section 2 to specify learning objectives. The learning objectives of each LO will be defined by sets of skill, knowledge and competence instances. After assessing a LO the user will acquire the knowledge, competence and skills defined within the learning objectives. The allocation of learning objectives to LOs can be formalized by a relation LOR that will be defined in the following. Let LO be a set of learning objects, $K = 9^n$ be a set knowledge instances with level description, $S = 9^n$ be a set skills with level description and $C = 9^n$ be a set competences including level description, then $LOR \subseteq LO \times K \times S \times C$. The function $\varsigma : LO \rightarrow \mathbb{R}$ yields the costs of each LO. To facilitate further processing available LORs are stored in a repository according to competence, skill and knowledge instances (see **Table 2**). Each LO listed in the right column instructs and trains the competence, skills or knowledge instance listed in the corresponding row. Applying the abbreviated notation proposed in section 2 the learning objectives of a learning object LO_1 are expressed as $\big(LO_1, (C_2\colon 3), (S_3\colon 6), (K_3\colon 5)\big)$.

Table 2. - Sample Learning Object Repository

Learning Objective	Learning Objects
$(C_2{:}3)$	LOR_1, LOR_3
$(C_4{:}3)$	LOR_2, LOR_3
$(K_3{:}5)$	LOR_1, LOR_2
$(K_3{:}6)$	LOR_2
$(S_2{:}7)$	LOR_3, LOR_4
$(S_3{:}6)$	LOR_1, LOR_4

4 Capability/Objective-Based Mapping of Learning Objects

After performing the requirements fulfilment calculation (see section 2) personal that qualifies for further education can be identified. Before relevant LOs can be selected the actual capability gap of each employee has to be determined. Due to the fact that the capability profile K and the task's capability requirement A use the same model, capabilities to be trained can be identified by a comparison function (similar to section 2.2). Thus employee K_1, for instance, has a set of missing capabilities $\big((C_2{:}3), (S_2{:}7), (K_3{:}6)\big)$ in order to execute task A_1. Each objective to be trained can be mapped to LOR elements (as introduced in section 3). The subset displayed in Table 3 lists all LORs that train capabilities that have been identified before. Each LO can train one or more capabilities and each capability can be trained by one or more LOs.

Table 3. - Subset of Learning Repository

Learning Objective	Learning Objects
$(\boldsymbol{C_2{:}3})$	LOR_1, LOR_3
$(\boldsymbol{S_2{:}7})$	LOR_3, LOR_4
$(\boldsymbol{K_3{:}6})$	LOR_1, LOR_2

In order to identify an efficient solution to further educate existing staff, we are interested to find a combination of LOR at minimum costs. To calculate the solution this paper proposes a linear integer program.

$$\min c^{\intercal}x$$
$$s.t.\ Vx \geq b$$

The objective function $c^{\intercal}x$ is minimized for the inner product of vertices c and x. Vector $c = (c_1, \dots, c_n)$ contains costs of relevant LOs and vector $x = (x_1, \dots, x_n)^{\intercal}$ with x_n being a binary variables $x_n \in \{0,1\}$ meaning

$$x_n \coloneqq \begin{cases} 1\ if\ LOR_n\ is\ selected\ for\ the\ solution \\ 0\ if\ else \end{cases}$$

The inequation $Vx \geq b$ consisting of matrix V and vertices x and b defines constraints of the objective function, so that each row of V represents a constraint stating a capability that has to be covered by a LO. If LO_n trains capability m (LO_n has a learning objective m) there is a binary coefficient $v_{mn} \in \{0,1\}$ with

$$v_{mn} \coloneqq \begin{cases} 1 \; if \; LOR \; n \; trains \; capability \; m \\ 0 \; else \end{cases}$$

Vector $b = (1, \ldots, 1)^{\mathrm{T}}$ states that each constraint has to be fulfilled so that each capability will be at least trained by one LO. Matrix V presents constraints of the sample given in Table 3.

$$V = \begin{pmatrix} 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 \end{pmatrix}, \qquad x = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix}, \qquad b = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$$

In order to calculate a selection of LOs with minimum costs the following simple algorithm is proposed. Let $n \in \mathbb{N}$ be the number of relevant LOs, $S \coloneqq \{0,1\}^n - \emptyset$ be a set of possible LO choices and $V_{set} \coloneqq \{v | v \; is \; row \; vector \; of \; V\}$ be a set of capability constraints then S, V_{set} and c are the input for the following algorithm.

Input: S, V_{set}, c
Output: $minCosts, solutionIndex$

```
minCost ← ∅
solutionIndex ← ∅
constraints ← ∅
for i = 1 to |S| do
    for j = 1 to |V_set| do
        if (v_j · s_i ≥ 1) then
            constraint ← true
        else constraint ← false
            break
    if (constraints = true and c · s_i < minCost or minCost = ∅) then
        minCosts ← c · s_i
        solutionIndex ← i
```

The algorithm essentially consists of two loops. Whereas the first loop iterates all possible choices of selecting LOs the second loop iterates all constraints that are subject to the objective function. If no constraint is violated, minimum cost are calculated and updated. In addition the algorithm stores and returns index of the solution. For n LOs the algorithm works with an upper bound of $O(2^n)$.

5 Case Study

For demonstration the algorithm presented before will now be employed in a business process scenario. Let's assume the business process given in Fig. 1 has to be executed eight times in parallel since three projects with eight pending bug requests have to be handled. Furthermore the company which is in charge of these tasks has five resources (A to E) and five job applicants available (K to O).

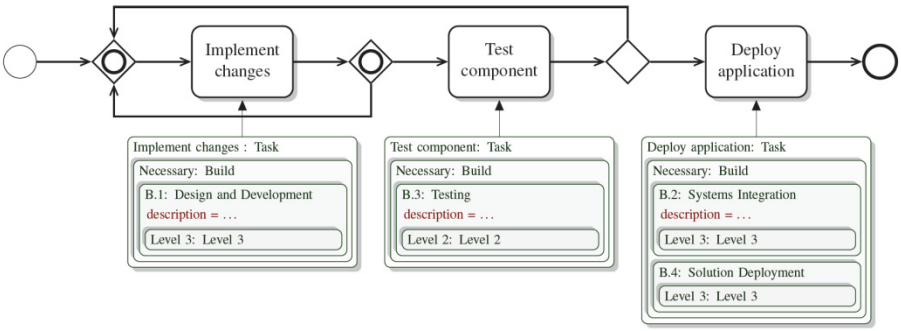

Fig. 1. Process tasks tagged with requirements profiles

Table 4 does list capability profiles of internal resources (A to E) and potential recruits (K to O). Each resource is capable of handling two tasks at one point of time; furthermore the execute task threshold δ_{task} is set to 2 (the training threshold δ_{train} is set to 4). With this setup we will demonstrate the choice of resources and further education as suggested by algorithms given above.

Table 4. - Available Resources

Resource	Type	Profile	Φ_{ij}
A	Internal	$(C_2:3), (C_4:3)$	0,00
B	Internal	$(C_2:0), (C_4:1)$	3,61
C	Internal	$(C_2:3), (C_4:3)$	0,00
D	Internal	$(C_2:4), (C_4:3)$	0,00
E	Internal	$(C_2:0), (C_4:0)$	11,31
K	External	$(C_2:0), (C_4:0)$	11,31
L	External	$(C_2:1), (C_4:2)$	2,24
M	External	$(C_2:3), (C_4:2)$	1,00
N	External	$(C_2:3), (C_4:3)$	0,00
O	External	$(C_2:2), (C_4:2)$	1,41

According to threshold δ_{task} =2 only the resources A, C and D as well as potential recruits M, N and O are capable to execute task *deploy application*. Due to $\delta_{train} = 4$ resource B will be trained to handle the tasks. Resource B lacks the competence *System Integration* $(C_2:3)$ and *Solution Deployment* $(C_4:3)$ (see Fig. 1 and Table 4). Based on learning objectives in Table 2 the following set of learning objects LO_1, LO_2 and LO_3, that train missing capabilities, is identified. Assuming the following learning objects' costs $LO_1 = 2$, $LO_2 = 1$ and $LO_3 = 4$ the algorithm presented above calculates a solution consisting of learning objects LO_1 and LO_2 with minimum costs of 3. If the training threshold δ_{train} would be set to 3 none of the internal resources would be trained. Instead, resource M with $\Phi = 0$ would be recruited to handle the tasks.

6 Conclusion and Outlook

The challenge to enrich, update and match available information of resources (profiles) and task requirements has to be addressed in order to support decisions regarding further education and recruiting. On basis of RML resource profiles can be described and analysed. In this article we demonstrated how capability descriptions, as given in RML, can be reused for the modelling of business process relevant requirements and formal description of learning objectives. This demonstrates the potential of the combination and extension of known concepts in resource modelling and business process management. The formal description of capabilities and learning objectives enables capability driven execution of business processes as well as dynamic acquisition of learning objects that train required capabilities. Especially the domain of e-learning does benefit from dynamic allocation of learning objects and thereby simplifies the process of further education.

In order to facilitate dynamic allocation of learning objects dependent on learning objectives, task requirements and resource profiles we introduced an algorithm to select relevant learning objects at minimum cost in a given situation. However, at this stage the algorithm does only optimize the costs of further education, thus additional criteria like the time to conduct learning objects, a minimum set of learning objects to cover a set of learning objectives or a minimum set of learning objects for more than one person are not taken into account yet. As next steps we plan to simulate larger scenarios to evaluate our algorithm. Furthermore we intend to cover additional criteria while developing a more efficient heuristic algorithm as well as stop criteria to improve the algorithm's upper bound of $O(2^n)$. Future findings will continuously be evaluated in simulations and benchmarked against optimum results.

With the intention of practical enactment of our approach, we furthermore intend to introduce a description format for learning objectives that can be used alongside e-learning environments. This is essential to our goal of an integrated decision support system that enables forecasting lacks of capabilities and provides recommendations for action. Additionally we will strive to investigate fuzzy determination of requirements by analysis of historical data (e.g. by log information about given business processes), therefore we will evaluate mechanisms of process mining and related disciplines (e.g. as described in [6, 8]).

References

1. Baumgartner, P., Häfele, H., Maier-Häfele, K.: E-Learning Standards aus didaktischer Perspektive. In: Bachmann, G., Haefeli, O., Kindt, M. (eds.) Campus 2002: Die virtuelle Hochschule in der Konsolidierungsphase, pp. 277–286. Waxmann Münster (2002)
2. Du, W., Shan, M.C.: Enterprise Workflow Resource Management. In: Proceedings of the Ninth International Workshop on Research Issues on Data Engineering: Information Technology for Virtual Enterprises, RIDE 1999 (1999)
3. Hlaoittinun, O., Bonjour, E., Dulmet, M.: Managing the competencies of team members in design projects through multi-period task assignment. In: Camarinha-Matos, L.M., Boucher, X., Afsarmanesh, H. (eds.) PRO-VE 2010. IFIP AICT, vol. 336, pp. 338–345. Springer, Heidelberg (2010)

4. IEEE Standard for Learning Object Metadata, IEEE Std 1484.12.1-2002, pp. i–32 (2002), doi: 10.1109/IEEESTD.2002.94128
5. Knolmayer, G.: E-Learning-Objects. Wirtschaftsinformatik 46 (2004)
6. Moe, W., Rozinat, A., van der Aalst, W.M.P., Hofstede, A., Fidge, C.: Process Mining and Simulation. In: Modern Business Process Automation, pp. 437–457 (2010)
7. Oberweis, A., Schuster, T.: A meta-model based approach to the description of resources and skills. In: Proceedings of the 16th Americas Conference on Information Systems, AMCIS 2010 (2010)
8. Sun, P., Tao, S., Yan, X., Anerousis, N., Chen, Y.: Content-Aware Resolution Sequence Mining for Ticket Routing. In: Hull, R., Mendling, J., Tai, S. (eds.) BPM 2010. LNCS, vol. 6336, pp. 243–259. Springer, Heidelberg (2010)
9. Polsani, P.: Use and Abuse of Reusable Learning Objects. In: E-education: Design and Evaluation, North America, vol. 3 (2003), http://journals.tdl.org/jodi/article/view/89/88
10. Roman, B., Grassmann, O., von Zedtwitz, M.: Managing Knowledge and Human Resources. In: Managing Global Innovation, pp. 273–288 (2008)
11. Russell, N., van der Aalst, W.M.P., ter Hofstede, A.H.M., Edmond, D.: Workflow Resource Patterns: Identification, Representation and Tool Support. In: Pastor, Ó., Falcão e Cunha, J. (eds.) CAiSE 2005. LNCS, vol. 3520, pp. 216–232. Springer, Heidelberg (2005), http://ieeexplore.ieee.org/servlet/opac?punumber=8032
12. van der Aalst, W.M.P., Kumar, A.: Xml-based schema definition for support of interorganizational workflow. Information Systems Research 14(1), 23–46 (2003)
13. Wiesner, A.: Activity Tree Harvesting: Entdeckung, Analyse und Verwertung der Nutzungskontexte SCORM-konformer Lernobjekte, pp. 75–77. KIT Scientific Publishing Karlsruhe (2010)
14. Wiley, D.A.: Connecting learning objects to instructional design theory: A definition, a metaphor, and a taxonomy. In: GPCE 2003. Learning Technology, vol. 2830, pp. 1–35 (2000)
15. Winterton, J., Delamare-Le Deist, F., Stringfellow, E.: Typology of knowledge, skills and competences: clarification of the concept and prototype. In: Cedefop (2005)
16. zur Mühlen, M.: Organizational management in workflow applications—issues and perspectives. In: Information Technology and Management, vol. 5, pp. 271–291. Springer-Author Discount, Appendix (2004)

Considerations in Introducing a Blended Learning Environment Dedicated to Training Computer Network Administrators

Bartosz Marcinkowski and Piotr Ostrowski

University of Gdansk, Department of Business Informatics
Piaskowa 9, 81-864 Sopot, Poland
{bartosz.marcinkowski,piotr.ostrowski}@ug.edu.pl

Abstract. Increasing the flexibility of ICT teaching process may be achieved by extending the scope of application of e-learning tools. In order to reach candidates who may not fit the rigid course schedules, Cisco Academies consider application of simulation, emulation as well as remote equipment management environments. The goal of the current article to examine the validity and the economic viability of extending usage of e-learning components by Academy Support Center / Instructor Training Center at the University of Gdansk (UG) beyond knowledge transfer domain and covering skills development domain. Questionnaire survey was used to capture learning preferences of all the candidates to take part in networking courses in 2012 fall. Selected survey results were analyzed and discussed in the paper.

Keywords: computer network, e-learning, simulation, emulation, Cisco, survey.

1 Introduction

Due to the recent economic stagnation – combined with high competition observed on the European market of certified training for networking professionals – it is vital to maintain quality, seek new content-related development areas as well as incorporate innovations within training process itself. It is e-learning that is among such innovations aimed at increasing the flexibility of teaching process. Education institutions in Poland largely implemented world achievements in the field of e-learning and in many cases recorded significant successes in this field. The purpose of this article to examine the validity and the economic viability of extending usage of e-learning components by Academy Support Center / Instructor Training Center at the University of Gdansk (UG) beyond knowledge transfer domain and covering skills development domain. The academy (UG, 2013) is a partner in Cisco Networking Academy Program that is one of the leading network administration-oriented training systems on a global scale. A questionnaire survey was used to capture the preferences of respondents (who were in fact candidates to take part in networking courses in 2012 fall). No sampling methods were used – all members of the population were

S. Wrycza (Ed.): EuroSymposium 2013, LNBIP 161, pp. 85–93, 2013.

asked to provide opinions within a given timeframe. From 53 questionnaire surveys that were handed to potential participants, 40 were filled in and returned. Such surveys are planned and executed by UG on a regular basis (see Marcinkowski and Ostrowski, 2010, 2012).

It should be noted that the market for professional training regarding network administration training is so non-typical that it has a relatively significant entry barriers in comparison to other educational initiatives (just to mention the cost of network equipment or the expenses necessary to secure industrial certification at least two instructors in each track and renew such certification every two years). Activities that are carried out by students on the networking hardware are crucial to the success of the training. Having said that, teams responsible for authoring e-learning content that supports classroom instruction and development of practical skills include more and more skill-oriented components feasible outside the training room. Such components are willingly used by instructors to expand core of the courses and as a tool of consolidating skill set acquired in the classroom. Such an approach is intended primarily to make the training process more flexible and therefore meet time constraints of potential customers – it is a paradox that in the case of the more specialized training it is challenging to ensure that the classroom is occupied frequently enough to justify investments in expensive network switches/routers while there is a shortage of convenient dates for students working in the industry. While such restrictions do not apply to academies located by colleges and universities (owing to the fact that the equipment supports teaching of regular students as well; moreover, dates that are convenient for university students do not collide with dates preferred by industry students and vice versa), the situation of institutions that do not have a balanced customer base leaves absolutely no margin for error while selecting a training path to be invested in.

2 Simulation/Remote Equipment Management Environments

Technical means of supplementing classroom-executed skills development (or even replacing practical exercises with networking hardware in extreme cases) include first of all simulators and emulators used in conjunction with pre-prepared network topologies. It should be noted though that implementing any form of simulation is usually a compromise and comes with a risk of impoverishing skill set acquired. Hence, networking academies vitally interested in adopting flexible timetables may consider professional environments for remote hardware management. Having said that, it should be noted that even in such case mastering physical layer-related issues may prove to be challenging and special attention of couches is required in this respect.

When network simulating software is concerned, market research was limited to easily available solutions for simulating Cisco network devices, as devices offered by this very manufacturer are of interest to the academy. Cisco Systems Inc. does not seem to see simulators as a threat to their learning initiative as the corporation provides its own solution, which is Cisco Packet Tracer. The Packet Tracer is a

network simulation and visualization program that allows students to experiment with network behavior and ask "what if" questions; it supplements physical equipment in the classroom by allowing students to create a network with an almost unlimited number of devices, encouraging practice, discovery, and troubleshooting (Cisco Systems Inc., 2013). Selecting the Packet Tracer as recommended network simulation software by an academy is a reasonable solution as that does not involve additional licensing costs – the tool is available free of charge to instructors, students, alumni, and administrators that are registered Academy Connection users. As mentioned earlier, Cisco learning initiative evolves to intensify usage of simulation software by including pre-prepared Packet Tracer network topologies in the courses' curricula as well as proposing SBA practice exams (executed online via Academy Connection) as an alternative to hand-on practical exams concluding each course/semester. Such alternative may be found attractive in respect of formal quantitative requirements regarding the networking hardware – while UG decided to assign each student an individual network device in order to comfortably carry out routine, equipment sharing is a common practice.

It is Boson NetSim that is an advanced competitive solution to the Packet Tracer. It is a proprietary software that simulates Cisco networking hardware and software and is designed to aid the user in learning the Cisco IOS command structure (Boson Software LLC, 2012) not only on Cisco Certified Network Associate (CCNA) level, but on the Cisco Certified Network Professional (CCNP) as well. Depending on the edition purchased (NetSim for CCENT / NetSim for CCNA / NetSim for CCNP) it supports an extensive functionality, up to 48 device types as well as up to 196 lab activities.

Instructors interested in providing network device management experience indistinguishable from configuring actual hardware via a virtual terminal connection ought to go beyond simulation and consider hardware emulators. It is Dynamips that is popular software that emulates Cisco IOS on a traditional PC (Fillot, 2007). It is commonly used in conjunction with a text-based front ends such as Dynagen (Anuzelli, 2007) or graphical ones, of which the leading solution is GNS3 (Grossmann et al., 2013). Such tool packages support networking technologies and features unattainable for conventional simulator and may be obtained free-of-charge. On the other hand, users are to provide their own IOS images and relatively powerful PCs, as the system requirements of emulation software are significant. It should be stressed that using IOS images on emulation software requires resolving IOS license-related legal issues.

Last but not least, Cisco Academy may take a completely different approach to implementing e-learning components regarding hand-on exercises. It may support ITC training with real IT equipment, virtual machines and lab content on the Internet using dedicated, remote equipment management environments such as NETLAB+ (Network Development Group, 2013). Such environments enable not only remote access to hardware in academy's possession, but allow for automation regarding hardware management, advanced scheduling capabilities and centralized management.

3 Selected Survey Results

For the current paper, a set of organization- and cost-related questions were selected. The first issue is based on the assumption that all the courses shall be held in accordance with traditional or blended-learning regime. In such context, the dominant preference of the respondents is to organize courses on workday evenings. This option was selected by 62.5% of professional training participants (Fig. 1.). While the intensity of events is derived from the individual group members needs and can be flexibly adapted to the needs, there are a very limited number of cases in which more than one meeting per week would be advised. Fridays are the least preferred workday for running courses. For weekend training sessions (which were the first choice of the 37.5% of the respondents) no preference regarding a dominant day or time was indicated.

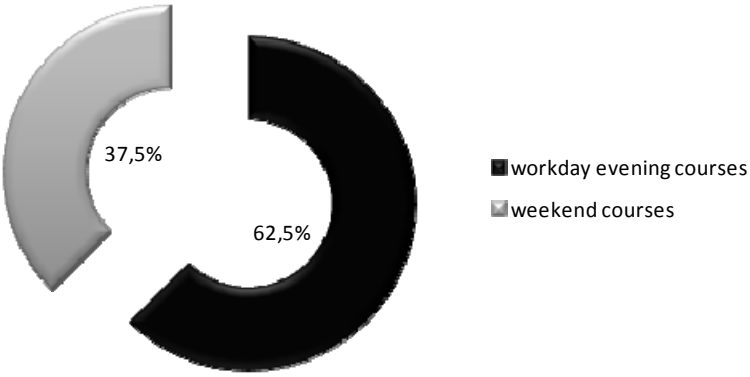

Fig. 1. Preferences for scheduling courses

Apart from the well-embedded in practice arrangement of Cisco courses offerings that include an extensive e-learning component addressing comprehensive theory-oriented content available online as a basis to conduct practical exercises using physical network equipment and/or network simulator (blended learning), the expectations of people registered for training vary greatly. While a total of 84.4% of people prefer or simply tolerate an online component, 15.6% of respondents are still faithful to the traditional teaching methods only (Fig. 2). Experience shows that course participants within this last group quickly adapt to blended methods of professional training, trading-off the typical presentations and lectures for possibility to take part in individualized problematic discussions after studying the content on their own. This group is often, however, inherently averse to methods based on the simulator. It should be noted that the discomfort related to the non-traditional forms of training often originates from language barriers – due to the fact that the most popular CCNA Exploration course is generally localized only partly. First two semesters, corresponding to the Entry-level certification are the main focus for translation projects. Associate-level (semesters three & four) is rarely localized.

The vast majority of respondents asked about the admissibility of running the course in a completely remote form were reluctant to such a form. Only 18.8% of students chose this as the most desirable form. Characteristically, the response is consistent with the overall, binary evaluation of willingness to register for training that would be carried out solely in e-learning form. Therefore the group of e-learning mode supporters turns out to be very airtight. Moreover, other detailed questions regarding individual preferences for optimal teaching methods and tools without introducing any additional incentives confirmed the initial declarations. It was the blended learning – a de facto standard of Cisco Academy offerings – that was assessed as the optimal form for carrying out courses by 34.4% of participants. Additionally, 31.3% of them did not state their preference. It should be noted, however, that this group previously ruled out the solely e-learning mode – so in this case the presumed acceptance of blended learning methods is assumed.

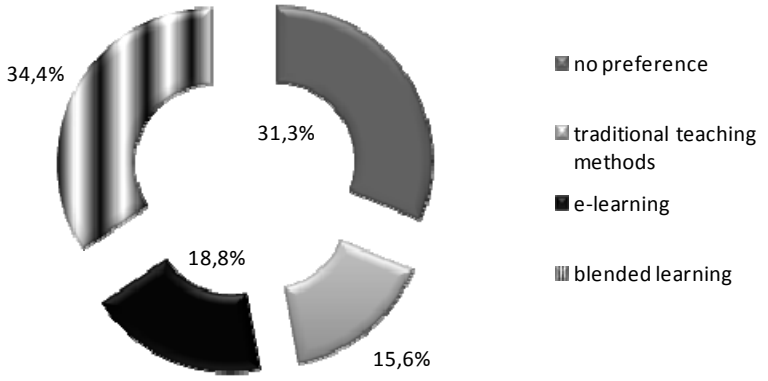

Fig. 2. Preferences regarding teaching methods

Building upon results synthetically included in Fig. 2, participants of the courses that are held on weekends are particularly reluctant to exclusively online education – only as much as 16.7% of the respondents preferring to develop skills on Saturdays and Sundays would be willing to take a course in all remotely. Additionally, most of the people that expected the lowest course intensity originated from this group. They often emphasized the importance of direct contact with the networking hardware, the possibility for tailoring network topologies to individual needs as well as more direct feedback regarding skills gained than in the case of connecting with the same equipment remotely using specialized software.

Respondents were also asked about the expected level of discount pricing in case they registered for entirely remote CCNA Exploration editions (Fig. 3). The participants were not presented predefined options, but were supposed to specify the percentage of their own.

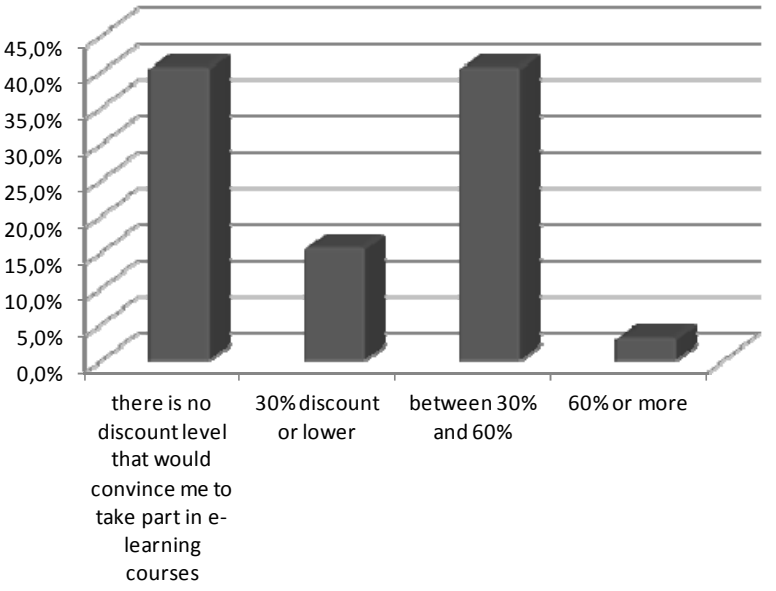

Fig. 3. Expected levels of discount pricing for 100% e-learning courses

40.6% of the respondents confirmed the previously declared lack of interest in exclusively remote training mode, indicating that no discount level shall affect their decision. After confronting the results of a binary indication of eligibility to take training in strictly remote mode, about half previously anti-e-learning course candidates would be convinced with price incentives and is willing to join the remote courses. At the same time, respondents expected high levels of discounts – only 15.6% of them would be willing to accept discount of 30% or lower. In addition, almost every second of those requiring reasonable discount levels is a person who by definition was determined to take on e-learning courses and had not considered its decision in terms of cost. The same percentage of candidates that reject the remote mode expects discounts between 30 and 60%, and 3.1% – more than 60%. Due to the fact that the overall costs of delivering remote training within configuration and maintenance of network equipment domain differ only slightly from traditional forms of training (in case of conducting a limited number of courses' editions using remote access to the physical network equipment such costs may even exceed the costs of traditional training), high price-related pressure restrict the market for e-learning to a large extent.

At the same time it should be noted that across the analyzed population the respondents were in general agreement as to the expected price levels (Fig. 4). Median result of the entire study population was exactly 50%. Interestingly, people with pre-exclusion to study online and – at the same time – declaring to change their decision should discounts system be introduced, expected statistically lower discount levels than those opting for this form from the very beginning. For skeptics, it was in fact an average of 40.8% (standard deviation 13.2%), while for pro-e-learning people – as much as half the price (standard deviation 6.3%).

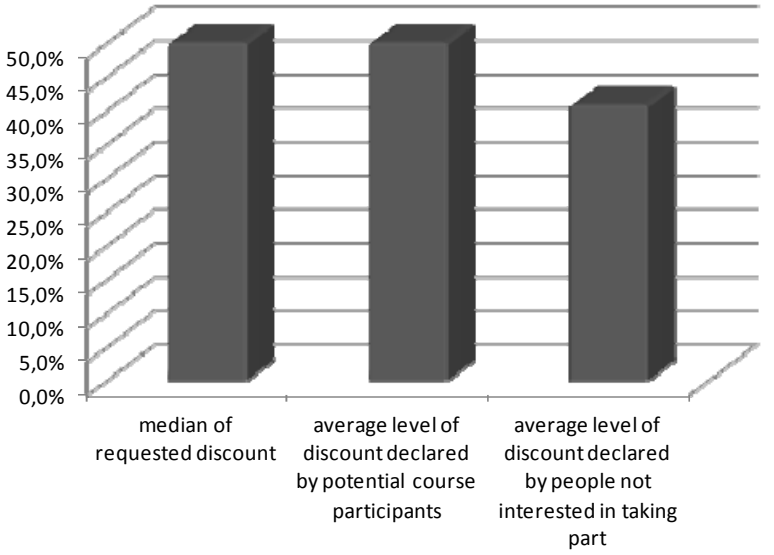

Fig. 4. Averages for expected price levels

It is a form of conducting hand-on exercises that is a crucial factor for the implementation of online professional training. Such exercises provide in fact the greatest added value for the participants – especially for those currently working in the industry and thus possessing significant theoretical foundations. What may be a bit surprising in such context is an extraordinarily high level of acceptance for the Cisco Packet Tracer tool as primary software to perform exercises (see Fig. 5). It was preferred by 71.4% of individuals who have decided to let go of direct interaction with the networking hardware. It should be noted that this product, while being a solution that is mature, continuously improved and offering a great deal of physical equipment's functionality, is only a simulator – not an emulator – and still not sophisticated enough to compete on an equal footing with the actual hardware as a teaching tool.

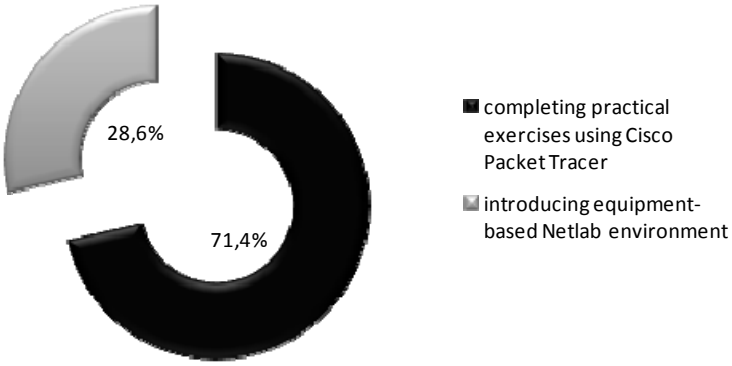

Fig. 5. Preferences regarding tools for out-of-classroom practical training

From the professional training's supplier point of view it is beneficial news, since it does not lead to the formation of additional licensing costs and at additionally does not involve physical equipment of the academy, which may be at the same time intended to perform other tasks of teaching. On the other hand, 28.6% of respondents expect to implement environmentally sound management of remote access to the physical hardware and the tasks performed on the equipment (for instance NETLAB+). It is significant that these candidates have lower expectations regarding discounts, ranking in the majority in the first compartment (which is up to 30%).

An alternative approach, used especially by commercial training centers, is to separate the issue of online knowledge transfer, participation in assessments and other components of the training from all-day (or even a few days) long practical sessions, appointed every semester. It is as much as 12.5% of the candidates that allow practical sessions only on the academy's equipment while pursuing the course on the basis of remote teaching. That makes this form inferior in terms of attractiveness compared to both the traditional mode of education at the academy's premises, as well as considered variants of e-learning.

4 Summary

Both authors' practical experience and the results of the survey clearly indicate that offerings of courses conducted in full accordance with e-learning principles should be thoroughly studied in terms of risk. While the online transfer of ICT-related knowledge is enthusiastically accepted as well as course organization that allows studying the course content at homes based upon previously imposed schedule works fine in practice, applying the same principles to hand-on exercises shall be successful only in some cases. A typical student is in fact expecting a blended formula, in which he/her is able to verify knowledge acquired and systematically develop practical skills over a series of classroom meetings moderated or run by certified instructors. Thus, the implementation of 100% e-learning initiatives is recommended to be preceded by its own preference research within the recruitment process and based on e-learning environments that do not involve additional licensing costs. Blended form is also actively supported by Cisco by offering useful tools and SBA exams, what does not make considering other simulation environments pointless – while Packet Tracer proves to be useful on Associate level, it lacks many features required to act as a primary tool on higher levels. In this context it would be wise to take a closer look at GNS3 emulation environment, which provides added value nearly as high as physical networking hardware. Characteristically, there was no significant advantage of remote equipment management environments reported over emulators or even ordinary hardware simulators. Ipso facto, the implementation of such environments by UG was deemed uneconomic and discontinued.

References

1. Academy Support Center / Instructor Training Center at the University of Gdansk, cna.gda.pl (2013)
2. Anuzelli, G.: Dynamips / Dynagen Tutorial. Documentation Revision 1.11.7 (2007), http://dynagen.org/tutorial.htm
3. Boson Software LLC: NetSim 8. User Manual (2012), http://www.boson.com/Files/Support/NetSim-8-User-Manual.pdf
4. Cisco Systems Inc.: Cisco Packet Tracer (2013), http://www.cisco.com/web/learning/netacad/course_catalog/PacketTracer.html
5. Fillot, C.: Cisco 7200 Simulator (2007), http://www.ipflow.utc.fr/index.php/Cisco_7200_Simulator
6. Grossmann, J., Marsili, B., Goudjil, C., Baum, P., Benichou, M., Boudigues, A., Menetrier, J., Michel, M., Millet, A.: GNS3 Documentation (2013), http://www.gns3.net/documentation
7. Marcinkowski B., Ostrowski P.: Activities and Achievements of Pomeranian Regional Cisco Academy in Preparing ITC Specialists for Pomeranian Market between 2002 and 2007 (in Polish); Goliński J. (ed.): Modern Aspects of Information. Warsaw School of Economics Press (2010) (in Polish)
8. Marcinkowski, B., Ostrowski, P.: Professional Training Regarding Network Infrastructure Management Assessed, "E-mentor" (5) (2012) (in Polish)
9. Network Development Group: NETLAB+ Product Overview (2013), http://www.netdevgroup.com/products

Author Index

Bajaj, Akhilesh 39
Becker, Christoph 76

Doucek, Petr 17
Dubauskaite, Ruta 1

Jatkiewicz, Przemyslaw 50

Kalibatiene, Diana 1

Marcinkowski, Bartosz 85
Maryska, Milos 17
Maślankowski, Jacek 66

Nedomova, Lea 17

Ostrowski, Piotr 85

Rahimian, Firoozeh 39

Schuster, Thomas 76
Soja, Piotr 27

Trąbka, Jan 27

Vasilecas, Olegas 1