WIRTSCHAFTSINFORMATIK
UNIVERSITÄT ROSTOCK

**Birger Lantow**
**Kurt Sandkuhl**
**Ulf Seigerroth**
**(Eds.)**

# 6th International Workshop on Information Logistics, Knowledge Supply and Ontologies in Information Systems

**12th International Conference on Perspectives in Business Informatics Research**
**Warsaw, Poland, September 23-25, 2013**
**Workshop Proceedings**

Universität Rostock
Traditio et Innovatio

JÖNKÖPING UNIVERSITY

**Title**
6th International Workshop on Information Logistics, Knowledge Supply and Ontologies in Information Systems

**Sub-title**
12th International Conference on Perspectives in Business Informatics Research
Warsaw, Poland, September 23-25, 2013
Workshop Proceedings

**Volume Editors**
Birger Lantow
The University of Rostock
Institute of Computer Science
Albert-Einstein-Str. 22
18 057 Rostock, Germany
E-mail: birger.lantow@uni-rostock.de

Kurt Sandkuhl
The University of Rostock
Institute of Computer Science
Albert-Einstein-Str. 22
18 057 Rostock, Germany
E-mail: kurt.sandkuhl@uni-rostock.de

Ulf Seigerroth
Jönköping University
School of Engineering
Box 1026
55 111 Jönköping, Sweden
E-mail: ulf.seigerroth@jth.hj.se

# Preface

The field of information logistics and knowledge supply investigates methods, technologies and applications supporting demand-oriented and efficient information and knowledge provision in enterprises. In particular in knowledge-intensive industry and service sectors, information is a major factor in production processes, and knowledge reflects an important asset of the enterprise. Similarly, public organizations and governmental bodies are dependent on accurate and timely information supply for efficient and high quality processes and services. Intelligent information supply has become an important issue that is characterized by just-in-time, demand-oriented and context-sensitive information.

Ontologies are a knowledge representation technique of growing importance - not only for intelligent information supply but equally for information systems and industrial applications. In conjunction with evolving semantic technologies for ontology engineering, representation and query, new ways open up for ontologies and information systems (IS) integration, combination and use. In the structural perspective, ontologies can provide means to structure, store and access generic IS content. In the temporal perspective, ontologies can guide the development of new IS. They may help to choose appropriate processes, algorithms, rules, and software components depending on the requirements.

ILOG 2013 — the 6$^{th}$ International Workshop on Information Logistics, Knowledge Supply and Ontologies in Information Systems — had the aim to bring together people who have a strong interest in the innovative use these technologies and approaches in the context of enterprises and public organizations. The workshop took place on September 23, 2013, in Warsaw (Poland) co-located with the 12$^{th}$ International Conference on Perspectives in Business Informatics Research (BIR 2013). Based on at least three reviews per submission the international Program Committee selected 7 high-quality papers for inclusion in this volume. The authors of these papers include both researchers and practitioners from different disciplines. The ILOG 2013 program reflects different facets of the workshop topics, including organizational and social issues, as well as methodical and technical aspects related to information logistics and information systems.

We dedicate special thanks to the members of the international Program Committee for promoting the conference, their support in attracting high-quality submissions, and for providing excellent reviews of the submissions. Without their committed work a high-quality workshop like ILOG 2013 would not have been possible. Our thanks also include the external reviewers supporting the paper selection process.

September, 2013

Birger Lantow
Kurt Sandkuhl
Ulf Seigerroth

# Organization

## ILOG 2013 Chairs

| | |
|---|---|
| Birger Lantow | University of Rostock, Germany |
| Kurt Sandkuhl | University of Rostock, Germany |
| Ulf Seigerroth | University of Jönköping, Sweden |

## ILOG 2013 Program Committee

Witold Abramovicz, Poznan University of Economics, Poland
Clara Bassano, “Parthenope” University of Naples, Italy
Edward Babkin, Higher School of Economics Nizhny Novgorod, Russia
Andreas Billig, Fraunhofer FOKUS, Germany
Eva Blomqvist, Linköping University, Sweden
Andreas Both, Unister Holding, Germany
Susanne Busse, FH Brandenburg, Germany
Olov Candell, Saab Group, Sweden
Albertas Caplinskas, Vilnius University, Lithuania
Wolfgang Deiters, Fraunhofer ISST, Germany
Barbara Dinter, Technical University Chemnitz, Germany
Henrik Eriksson, Linköping University, Sweden
Janis Grabis, Riga Technical University, Latvia
Darek Haftor, Linnæus University, Sweden
Karl Hammar, Linköping University, Sweden
Yanbo Han, Inst. Computing Techn., Chin. Acad. of Science, China
Tomasz Kaczmarek, Poznan University of Economics (Poland)
Mira Kajko-Mattsson, Stockholm University, Sweden
Ramin Karim, Luleå University of Technology, Sweden
Marite Kirikova, Riga Technical University, Latvia
Ralf-D. Kutsche, Technical University Berlin, Germany
Birger Lantow, Rostock University, Germany
Tatiana Levashova, St. Petersburg Institute Inform. and Autom., Russia
Anita Mirijamdotter, Linnæus University, Sweden
Stefan Pfennigschmidt, Fraunhofer FOKUS, Germany
Paolo Piciocchi, University of Salerno (Italy)
Kurt Sandkuhl, Rostock University, Germany
Ulf Seigerroth, Jönköping University, Sweden
Nikolay Shilov, St. Petersburg Institute Inform. and Autom., Russia
Alexander Smirnov, St. Petersburg Institute Inform. and Autom., Russia
Janis Stirna, Stockholm University, Sweden
Renate Strazdina, Ernst and Young, Latvia
Vladimir Tarasov, Jönköping University, Sweden
Jan Willem, Nashuatec/Ricoh Group, Netherlands
Oleg Zaikin, West Pomeranian University of Technology, Poland

# Table of Contents

## Information Logistics Concepts, Technologies and Applications



## Ontologies and Information Systems

# Joint Optimization of Physical and Information Flows in Supply Chains

Jānis Grabis

Riga Technical University, Kalku 1, Riga, LV-1658, Latvia
grabis@rtu.lv

**Abstract.** Supply chain units are connected by both physical and information flows. Electronic services are used to perform a large part of activities in modern supply chains. The traditional supply chain configuration models pay the main attention to optimization of the physical flows. However, it is equally important to ensure that the physical units have appropriate information processing and logistics capabilities. Therefore, a model for joint configuration of the physical and information flows in e-retailing supply chains is elaborated in this paper. The model simultaneously identifies appropriate suppliers, selects the third party logistics service provider and selects appropriate electronic information processing services. The typical services used in global e-retailing are identified. The services are characterized by their functionality and quality of service measurements. The preliminary experimental studies demonstrate interdependencies between the physical and information flows.

**Key words:** supply chain, information flow, service selection

## 1 Introduction

A supply chain is the network of interrelated companies collaborating to serve its customers. It has long been acknowledged that supply chain management concerns with both physical movement of products from suppliers to customers as well as with information flows for synchronizing the supply chain management processes [1]. Meanwhile the distinction between the physical and information flows is becoming less visible in modern supply chains as many supply chain functions are digitalized. Nevertheless, supply chain management methods continue to analyze both flows separately.

This paper attempts to elaborate a supply chain configuration model dealing with both physical and information flows. The supply chain configuration is one of the key supply chain management sub-problems dealing with selection of supply chain units and establishing connections among the units [2]. The joint supply chain configuration implies that the model selects units and establishes the connections to ensure movement of physical goods as well as integrated information processing. The model explores configuration of an e-retailing supply chain, where the e-retailer identifies appropriate suppliers and selects a third-party logistics services provider and simultaneously selects appropriate web services for enabling the information flow in the supply chain. The model combines the traditional supply chain configuration problem [3] with the web service selection problem [4]. The contribution to the field of information logistics is demonstration of mutual interdependencies of the physical supply chain configuration and the web service selection. It is shown that the physical supply chain structure depends upon availability of appropriate information processing services ensuring information integration and information logistics in electronic supply chains.

The paper expands a traditional information sharing approach [5] of studying information flows in supply chain. Klein and Rai [6] suggest that a strategic approach is needed to information integration in supply chains, and Swaminathan and Tayur [7] identifies opportunities for using emerging information technologies in e-business supply chain. Improvements in information logistics is one of options for improving information flows. That is also confirmed by findings that information accuracy and relevance is among the key factors affecting web-site quality in e-business [8]. A hub based approach can be used in integrated the physical and information flows in supply chains [9]. However, in the case of highly distributed and heterogeneous supply chains as in e-retailing, a service oriented approach could be a more attractive option [10].

The rest of the paper is organized as follows. The physical and information flows in the e-retailing supply chain are described in Section 2. These are represented in a form of business process models. These models capture the flows at the strategic level. A mathematical supply chain model for joint optimization of the physical and information flow is formulated in Section 3. Section 4 reports results of the experimental studies conducted to demonstrate interdependencies between the physical and information flows. Section 5 concludes.

## 2 Supply Chain Flows

A model for joint design of the physical and information flows is developed for supply chains, where a significant part of supply chain activities take place in an electronic form. Supply chains by e-retailers such as Amazon.com, Macy's[1] belong to this group of supply chains. The physical flow represents the flow of products from suppliers to e-retailers facilities and final delivery of products to end-customers usually done by a 3PL provider. The information flow represents different on-line services to customers and supply chain partners. These services include product information services, payment services, insurance services, shipment tracking services and others. The services can be provided by the same partners providing the physical processing or by partners specializing in delivering electronic information processing services. For instance, Borderfree[2] acts as an integrator for e-retailers providing end-to-end information processing services.

[1] http://www.amazon.com, http://www.macys.com
[2] http://www.borderfree.com

### 2.1 Physical and Internal Information Flow

A business process model is used to represent the physical and information flows and their processing in supply chains. It is assumed that the physical data flow and supply chain units mainly dealing with processing of physical products are represented as a single entity while supply chain units mainly dealing with information processing are represented as independent units. Therefore, the physical supply chain units are represented in the business process as lanes in a single pool (Fig. 1), and the electronic supply chain units are represented as separate pools (see Section 2.2).

The physical flow of products is initiated by detecting the customer demand without specifying how the demand is detected. Suppliers are responsible for supplying the products. The e-retailer is the focal unit in the e-retailing supply chain and its main task is to sell products to customers. The e-retailer can also operate storage and distribution facilities (see [11] on various options for products processing in e-retailing supply chains). The 3PL providers are responsible for delivery of products to the customers. The internal information flow accompanies the physical flow of products. It is referred as to the internal information flow because information processing is perceived as an essential part of the physical products processing tasks. Data objects are used to represent internal supply chain information flows. Only the main data objects such as sales order, purchasing requisition, delivery note and delivery confirmation are referenced in the model. This supply chain representation does not include the reverse supply chain flow for simplicity.

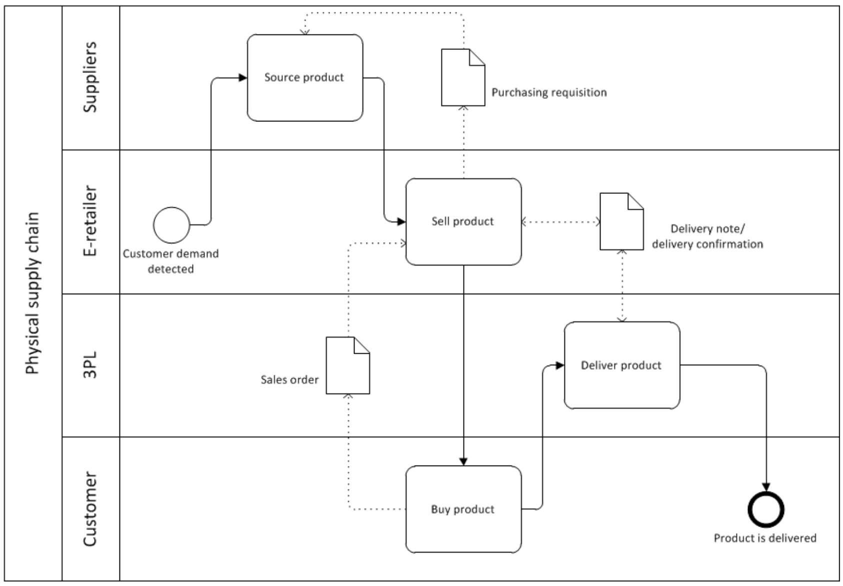


**Fig. 1.** The physical flow in supply chains

### 2.2 Message and Integrated Flow

An integrated physical and information flow model is created in order to capture interrelationship among the physical and electronic supply chain units. The electronic supply chain units are represented each in a separate pool named as a service unit with a specific type. These pools represent abstract service providers. The actual service providers can provide several of the services required, some of them act as service aggregators and some services can be provided by the physical supply chain units. The interrelationships are shown as message flows among the pools. The message flow shows only purely electronic information processing activities. For instance, a shipment activity includes shipment data processing, shipment confirmation and other operations but these information processing activities are perceived as an essential part of the physical activities and are included in the Deliver product task.

The model defines main types of the electronic service units present in the e-retailing supply chains in the global setting. These types include:

- Product information services – detailed information possibly aggregated from multiple sources is provided about each product offered by the e-retailer
- Import/export services – checks on import and export restrictions from one country to another for certain products, i.e., the service rejects selling a product in certain countries where specific licensing rules are applicable
- Customs and taxes services – calculation of appropriate taxes depending upon the customer location is performed
- Payment services – multi-currency processing of payments using different payment channels is performed and restrictions concerning availability of the payment channels are applied, e.g., credit cards only from specified countries are accepted
- Shipment services – if multiple shipment modes are available the most appropriate alternative with regards to the destination and delivery time is determined and shipment tracking is provided independently of the 3PL provider is provided, especially, if multiple logistics providers are used for delivery.

The list of service types is not exhaustive and other types of services can be used such as fraud detection and shipment insurance. Fig. 2 shows the physical e-retailing supply chain process along with the necessary electronic services. The expansion of the Buy product task is given in Fig. 3. The message flow for this task is shown only at the sub-process level. The product information service provides information to the Sell product task and is responsible for providing as rich information about the product as possible. The shipment service is invoked during the product delivery to provide opportunities for tracking the product delivery.

Majority of the message flows are associated with the Buy product task. The services are invoked to provide an accurate estimate of the total ordering costs for the customer. For local e-retailers, this operation usually is straightforward but much more comprehensive information should be gathered for global e-retailers. The message flows should ensure information about applicable taxes, import/export restrictions, delivery options and international payment processing. This information is specific to the customer location.

In the given process model it is assumed that the e-retailer manages both the physical and electronic sales process by itself. Another possibility is that a traditional retailer deals only with the physical sales while the electronic part is provided by a sales service provider.

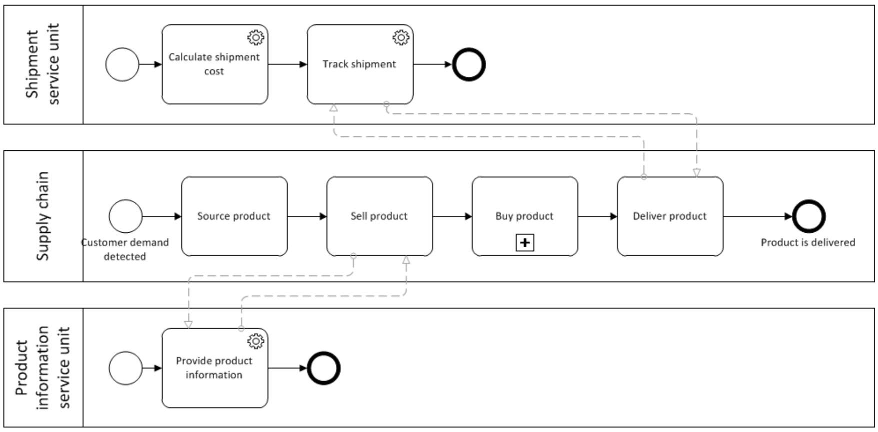


**Fig. 2.** The message flow between the supply chain and the product information and shipment services

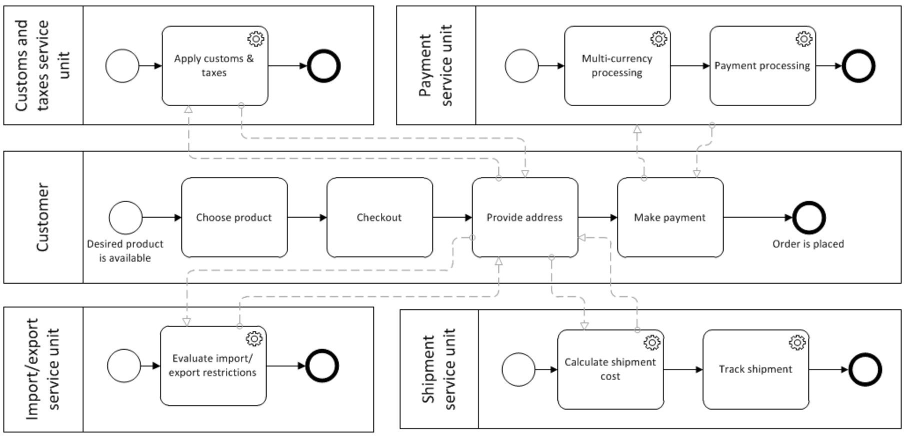


**Fig. 3.** The message flow for the Buy product task

## 3 Model

The supply chain business process models show interactions among the physical and electronic supply chain units. The supply chain configuration problem is to select suitable physical and electronic supply chain units to optimize supply chain performance. In the case of e-retailing supply chain, products' suppliers, third party logistic provider and web services for information processing are selected. The supply chain performance is measured by supply chain profitability and customer satisfaction affected by efficiency of information processing. The profitability is calculated as revenues from product sales after deducting sales expenses minus sourcing, delivery and unit setup costs. The information processing efficiency is calculated as a weighted sum of web service Quality-of-Service (QoS) criteria, namely, response time, error rate and reliability, which are among the most frequently used QoS criteria [4].

The mathematical formulation of the model consists of the objective function (Eq. 1) and constraints (Eq. 7-13), the notations used are defined in Table 1. The weights $w_1$ and $w_2$ are used to combine the physical units selection and web service selection criteria in a single objective function. Eq. 6 evaluates the information processing efficiency for the selected web services. Importance of the each QoS criteria is determined by the weight factor $v_i$. Eq. 5 evaluates the fixed cost incurred by incorporating physical or electronic units in the supply chain.

$$P(\mathbf{X}, \mathbf{Y}, \mathbf{Y}) = w_1(R - C_1 - C_2 - C_3) + w_2 L \rightarrow max \quad (1)$$

$$R = \sum_{i=1}^{N_p} \sum_{j=1}^{N_c} \sigma_{ij} S_{ij} \quad (2)$$

$$C_1 = \sum_{i=1}^{N_p} \sum_{j=1}^{N_v} \pi_{ij} Q_{ij} \quad (3)$$

$$C_2 = \sum_{i=1}^{N_p} \sum_{j=1}^{N_l} \sum_{k=1}^{N_c} \delta_{ijk} U_{ijk} \quad (4)$$

$$C_3 = \sum_{i=1}^{N_v} \lambda_i^1 X_i + \sum_{i=1}^{N_v} \lambda_i^2 Y_i + \sum_{i=1}^{N_v} \lambda_i^3 Z_i \quad (5)$$

$$L = \sum_{i=1}^{N_s} \sum_{j=1}^{3} v_j \beta_{ij} Y_i \quad (6)$$

$$S_{ij} \leq d_{ij}, i = 1, ..., N_p, j = 1, ..., N_c \quad (7)$$

$$\sum_{j=1}^{N_l} U_{ijk} \leq S_{ik}, i = 1, ..., N_p, k = 1, ..., N_c \quad (8)$$

**Table 1.** Notation

| Notation | Description |
|---|---|
| $N_p$ | number or products |
| $N_c$ | number of countries where customer are located |
| $N_v$ | number of potential suppliers |
| $N_l$ | number of potential 3PL providers |
| $N_s$ | number of potential services |
| $P$ | e-retailer's profit |
| $X_i \in \{0,1\}$ | a decision variable indicating whether the $i$th supplier is selected or not |
| $Y_i \in \{0,1\}$ | a decision variable indicating whether the $i$th service is selected or not |
| $Z_i \in \{0,1\}$ | a decision variable indicating whether the $i$th 3PL provider is selected or not |
| $S_{ij}$ | a decision variable determining the quantity of the $i$th product sold to customer in the $j$th country |
| $Q_{ij}$ | a decision variable determining the quantity of the $i$th product sourced from the $j$th supplier |
| $U_{ijk}$ | a decision variable determining the quantity of the $i$th product delivered by the $j$th 3PL provider to the $k$th country |
| $\sigma_{ij}$ | revenues from each item of the $i$ product sold in the $j$th country |
| $\pi_{ij}$ | purchasing prices of the $i$th product from the $j$th supplier |
| $\delta_{ijk}$ | delivery cost for the $i$th product by the $j$th 3PL provide to the $k$th country |
| $\lambda_i^1$ | the setup cost for the $i$th supplier |
| $\lambda_i^2$ | the setup cost for the $i$th service |
| $\lambda_i^3$ | the setup cost for the $i$th 3PL provider |
| $\beta ij$ | the value of the $j$th QoS attribute for the $i$th service |
| $d_{ij}$ | demand for the $i$th product in the $j$th country |
| $\gamma_{ij}$ | equals to one if the $i$th service supports the $j$th function and zero if not |
| $\tau_{ij}$ | equals to one if the $i$th service is available in the $j$th country and zero if not |
| $M$ | a large number |

$$\sum_{j=1}^{N_c} S_{ij} \leq \sum_{k=1}^{N_v} Q_{ik}, i = 1, ..., N_p \tag{9}$$

$$\sum_{i=1}^{N_s} \gamma_{ij} Y_i = 1, j = 1, ..., N_f \tag{10}$$

$$\sum_{l=1}^{N_p} S_{li} \leq \sum_{k=1}^{N_s} \gamma_{kj} \tau_{ki} Y_k M, i = 1, ..., N_c, , j = 1, ..., N_f \tag{11}$$

$$\sum_{i=1}^{N_p} Q_{ij} \leq X_j M, j = 1, ..., N_v \tag{12}$$

$$\sum_{i=1}^{N_p}\sum_{k=1}^{N_c} U_{ijk} \leq Z_j M, j = 1, ..., N_l \tag{13}$$

The constraint Eq. 7 ensures that the sales do not exceed the demand. The sales-delivery balance is enforced by the constraint Eq. 8. The sales-supplies balance is enforced by Eq. 9 stating that the products must be purchased from suppliers in order to sell them to the customers. Eq. 10 specifies that the services should be selected to satisfy all the required information processing functions. The constrains Eq. 11-13 ensure that suppliers, providers and services, respectively, should be included in the supply chain if they perform any activities (e.g., products are supplied by the given supplier). The constraint (12) ties together the physical and information flows by requiring that the products cannot be physically delivered if appropriate information services are not available.

## 4 Experimental

The experimental studies are conducted to demonstrate interdependencies between physical and information supply chain configuration decisions and to investigate impact of the weights $w_1$and $w_2$ on the configuration results. In order to check the first aspect, the supply chain configuration is performed without taking into account the information flows (EXP1). Technically, it means that $w_2$ is set to zero and the constraints (9) and (10) are ignored. The results of EXP1 are compared with an experiment (EXP2) where the physical and information flows are considered simultaneously. It is argued that the joint configuration has a significant impact on supply chain configuration if different suppliers or 3PL providers are selected.

### 4.1 Design of Experiments

A test supply chain configuration problem is set up for the experimental purposes. The dimensions of this supply chain are given by $N_p = 10, N_c = 30, N_v = 8, N_l = 3$ and $N_s = 10$. The services should provide seven functions. The services vary from highly specialized providing just one function to aggregators providing all functions. Some of the services are available in all countries while others are limited just to selected countries. The demand is randomly generated. However, the average demand for certain products is country dependent, and some supplier are able to produce these cheaper than others. The QoS characteristics are also randomly generated thought they are correlated with a number of functions the service provides (i.e., than more functions than worse performance). The model is solved using a commercially available mathematical programing software.

### 4.2 Results

The experiments EXP1 and EXP2 are carried out for the test supply chain. Fig. 4 shows all supply chain unit evaluated during the configuration and the units selected are shaded. $P' = R - C_1 - C_2 - C_3$ measures the supply chain performance in each experiments. It can be observed that different supply chain configurations are obtained in both experiments. In EXP1, the supply chain is able to serve all customers. The electronic information flows are provided by a combination of three web services and a 3PL provider, which provides more uniform delivery costs around the world, is selected. In EXP2, the information processing is performed by a single aggregator, which covers all but four countries. One additional supplier is present in the results of EXP1 compared to the results of EXP2. This supplier is able to supply all products but it specializes in the products most frequently order by the customers in the countries not served in EXP2. The supply chain performance is substantially affected by taking into account interdependencies between the physical and information flows for the given test supply chain.

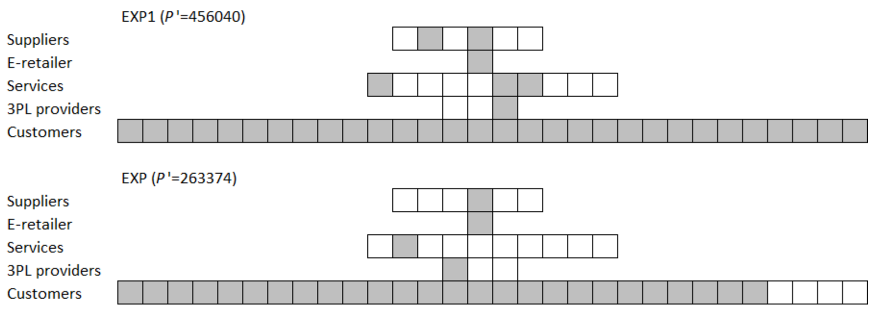


**Fig. 4.** The supply chain configurations obtained in experiments EXP1 and EXP2

The relative value of the weight factor $w_2$ characterizing the importance of QoS criteria in optimization is varied in order to evaluate sensitive of the results. The test supply chain used in the paper is quite insensitive to this factor. The QoS criteria had significant impact on the configuration results only for values $w_2$ exceeding $10^5$ (the cost related factors and quality related factors have vastly different scales).

## 5 Conclusion

A model for joint optimization of the physical and information flows in e-retailing supply chains has been elaborated. The model ensures that the physical supply chain units have appropriate information processing capabilities at their disposal. The importance of the joint optimization increases along with a growing number of electronic services available over the Internet.

The formulated optimization model defines relationships between the physical and information flows and takes into account QoS requirements for efficient information processing. Preliminary experimental results show that the information flows indeed affect selection of appropriate physical supply chain units.

However, the QoS requirements have minor impact of the supply chain configuration decisions for the test supply chain analyzed in the paper. An alternative approach to including QoS criteria directly in the objective model would specification of minimum quality requirements in the form of constraints. That would also alleviate the problem of selecting appropriate weights for multi-criteria optimization. The QoS characteristics also have impact on customer demand what could also be represented in the optimization model.

## References

1. Lambert, D.M., Cooper, M.C., Pagh, J.D.: Supply Chain Management: Implementation Issues and Research Opportunities, International Journal of Logistics Management, The, Vol. 9 Iss: 2, pp.1 - 20 (1998)
2. Chandra, C., Grabis, J.: Supply Chain Configuration: Concepts, Solutions, and Applications, Springer: New York (2008)
3. Melo, M.T., Nickel, S., Saldanha-da-Gama, F.: Facility location and supply chain management - A review, European Journal of Operational Research, vol. 196, no. 2, pp. 401-412 (2009)
4. Strunk, A.: QoS-aware service composition: A survey, Proceedings of the 8th IEEE European Conference on Web Services, ECOWS 2010, pp. 67-74 (2010)
5. Lee, H.L., Whang, S.: Information sharing in a supply chain. International Journal of Technology Management 20, 373-387 (2000)
6. Klein, R., Rai, A.: Interfirm strategic information flows in logistics supply chain relationships. MIS Quarterly: Management Information Systems 33, 735-762 (2009)
7. Swaminathan, J.M., Tayur, S.R.: Models for supply chains in e-business. Management Science 49, 1387-1406 (2003)
8. Hernandez, B., Jimenez, J.,Martin, M.J.: Key website factors in e-business strategy. International Journal of Information Management 29, 362-371 (2009)
9. Trappey, C.V., Trappey, A.J.C., Lin, G.Y.P., Liu, C.S., Lee, W.T.: Business and logistics hub integration to facilitate global supply chain linkage. Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture 221, 1221-1233 (2007)
10. Candido, G., Barata, J., Colombo, A.W., Jammes, F.: SOA in reconfigurable supply chains: A research roadmap. Engineering Applications of Artificial Intelligence 22, 939-949 (2009)
11. Chopra, S., Meindl, P.: Supply Chain Management: London, Prentice Hall (2012)

# Usability Evaluation of Method Handbook

Laura Heinermann[1], Dirk Stamer[1], Kurt Sandkuhl[1,2]

[1] Rostock University, Institute of Computer Science
Albert-Einstein-Str. 22, 18059 Rostock, Germany
[Laura.Heinermann, Dirk.Stamer, Kurt.Sandkuhl]@uni-rostock.de

[2] School of Engineering at Jönköping University,
P.O. Box 1026, 55111 Jönköping, Sweden
Kurt.Sandkuhl@jth.hj.se

**Abstract.** In enterprise modelling and information systems development, methods contribute to systematic work processes and to improving the quality of modelling results. Information Demand Analysis (IDA) is a method, which recently was developed for the purpose of optimizing the information flow in the field of information logistics. In order to contribute to improvement of the IDA method, the focus of this paper is to evaluate the usability of the IDA method handbook. For this purpose, an approach for usability evaluation of the handbook is proposed and applied. The main contributions are (1) an approach how to apply the concept of usability when evaluating a method handbook, (2) experiences from using this approach in a real world case, and (3) recommendation for improving the IDA method handbook with respect to usability.

**Keywords:** information demand modelling, method usability, information logistics, practice of modelling

## 1 Introduction

In enterprise modelling and information systems development, methods contribute to systematic work processes and to improving the quality of modelling results [1, 2] by providing procedural guidelines and notations for representing modelling results [3]. Information Demand Analysis (IDA) is a method, which was developed for the purpose of optimizing the information flow in the field of information logistics (see section 2). In an effort to provide this method, the creators of IDA released a handbook – the "IDA-User Guide" – that intends to provide guidelines and instructions for conducting information demand analysis. Although the IDA method has already been applied in several organisations, the usability of its handbook in fulfilling the intended purposes is still questioned. In utilising the handbook as applied to a real world situation, we examine the usability of the handbook from the view of a user and intend to contribute to its improvement. In the following we use "IDA-User Guide" equivalent to the term "Handbook".

The research on usability evaluation of user-oriented products belongs to the most productive areas of research in the field of information systems. This is most certainly inevitable since the growth of technology cannot be separated from the dynamic of the human factor, which demands the quality of use. Critical research is being conducted to examine the soundness of the usability concept as well as the suitability of the evaluation methods applied. However, these works deal more with the principles of usability for software products and less with the usability of an analysis method in information logistics, let alone in the area of information demand. The focus of this work is the application of usability to a method handbook. Other approaches in this area, like the SEQUAL approach [4], aim to evaluate the quality of methods on the overall.

Furthermore, we performed a literature analysis regarding IDA. Few results were found in the search of particular literature on the evaluation of IDA method. This situation is understandable because the method for information demand analysis is still in its preliminary stage. The only script that is found on this related issue is an undergraduate thesis written three years ago by a group of students at Växjö University [5]. The emphasis of their work focussed on the requirements and not on the usability of the method. Their object of research was the previous version of the method handbook.

These factors motivated us to begin an initial work on the usability of the handbook for information demand analysis. The main contributions of this paper are (1) an approach how to apply the concept of usability when evaluating a method handbook, (2) experiences from using this approach in a real world case, and (3) recommendation for improving the IDA method handbook with respect to usability.

The remaining part of the paper is structured as follows: section 1 introduces the background of this paper from the area of information demand analysis and information logistics; section 3 discusses related work with respect to usability and the approach used for usability evaluation of the handbook; section 4 contains the evaluation of the IDA method handbook; and section 5 draws conclusions.

## 2 Information Demand Analysis

Accurate and readily available information is essential in decision-making situations, problem solving and knowledge-intensive work. Recent studies show that information overload is perceived as a problem in industrial enterprises [6]. An example of a problem in relation to information overload is, in relation to different roles, to find the right information needed for a work task [7]. It is expected that an improved information supply would contribute significantly to saving time and most likely to improving productivity.

The research field information logistics addresses the before mentioned challenge in information supply. The main objective of information logistics is improved information provision and information flow. This is based on demands with respect to the content, the time of delivery, the location, the presentation and the quality of information. The scope can be a single person, a target group, a machine/facility or any kind of networked organisation. The research field of information logistics explores, develops, and implements concepts, methods, technologies, and solutions for the aforementioned purpose.

A core subject of demand-oriented information supply is how to capture the needs and preferences of a user in order to get a fairly complete picture of the demand in question. This requires an understanding of what information demand is and a method for capturing and analysing information demand. The term information demand (section 2.1) and the method for information demand analysis (section 2.2) will be briefly discussed in this section.

### 2.1 Information Demand

The understanding and definition of the term 'information demand' used in this paper is based on an investigation performed from 2005-2007 [8]. The main objective of this investigation was to identify the connection between information use and different work-related aspects, such as work processes, resources, and organisational structures, in order to achieve a better understanding of the term information demand. The investigation comprised 27 interviews with individuals from three different organisations, a sample which represented all levels of the investigated organisations, i.e. from top-level management via middle management down to production- and administrative personnel. Among the results of the investigation was a definition of information demand as well as a conceptualisation of this term. Information demand will be used throughout this paper with the following meaning: "*Information demand is the constantly changing need for relevant, current, accurate, reliable, and integrated information to support (business) activities, when ever and where ever it is needed.*"[8]

Information demand has a strong relation to the context in which such a demand exists [8]. The organisational role having the demand and for what task the information is demanded as well as the setting in which such tasks are performed are important aspects for understanding information demand. Thus, the concept of information demand context has been defined both conceptually and as the core of the method with respect to modelling, evaluating and analysing of information demand: "*An Information Demand Context is the formalised representation of information about the setting in which information demands exist and comprises the organisational role of the party having the demand, work tasks related, and any resources and informal information exchange channels available, to that role.*"[8]

Based on the above results, a method for analysing information demands and capturing information demand context has been developed as part of the research project infoFLOW-2 during 2010 – 2012 [8]. This method is documented in a method handbook, which describes the work procedures to be performed, the notation for documenting information demand contexts, and the concepts and aspects to be taken into account during analysis. The different phases of this method will be introduced in the next section.

### 2.2 Method for Information Demand Analysis

In Figure 1 an overview of a framework for achieving such an understanding is presented, which will be described in this section. Since context is considered central to information demand analysis, method support for modelling such contexts is at the core of the framework as highlighted in the framework below. However, in order to be able to perform any meaningful context modelling, a clear scope is needed. Consequently, the information demand analysis starts with scoping activities.

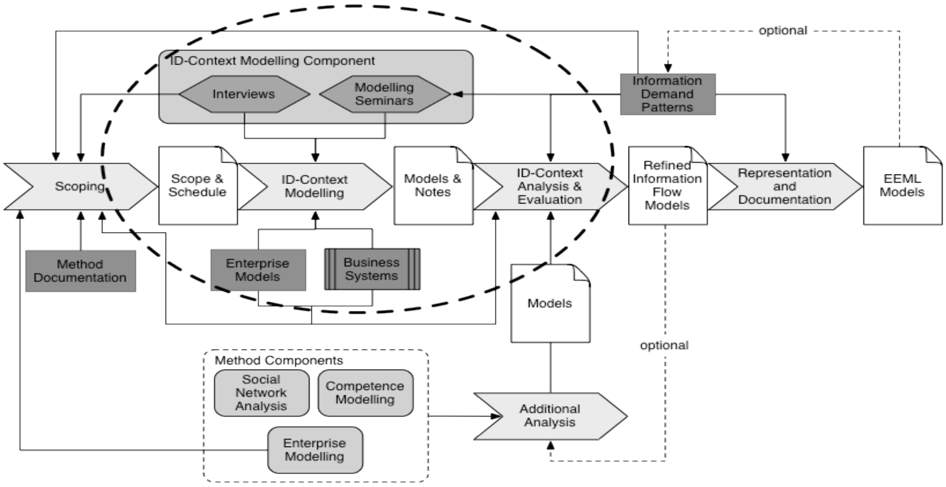


**Figure 1: An overview of the process of analysing information demand**

*Scoping* is the process of defining the area of analysis and is done with the purpose of selecting parts of an organisation to be subjected to analysis. This phase also includes the identification of the roles (individuals) relevant for the continued information demand analysis. Scoping also sets the scene for identification and understanding of the organisation's problems, goals, intentions, and expectations to motivate them to engage in the information demand analysis.

*Information Demand Context Modelling* is mainly performed through participative activities such as joint modelling seminars where the participants themselves are involved in the actual manufacturing of different models. This process is usually supported and facilitated by a method expert who could be an internal or external person. As illustrated in Figure 1 the conceptual focus is in this phase of information demand within a defined scope. The key to context modelling is to identify the interrelationship between roles, tasks, resources and information. No regard is given to the sequence of activities, resource availability, etc.

*Information Demand Context Analysis and Evaluation:* Once the necessary knowledge about the information demand contexts is obtained, it can be used for a number of different purposes. One purpose is evaluation where different aspects of information demand can be evaluated in relation to roles, tasks, resources and information. It is also suitable to address the results from the modelling session with respect to motivation and purposes expressed during the scoping activities. Focusing on information demand contexts provides only an initial view of information demand without any consideration given to such aspects as individual competence, organisational expectations and requirements in terms of goals, processes etc. Depending on the intentions behind the analysis further activities might be required. The method provides a number of method components supporting such activities. Since the main focus of the method presented here is on information demand, it utilises existing procedures and notations for such additional aspects rather than defining new ones. Consequently, if the method user wishes to investigate such additional aspects of information demand, he or she can do this by using subsets of the other methods, notations and languages.

## 3 Usability Evaluation Approach

In this section the approach taken to arrive at a suitable definition of the term usability in the context of the IDA User Guide is described. In addition, the proposed criteria for usability evaluation are outlined. Therefore, in section 3.1, Usability in the context of evaluating a handbook is defined. In section 3.2 the given definition is evolved into Evaluation Criteria.

### 3.1 Approach to defining Usability

The daily use of the term usability is often found reducing the meaning of the word simply to be the ease that one has in using a product. However, the dimensions of ease itself are so diverse that it is quite difficult to measure. Usability practitioners in the area of Human Computer Interaction have dealt with this situation since the early 1970s. During the course of usability research in the field of software engineering back in 1994, usability practitioners had complained that work on usability was too little, too late. Nigel Bevan believed that the reason for this complaint lay in the narrow view toward the term usability. Usability was often considered only in the term of ease of use while the term of ability to fulfil its intended purpose was neglected [9, 10].
Owing to the experiences of fellow usability researchers from the field of information systems, we are well aware of these important aspects of usability and are, therefore, taking these aforementioned aspects into consideration in our initial work on evaluating the usability of the IDA User Guide. Realizing the different contexts between that of software engineering and that of information demand, our attempt to determine a suitable definition of usability for this work is directed in approaching the interpretation of usability based on the context of use.
This approach is believed inevitable due to the absence of previous research on defining usability in the practical use of method to the best of our knowledge. Here we observe a definition of usability from a usability practitioner and anticipate how usability is interpreted by the International Organization for Standardization.
Jakob Nielsen defined usability by associating it with the following five attributes:

- *Learnability*: *The system should be easy to learn so that the user can rapidly start getting some work done with the system.*
- *Efficiency*: *The system should be efficient to use, so that once the user has learned the system, a high level of productivity is possible.*
- *Memorability: The system should be easy to remember, so that the casual user is able to return to the system after some period of not having used it, without having to learn everything all over again.*

- *Errors: The system should have a low error rate…catastrophic errors must not occur.*
- *Satisfaction: The system should be pleasant to use, so that users are subjectively satisfied when using it* [11].

The International Organization for Standardization issued ISO 9216 from 2001 to 2003 as the series of standards addressing software quality and replaced it by ISO/IEC 25010 in 2011. To understand the implicit value of usability beyond the applied categorization, we will consider both releases of the standard for our observation.

ISO 9216 constituted a software quality model which divides software quality into six general categories of characteristics: functionality, reliability, usability, effectiveness, maintainability and portability. Simultaneously the definition of usability is given as "the capability of the software product to be understood, learned and liked by the user, when used under specified conditions" [12].

In the final draft ISO/IEC FDIS 25010 a quality-in-use model constituted five characteristics whereas three of them, which tend to be relevant for our purposes are presented here with their definitions as follows:

- *Effectiveness*: *accuracy and completeness with which users achieve specified goals.*
- *Efficiency: resources expended in relation to the accuracy and completeness with which users achieve goals.*
- *Satisfaction: degree to which user needs are satisfied when a product or system is used in a specified context of use.*
- Freedom from risk
- Context coverage.

*Effectiveness* and *Efficiency* are not further divided into sub characteristics as *Satisfaction* is. However, we believe among the given sub characteristics of *Satisfaction*, *Usefulness* is the only one applicable to our context. It is interpreted as "degree to which a user is satisfied with their perceived achievement of pragmatic goals, including the results of use and the consequences of use" [13].

The concept of *Usability* was still recognized in the development of the standard and the characteristic *Usability* is categorized as one of the characteristics of the second quality model - product quality model. It is stated as "degree to which a product or system can be used by specified users to achieve specified goals with effectiveness, efficiency, and satisfaction in a specified context of use" with the sub characteristics: appropriateness recognisability, learnability, operability, user error protection and user interface aesthetics [13].

In both releases, Usability is to be associated with the perspective of *learnability, specified user, specified goal* and specified conditions or *specified context of use*. In addition, the presented quality of use model indicates the human factor approach, in the sense that the system is able to fulfil the needs of the user. Both aspects seem to be maintained as the backbone of the definition of usability in the past two decades.

These aspects of usability are now to be assessed with respect to their compliance with the nature of the handbook and the context of information demand analysis to which the handbook refers.

1. *Learnability: the system should be easy to learn so that the user can rapidly start getting some work done with the system.*

As stated in the outline of the handbook, section *Background,* "a method of information demand analysis therefore has to include all relevant aspects of enterprises, such as work tasks, organisational structures, processes, information sources and information receivers. This handbook describes exactly such a method in terms of content, structure, and use."[8] Consequently a holistic understanding of the content, structure, and use of the method is very important. Moreover, learning a concept, which is intangible unlike a visualized program, requires a very clear explanation and understandable description from the provider of the concept. Taking these matters into consideration, the comprehensibility of the handbook is an essential aspect for the usability of the handbook. Therefore, the term learnability is going to be replaced with the term comprehensibility and its interpretation is as follows: the handbook should be easy to understand so that the user can quickly perceive the overview of the method and finish the preparation for the analysis in a timely manner.

2. *Specified user.*

In the section *Purpose of the Handbook*, there are three different types of readers the handbook is intended for. This differentiation is nonetheless not a specifying classification based on some competency requirements in the area of information demand analysis. It is rather to serve as information about the particular intentions the handbook is able to facilitate. Since the IDA method itself is "user-intensive", the aspect of user should be adopted in the concept of usability of the handbook. However, the term "specified" is not proper in this case due to the abovementioned view and will be dropped therefore.

3. *Specified goal.*

Accommodating different types of user requires the handbook to be able to accommodate their different goals, which can be found within the stated purpose of the handbook. The specified goal refers to that of the user in using the handbook according to its stated purpose. Based on this view, this aspect of specified goal is relevant for the concept of usability of this handbook.

4. *Specified context of use.*

The Handbook described in the section *The Purpose of a Method for Information Demand Analysis* the context in which information demand analysis needs to be done. Hereby the context of use for the handbook is already specified and a specified context of use is not relevant for the usability of the handbook.

5. *Effectiveness* in the sense of *accuracy and completeness with which users achieve specified goals.*

The specific purpose of this handbook is to facilitate the information demand analysis. Consequently, the focus of the handbook is placed on the pragmatic aspects of information demand. Owing to this purpose, accuracy and completeness of the explanations, directions and references in the handbook is very important. Although it is somewhat subjective, it is absolutely relevant for the evaluation of quality of use.

6. *Efficiency* as in *resources expended in relation to the accuracy and completeness with which users achieve goals.*

The amount of resources required for the implementation of the handbook depends on various factors. Amongst them are the user profile, the perceived problem, and the context of the use. Even if the resources used can be recorded and the procedures repeated, there are still more variable parameters than fixed ones in the implementation of the handbook. A further specialization in the area of implementation would probably be necessary before integrating this aspect into the usability of the handbook. Hence, we find this aspect irrelevant for the consideration of current usability of the handbook.

7. *Usefulness* with respect to the *degree to which a user is satisfied with their perceived achievement of pragmatic goals, including the results of use and the consequences of use*.

The practical steps found in the fourth section of the handbook are categorized into prerequisites, activities and expected results. The expectation and pragmatic goals of user during each of these steps can hereby be evaluated regarding to the usefulness of the prerequisites, activities and expected results of each phase. Therefore usefulness is a relevant attribute for the concept of usability.

Under the consideration of the abovementioned analysis, we define the term usability of the handbook for Information Demand Analysis as follows:

*The quality of the handbook to be used with comprehensibility to achieve the specified goal with accuracy, completeness and usefulness for the user.*

### 3.2 Evaluation Criteria

By using the building elements of the before given definition we are going to propose criteria for evaluating the usability of the handbook.

*Attributes*

Among the said elements there are four attributes, which are directly related to the quality of use of the handbook. These attributes are as follows:

- Comprehensibility: the handbook should be easy to understand so that the user can quickly perceive the overview of the method and finish the preparation for the analysis in a timely manner.
- Accuracy: the essential point is communicated in the right section and each section presents its contents according to the correct priorities.
- Completeness: all the essential features are available. The pertained explanatory elements are likewise obtainable.
- Usefulness: the conceptual as well as practical descriptions available in the handbook do not only meet the user's fragmental expectations, but also accommodate the user's pragmatic goals.

*Selected Criteria*

In evaluating these attributes, certain criteria or parameters for the analysis are needed. There are certainly many facets pertaining to those attributes, but for this work on this particular handbook for information demand method, we limit our list of criteria. Based on our apprehension of the handbook, we assign the abovementioned attributes into the following aspects:

- Wording. It is one significant factor of comprehensibility in an expositional text such as a handbook.

- Contextual and visual illustration of the concept. Similar to wording, the contextual and visual illustration is also an influential factor of comprehensibility.
- The stated heading of each section and the main idea that emerges in the respective section are coherent to each other. This can serve as an aspect to evaluate the accuracy of the handbook.
- No parts of ideas, important notes or description pertaining to the existing ones are missing. This is an aspect of completeness.

The results of use after each phase meet the user's expectation and her or his goal. This is an aspect of usefulness.
We proposed here several attributes and selected criteria to evaluate the IDA User Guide. Due to the explorative advance we did not develop a measurement scale. This would be a further step towards applying usability concepts to a handbook.

# 4 Case Study and Evaluation of IDA User Guide

Most research projects conducted in university are financed by external funders, who make their decision to grant the fund or not based on the competitiveness of the submitted proposal for the respective project. The process of writing this proposal involves a great deal of information to be processed and procedures to be conducted by various areas of responsibility within the university. The information demand within this information-intensive process of proposal writing is to be identified with the method described in the handbook for information demand analysis. Our customer is the head of the proposal writing team. Interviews are conducted with the head of the proposal writing team, the financial administrator of the institution and the content manager. Furthermore, the role of method user is defined as the method-applying person who does the interviews as well. Problems that arose during the implementation of the handbook are recorded along with the evaluation of the usability criteria.
The following sections are structured according to the process of analysing information demand. Within the sections stated activities (proposed by IDA user guide), performed activities and problems, which occurred, are listed.

## 4.1 Scoping Phase

*Stated activities*

- Understanding and identifying the perceived problem.
- Defining the problematic area and the actors within it.

*Activities performed*

- Orientation interview with the head of the team about the process of proposal writing to get a basic understanding about the process.
- Prepared a mission statement with input from the interview.
- Review of the mission statement by the head of the team.

*Problems*

- Difficulties in describing the mission statement, especially the description of expected results and effects of the assignment by a misunderstanding.

### 4.2 Information Demand Context Modelling Phase

*Stated Activities*

- Short semi-structured interviews with each of the selected individuals in order to get an understanding of their expectations, goals and problems.
- Modelling seminars to collect the needed information to perform further work with the aim to get the participants to identify the different tasks performed within the scope and how this relates to information which is needed or used, grouped by role. Typically, this is done by writing down different tasks and information objects on pieces of paper and attaching them to large paper or plastic sheets and then connecting them to each other. The reasoning and motivation behind this approach is that it allows for easy restructuring and changes as the session progresses. An additional benefit of such a non-technical approach is that is easier to get the participants to contribute to the emerging model than it would be if they all were seated around a table watching the model emerge on the facilitator's computer screen.

*Activities performed*

- The first inquiry session was conducted with the head of the team. This lasted 15 minutes longer than the estimated duration of 60 minutes.
- For the session with the financial administrator, a sample of the model adopting the illustrated sample case was prepared with the same modelling tool that will be used for the interview. This sample was then shown to the informant at the beginning of the session. The reason for this was to provide the informant a better picture of what the result of the session was supposed to be. This sample was in German as to accommodate the preferable language of the informant for the session. The duration of this session was exactly as estimated.
- For the session with the content manager, the sample of the model was again shown to the informant at the beginning of the session. In spite of the use of the German sample, the language for the session was English. This session also finished within the estimated time.

*Problems*

- During the first session there were problems to classify the statements and responsibilities given by the informants.
- Regarding the respect to all three sessions: The first stated purpose was not fulfilled. The method user does not understand the character and personality of the individuals. Method user had difficulties identifying the information within the information situation.

### 4.3 Analysis and Documentation of Information Demand Phase

*Stated Activities*

- Transcription of the initial models into well-defined notation.
- During the transcribing of the models correlations between information flow in the current case and information flow in the previous cases can sometimes be identified. If an identified situation within the models is considered similar to already identified and solved situations from other cases it is reasonable to assume that the solutions applicable in that situation are also applicable in the current one. If no corresponding pattern can be found in the pattern collection despite the fact that it has been identified in a number of different cases, it is reasonable to assume that a candidate for a new and valid pattern has been identified.

*Activities performed*

- The models out of the first interview and modelling sessions were still lacking in connectors to most of the relationships. These three models were then analysed using the information from the documented interview.
- Constructing one new overall model out of all three, bilingually.
- To fill the gaps and get a confirmation for this developed model, a second meeting with each informant was then scheduled, which leads to their approval to the model. Furthermore, a similar information demand pattern addressing the process of proposal writing was found in the collection of patterns. Based on this pattern, the approved context graph was then digitalized.

*Problems*

- During transcription some flows were unclear because of the different treatments to the type of information. This should be described more clearly in the handbook.

## 5 Conclusions and Future Work

As shown in section 3 the concept of usability was applied, when evaluating a method handbook. The starting points were definitions of usability given by practitioners and the International Organization for Standardization. We extracted and analysed the relevant terms from the definitions. As a main contribution an adapted definition of usability for the Information Demand Analysis User Guide is proposed as the quality of the handbook to be used with comprehensibility to achieve the specified goal with accuracy, completeness and usefulness for the user. Furthermore we proposed evaluation criteria to measure the given attributes in the definition. Due to the explorative advance we did not develop a measurement scale. This will be a further step towards applying usability concepts to a handbook.

In section 4 we applied this approach in a real world case. This quality of use was evaluated by implementing the handbook in an information-intensive process within the process of proposal writing in the university. This leads to helpful insights and practical guidance in the internal process. Through its implementation, the information situation within the process of proposal writing became clearer, and suggestion for solving the problem related to information demand could be given. By using the definition of usability approached in this work, a continuation of the evaluation process in the future is possible.

Although the method user handbook follows in most aspects the usability requirements, but as another contribution some recommendations for improving the IDA method handbook with respect to usability could be given:

- Important notes on the procedures of asking questions during the interview.
- Additional explanation of the contextual use of identifying and differentiating.

This work attempted to apply one possible concept – usability – to improve the quality of a method handbook. As a further approach knowledge from cognitive psychology could be integrated as well.

The results of this usability evaluation may contribute as well to the acceptance of the Information Demand Analysis method in the networked organization and to the improvement of its quality.

## References

1. Brinkkemper, S.: Method engineering: engineering of information systems development methods and tools. Information and Software Technology. 38, 275–280 (1996).
2. Mirbel, I., Ralyté, J.: Situational method engineering: combining assembly-based and roadmap-driven approaches. Requirements Engineering 11. 58–78 (2005).
3. Goldkuhl, G., Lind, M., Seigerroth, U.: Method integration: the need for a learning perspective. IEE Proc., Softw. 145, 113–118 (1998).
4. Krogstie, J.: Model-Based Development and Evolution of Information Systems. Springer Verlag (2012).
5. Nyberg, C., Wass, S.: Does IDA meet the requirements? (2009).
6. Öhgren, A., Sandkuhl, K.: Information Overload in Industrial Enterprises - Results of an Empirical Investigation. Presented at the Proceedings ECIME 2008, London (2008).
7. Lundqvist, M., Sandkuhl, K., Seigerroth, U.: Modelling Information Demand in an Enterprise Context: Method, Notation, and Lessons Learned. International Journal of Information System Modelling and Design (IJSMD). 2, 75–94 (2011).
8. Lundqvist, M.: Information Demand and Use: Improving Information Flow within Small-scale Business Contexts, (2007).
9. Shackel, B., Richardson, S.J.: Human Factors for Informatics Usability. Cambridge University Press (1991).
10. Bevan, N.: Measuring usability as quality of use. Software Quality Journal. 4, 115–150 (1995).
11. Nielsen, J.: Usability Engineering. Academic Press (1993).
12. Abran, A., Khelifi, A., Suryn, W.: Usability Meanings and Interpretations in ISO Standards. Software Quality Journal. 11, 325–338 (2003).
13. ISO/IEC FDIS 25010: Systems and Software Engineering - Systems and Software Quality Requirements and Evaluation - Systems and Software Quality Models. (2010).

# TiEE – The Telemedical ILOG Event Engine: Optimization of Information Supply in Telemedicine

Sven Meister[1,*], Sven Schafer[1], Valentin Stahlmann[1]

[1]Fraunhofer Institute for Software and Systems Engineering, Dortmund, Germany
{sven.meister,sven.schafer,valentin.stahlmann}@isst.fraunhofer.de

**Abstract.** Ongoing problems in healthcare supply make it imperative to establish telemedicine as a feasible concept to ensure the quality of medical treatments and reduce costs. With TiEE, the Telemedical ILOG Event Engine, Fraunhofer ISST fosters the research to process streams of vital signs in telemedical scenarios using Complex Event Processing (CEP) in an Information Logistics (ILOG) manner. ILOG means to reduce the amount of information by preprocessing existing data and to route high-grade decision at the right time to the right person. As basic building blocks we developed the concepts of telemedical events (TE), Telemedical ILOG Listener (TIL) and TIL-Profiles on top of the event processing engine Esper [20–23]. TiEE analyzes the incoming patient specific streams of telemedical events and tries to detect relevant trend patterns. In a second step these got aggregated to higher level decisions.

**Keywords:** CEP, Telemedicine, Decision Support, Trend Recognition

## 1 Introduction

The continuous measuring of vital signs results in an unmanageable amount of data and furthermore information that could be deduced. Physicians therefore complain about the problem of information overload, which means that the information processing requirements exceed the information processing abilities [12]. Investigating solutions for solving the problem of information overload in telemedical scenarios at first means to reduce the amount (quantity) of data. Second, the time for acquiring information has to be reduced by distributing the right information, at the right time to the right place.

To cope with the problem of information overload in telemedical scenarios of continuous monitoring of vital signs we investigated the Telemedical ILOG Event Engine (TiEE). TiEE is based upon scientific outcomes of the following two research areas: Information Logistics (ILOG) and Complex Event Processing (CEP). The former is always related to the metaphor of transporting the right information at the right time to the right place [6, 14]. The latter enables one to process so called events in real-time by filtering, aggregating and transforming them into more complex events. Why is CEP an appropriate technology for ILOG processing in telemedicine? Every vital sign monitored by a telemedical application is some kind of event. Such an event could be related to additional information like the time of generation or the value of the vital sign itself. By analyzing all data and information related to such an event, under consideration of the history and order, we are able to reduce the amount of irrelevant vital signs. Furthermore we can aggregate events to higher order decisions, so called complex events.

Thus, the purpose of the paper is to show how information demand can be mapped on concepts of event processing to optimize information supply in telemedical scenarios by reducing the amount of information overload. In the following we will give an overview of the state of the art in ILOG and CEP. Furthermore we'll discuss the basic concepts we investigated to implement TiEE, which are telemedical events, Telemedical ILOG Listener (TIL) and Telemedical ILOG Listener Profiles (TIL-Profile). At the end we'll show first evaluation results based upon two different use cases.

## 2 Related Work

TiEE is focused on the two research areas complex event processing and information logistics to foster a fast, on-time processing of telemedical values. Basic definitions and concepts of CEP were developed and defined by Luckham, Chandy and Bates [2, 3, 18]. Citing them, an event is „an object that is a record of an activity in a system". A detailed overview about open questions and the current state of research in CEP is discussed in the Dagstuhl Seminar on „Event Processing" in 2010 [4]. In [17, 25] Lowe and Weber present ongoing work on applying CEP to health data using the event processing engine Esper in the context of the Stride project „Stanford Translational Research Integrated Database Environment". One basic open question is how to cope with the problem of heterogeneity of medical data to achieve an overall processing.

The need for information logistics is caused by an increasing number of situations of information overload. According to Wilson [27, 28] information overload expresses „that the flow of information […] is greater than can be managed effectively". ILOG is viewed as a research area to deliver the right information, in the right format at the right time to the right place and is partially used for information filtering or with context-models to optimize communication in the healthcare domain [7, 15, 26]. A broad overview of the state-of-the art research in information logistics is given by Haftor et al. [9] by analyzing 102 scientific publications. The link between CEP and ILOG is given by Chandy [4] by mentioning that „Disseminating and distributing is also about getting the right information to the right consumers at the right time.".

Alternative approaches to the usage of CEP and ILOG could be found in the research area of data streams. They could be used for real-time processing of data like shown within STREAM [1]. Geesen [8] gives an introduction how data streams could be used to process high frequent data in the area of Ambient Assisted Living. All approaches cope with the problem that data or information isn't standardized, developed concepts are not that modular and they don't take ILOG into account. Therefore many papers in both research areas give an outlook on using events and event based processing for real time data optimization.

## 3 TiEE – The Telemedical ILOG Event Engine

The reduction of information overload in telemedical scenarios requires to process telemedical information in the sense of aggregation, filtering as well as analysis of causal and temporal relationships. There are three basic requirements for the processing of vital signs in telemedical scenarios one should take into account [21]:

- Sensors for measuring vital signs act in a highly distributed manner. Every type, e.g. blood pressure or blood sugar concentration, is a result of a single sensor or telemedical application. The prospective solution should be able to process telemedical values from different sources to achieve an overall monitoring and decision making. This requires an overall description of a telemedical value, in sense of a vital sign, as well as methods to modular process different types of telemedical values depending on the actual medical situation of a patient.
- Monitoring of vital signs in telemedical scenarios produces a high amount of data which has to be processed in real-time. Not every single vital sign represents an important medical situation. The relevance depends upon the temporal ordering as well as the coincidence of different types of vital signs. So, a solution has to aggregate a set of those to higher order decisions.
- The delivery of those decisions mentioned above should be done in an intelligent way. So, a derived decision should be transported at the right time to the right place, e.g. prior to a physician a notification should be emitted to a family member.

With TiEE, the Telemedical ILOG Event Engine, Fraunhofer ISST investigated methods for event based processing of vital signs in telemedical scenarios considering the requirements mentioned above. We'll start by giving a broad introduction to the information logistics processing within TiEE. Afterwards we'll describe the three basic concepts telemedical event, TIL and TIL-Profile in more detail.

### 3.1 Information Logistics Processing

We started our discussion about information overload in telemedical scenarios by stating that it is not possible for a physician to deduce relevant information out of a stream of thousands of vital signs. Thus, realizing a support for clinical decision making one has to reduce the amount of data and deduce higher level, relevant information. Delivered relevant information should fulfill a given information demand, thus a physician needs a possibility to express its demand.

With TiEE we investigated a graphical demand modeling approach (Demand Modeling Language) upon the Clinical Algorithm Standard (CAS) [24] which is well known in medicine. Within CAS we have five elements (see **Fig. 1**): a start node (oval), a condition (rhombus), an activity (rectangle), a terminal node (circle) and a connector (arc). A condition is the evaluation of a given expression, like pulse higher than 150. The logical AND concatenation is modeled by writing conditions from left to right and the OR concatenation is realized by writing conditions down. The result of the evaluation of a set of conditions could be the terminal node, thus nothing will happen, or a red or green colored activity. An activity symbols the necessity to inform about a relevant situation. Relevance is divided into critical (red) and uncritical (green) situations.

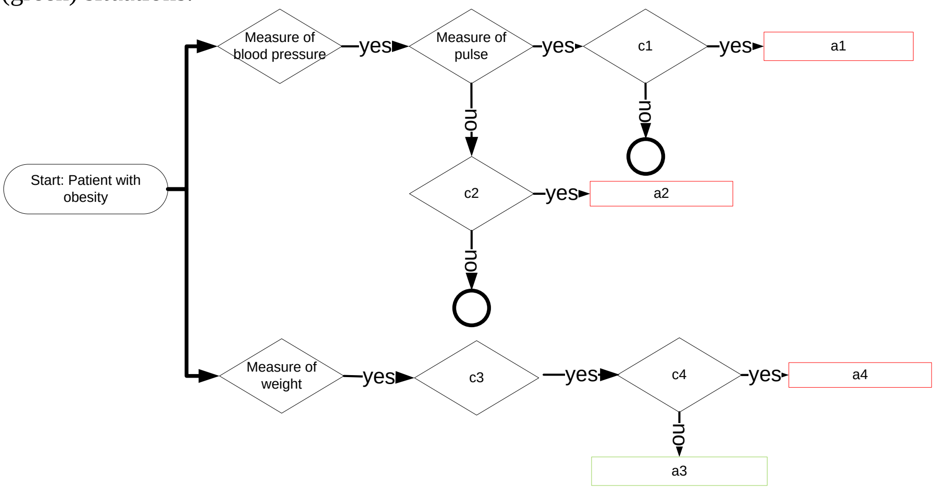


**Fig. 1.** Demand modeling using the Clinical Algorithm Standard.

Regarding the figure shown above we have a patient coping with obesity. He has to measure weight, pulse and blood pressure. Condition *c2* is only valid in case that the pulse isn't measured otherwise condition *c1* is valid which would lead to action *a1* or a terminal node.

The graphical model has to be transformed into processable data structures, thus we defined a set of structures upon the Extended Backus-Naur Form (EBNF) in such a way that the Demand Modeling Languages $DML \subset EPL$ is a subset of an Event Processing Language and further can be mapped to the extending concepts like TILs and TIL-Profiles. Below we sketched up the EBNF specification of the formalize demand and afterwards in **Table 1** examples according to **Fig. 1**.

```
demand    = "DEMAND" (condition | condition "→" condition)    CRLF ;
             "ACTION" activity "ELSE" activity ";" CRLF ;
```

**Table 1.** Expressions for condition *c1*, action *a1* and the demand *d1* according to the example shown in **Fig. 1**.

| **ID** | **Expression** |
|---|---|
| c1 | bloodpressure.BP1 = INCREASING AFTER pulse.BP1 = INCREASING |
| a1 | INFORMATION Critical increase of blood pressure and pulse. Please stay in contact with the patient.; |

| | | |
|---|---|---|
| | RELEVANCE | 80; |
| | CTITICALITY | 90; |
| | RECIPIENT | ID1 \| Servicemitarbeiter \| JMS \| service_in \| DIRECT; |
| D1 | **DEMAND** c1 **ACTION** a1 **ELSE** NIL; | |

Now, the above expressed and formalized demand has to be mapped on the concepts of TILs and TIL-Profiles bearing the Event Processing Language used within TiEE in mind. At first, an incoming telemedical event will be processed by a patient-specific TIL-Profile, like described in chapter 3.4. Afterwards a set of TILs is trying to detect characteristic patterns within the stream of telemedical events, so called Complex Trend Pattern Events (CTPE). A CTPE is an abstraction of a set of telemedical events (see also **Fig. 2**) and characterizes the progression of the measured values. Based upon the research of Charbon et al. [5, 10] we distinguish five basic types of trend pattern: slope, slope reverse, saltus up and saltus down steady.

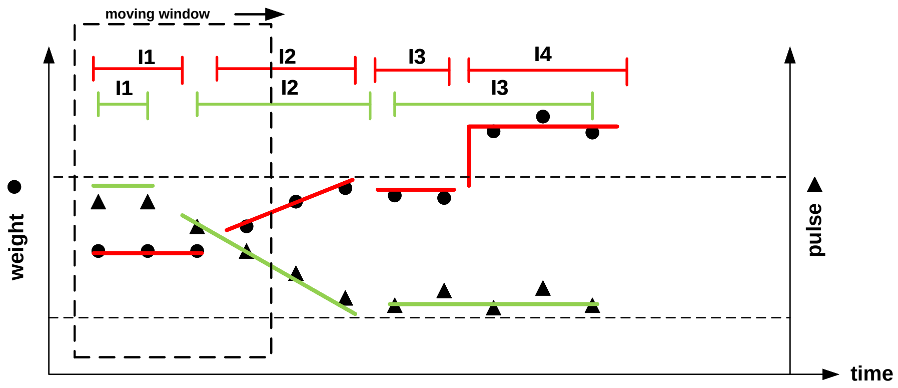


**Fig. 2.** Abstraction of incoming telemedical events by building intervals in Terms of CTPEs.

Thus, we derive an abstraction, the pattern, from a set of underlying measurements to reduce the amount of data and cope with the problem of information overload. The derived pattern is described using different types of parameters, e.g. the statistical spread or the amount of increase/decrease. As a basic feature of the TIL concept one can use any kind of algorithm for trend calculation as long as it fulfills the formal definitions of a TIL. CUSUM (cumulative sum) or ARIMA (autoregressive integrated moving average) based approaches are examples for processing time related data.

In summary, every TIL derives CTPEs for a specific telemedical event, i.e. a type of vital signs, and forwards them up to the referring TIL-Profile. Now, this TIL-Profile has to detect higher order, demand fulfilling patterns within the set of forwarded CTPEs. By using the formalized demand the processing within a TIL-Profile is organized as follows:

- Trends of same underlying types of vital signs: The repeated increase, decrease etc. of a set of vital signs could be abstracted to a single trend pattern.

- Trends of different underlying types of vital signs: It is obvious that there is a relation between weight and blood pressure in cases of cardiac decompensation. A TIL-profile has to detect the increase of both during a given time window and derive a new abstraction, emitting a new trend pattern.

Upon rules registered in the TIL-Profile, information logistics decisions are made to generate and send relevant information to a person.

### 3.2 Telemedical event and HL7 Telemedical Event Format

Within telemedical scenarios you'll find a lot of different sensors from various manufacturers to measure vital signs. The overall processing of vital signs using TiEE requires some concept to standardize the input. While TiEE is based upon the idea of complex event processing, every vital sign should be interpreted as an event. Therefore we defined the term telemedical event as a *measurement of a telemedical value and an instance of a telemedical event type, formatted in the HL7 Telemedical Event Format* [20]. The HL7 Telemedical Event Format is a message format we investigated to achieve an interoperable transportation of Telemedical Events. The refinement of this format is done by combining elements of the HL7 standard with such from the IEEE 11073 standards. HL7 is a widespread international standard for data exchange in the healthcare sector [11]. All HL7 V3 data types are based upon the HL7 Reference Information Model (RIM). In turn IEEE 11073 is a family of standards to harmonize the output of sensors using the IEEE 11073 Domain Information Modell (DIM) [13]. Using both standards we modeled a format that takes all attributes for complex event processing and ILOG processing of vital signs into account.

Formally a telemedical event is an n-tuple $e_i^T := (e_i, HL7_{Trans})$ where:

- $E^T := \bigcup_{i=1}^{\infty} e_i^T$ is the set of all telemedical events.
- $e_i \in E$ is an event based upon a telemedical event type $et^T \in ET^T$ .
- $HL7_{Trans}: E \rightarrow E^T$ is a transformation into the HL7 Telemedical Event Format.

Two telemedical events are identical $e_i^T \equiv e_j^T$ if and only if $e_i = e_j$ and $HL7_{Trans}(e_i) = HL7_{Trans}(e_j)$. $HL7_{Trans}$ is a function to transform a given event into the HL7 Telemedical Event Format.

### 3.3 Telemedical ILOG Listener (TIL)

A TIL-Profile realizes a patient specific filtering of the incoming telemedical events. The second step of filtering the high amount of events is done within the concept of a TIL. Related to CEP a TIL is some kind of Event Processing Agent, specialized for processing one type of telemedical values e.g. blood pressure events [23].

Besides the operation of filtering, a TIL encapsulates methods to detect patterns of interests in the stream of incoming events. Therefore different types of rules could be instantiated. Thus, a TIL is a modular piece of concept to encapsulate algorithms which are highly specialized to process one type of vital sign but is not specialized to a patient. That enables an easy reuse.

Formally a Telemedical ILOG Listener is defined as an n-tuple as follows $til := (et_{in}, ET_{out}, f_{in}, VL)$:

- $TIL := \bigcup_{n=1}^{\infty} til_n$ is the set of all TILs.
- $et_{in}$: The event type on which all instances $inst(et_{in}) = e_{in}$ are based upon. Initially this is the telemedical event type $et_{in} := et^T$.
- $ET_{out}$: Analogous to the definition of $et_{in}$, $ET_{out}$ is a set of event types $|ET_{out}| \geq 1$ which are permitted to be emitted as output.
- $f_{in}$: The filter function is based upon a boolean function $f_{in} \rightarrow E^T : (true, false)$. Given to functions $f_{in_1}$ and $f_{in_2}$ it is imperative that $f_{in_1} \equiv f_{in_2}$ are identical if and only if $vitaltype\ (f_{in_1}) = vitaltype\ (f_{in_2})$ that is, both functions relate to the same type of vital sign. Two mutually different functions $f_{in_1} \sqcap f_{in_2} = \emptyset$ are disjoint.
- VL: Every TIL consist of processing logic $VL$ which is a set of rules in terms of:

$$VL := \left\{ R \mid R = \left\{ \begin{matrix} f: E^T \rightarrow (true, false), \quad Filter \\ p: E \rightarrow E, \quad Pattern \\ t: E \rightarrow E, Transformation \end{matrix} \right. \right\}$$

Two TILs are identical $TIL_1 \equiv TIL_2$ if and only if $et_{in_1} \equiv et_{in_2}$, $ET_{out_1} \equiv ET_{out_2}$, $f_{in_1} \equiv f_{in_2}$ and $VL_1 \equiv VL_2$ .

### 3.4 Telemedical ILOG Listener –Profile (TIL-Profile)

Every medical situation represents an individual moment in lifetime. Thus, TiEE has to offer functionalities for a patient specific processing of incoming telemedical events, which very fast can be adapted to a new situation. Therefore we investigated the term TIL-Profile. A TIL-Profile realizes a patient specific filtering of telemedical events thus it reduces the amount of data. Afterwards the event is processed, depending on the type of vital sign, by one of the TIL's (see section 3.3) connected to this profile. For every type of vital sign one has to register one TIL. In the following the output of the TIL's is processed within the TIL-Profile by additional filtering, pattern detection and transformation into higher order decisions. That means that a TIL-Profile correlates the progression of different types of vital signs, e.g. blood pressure and weight, detects a medical situation of relevance and derives a higher order medical decision.

Formally a Telemedical ILOG Listener Profile is defined as an n-tuple as follows $til_{Profil} := (et_{in}, ET_{out}, f_{in}, TIL, VL)$:

- $TIL_{Profil} := \bigcup_{n=1}^{\infty} til_{Profil_n}$ set of all TIL-Profiles.

- $et_{in}$: The event type on which all instances $inst(et_{in}) = e_{in}$ are based upon. Initially this is the telemedical event type $et_{in} := et^{T}$.
- $ET_{out}$: Analogous to the definition of $et_{in}$, $ET_{out}$ is a set of event types $|ET_{out}| \geq 1$ which are permitted to be emitted as output.
- $f_{in}$: The filter function is based upon a boolean function $f_{in} \rightarrow E^{T} : (true, false)$. Given to functions $f_{in_1}$ and $f_{in_2}$ it is imperative that $f_{in_1} \equiv f_{in_2}$ are identical if and only if $patient\ (f_{in_1}) = patient\ (f_{in_2})$ that is, both functions relate to the same patient. Two mutually different functions $f_{in_1} \sqcap f_{in_2} = \emptyset$ are disjoint.
- TIL: The set of registered TIL's $TIL := \{til_1, \cdots, til_N\}$ in the TIL-Profile where $|TIL| \geq 1$, so at least one TIL has to be registered.
- VL: Every TIL-Profile consist of processing logic $VL$ which is a set of rules in terms of:

$$VL := \left\{ R \mid R = \left\{ \begin{matrix} f: E^{T} \rightarrow (true, false), & Filter \\ p: E \rightarrow E, & Pattern \\ t: E \rightarrow E, Transformation & \end{matrix} \right\} \right.$$

Two TIL-Profiles are identical $til_{Profil_1} \equiv til_{Profil_2}$ if and only if $et_{in_1} \equiv et_{in_2}$, $ET_{out_1} \equiv ET_{out_2}$, $f_{in_1} \equiv f_{in_2}$, $TIL_1 \equiv TIL_2$ and $VL_1 \equiv VL_2$.

### 3.5 Architectural insights into TiEE

The architecture of TiEE is based upon the event processing engine Esper which is commonly used in many commercial products. **Fig. 3** gives a broad overview about the main components and concepts like described in the past sections.

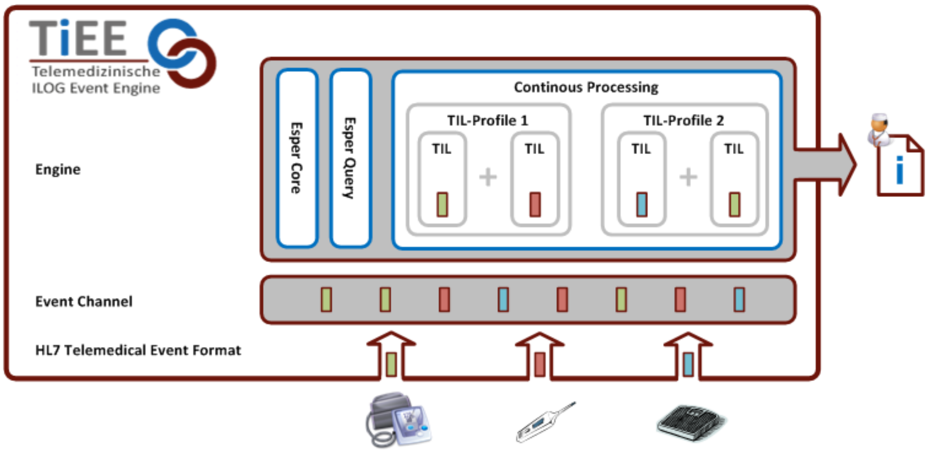


**Fig. 3.** TiEE architectural overview based upon the event processing engine Esper.

Starting at the bottom we have some kind of vital sign sensors from different manufactures. The sensors are connected to TiEE through Bluetooth HDP, supporting the IEEE 11073 standards family. To achieve an overall processing a single vital sign is interpreted as a telemedical event and will be encapsulated in the HL7 Telemedical Event Format. All events get transported to the event channel ordered by time. Above the channel we build the engine, introducing TIL-Profiles and TIL using Esper core and Esper queries of the Esper engine.

To realize a patient specific filtering like required by the TIL-Profile we extended Esper with the concept of PES, a patient-individual event stream. A PES formally represents an individual event stream that supports and bundles all types of events of one single patient. Thus there is a bijection between the patient and PES. Accordingly, any PES are pairwise different, in other words all PES are distinct, as each PES is characterized by a different patient, $PES_n \neq PES_m \; \forall \, n, m$ where $n \neq m$. The concept of a PES is based on the Variant Stream of Esper. All events of one single patient will be redirected in separate PES, with the help of a filter criterion, which checks for an individual patient identification. Likewise, a PES serves as the event source for all TILs in one TIL-Profile.

Furthermore we implemented methods and services to administrate TiEE. The developed services serve the purpose to modify the current system afterwards. So it is possible to add or remove main components like TIL and TIL-Profiles. In more detail the services allow to add, modify and remove patients, TILs and statements. Also, the services provide information about the components within the system. All services are designed as REST services and work with JSON objects.

## 4 Evaluation

The evaluation of TiEE is done upon two different data sets gained from two different projects. The first project is FitPit, the Fitness Cockpit, a solution to optimize preventive and rehabilitative trainings developed at Fraunhofer ISST [19]. Pulse and oxygen saturation as well as weight and blood pressure will be measured at the beginning and at the end of the training. In total we recorded 2450 measurements gained from 10 patients in terms of a long term measurement. The second project is the Vital Signs Dataset of the University of Queensland [16]. They recorded over 10 vital sign parameters from 32 patients undergoing anesthesia. The data was recorded with a solution of 10ms. Thus, in total we have around 240.000 measurements per patient, depending on the duration of the surgery. Now, before we can apply the datasets above, we have to define the main questions that have to be evaluated using TiEE:

- Question 1: Is TiEE capable of reducing the amount of incoming data and process them according to the defined information demand?
- Question 2: Is TiEE capable to process long-term trends as well as high frequent data?
- Question 3: Does the usage of TiEE reduce the overall implementation efforts for an analytical infrastructure?

What we won't try to answer at this point is, if TiEE is capable of being better in decision making compared to a physician. To answer the first question we started with modeling the information demand using CAS like mentioned before. Thus, we can prove that it is possible to map some kind of formalized demand to processing rules. To show the amount of data reduction we calculated the percentage of CTPEs per Minute. Within the FitPit scenario we had 1/11520 CTPEs per minute, respectively one trend within eight days. This very low value is caused by the long-term characteristic of the use case. In **Fig. 4** we show the ratio of incomings events and fired activities. Thus, there is a recognizable reduction of data or irrelevant information because activities are only fired according to a formalized information demand.

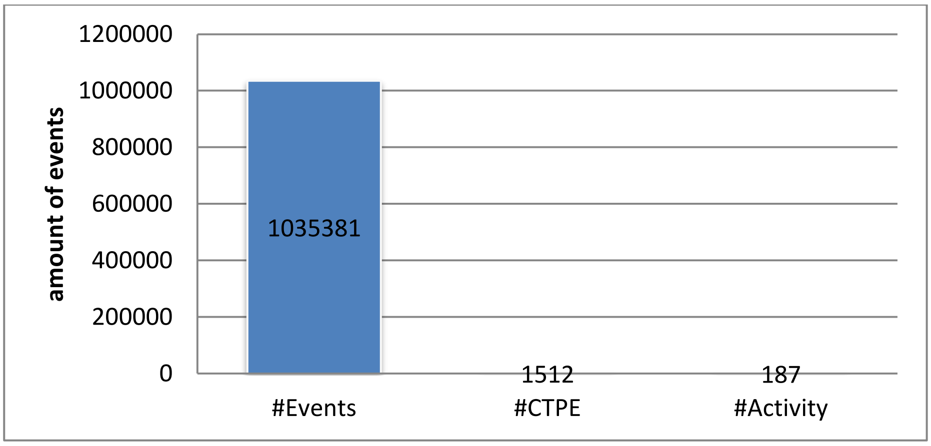


**Fig. 4.** Ratio between incoming events, generated CTPEs and fired activities of one patient of the University of Queensland data set.

The high-frequent data of the second use case produces 8.3 CTPEs per minute. It is possible to optimize or modify the percentage of CTPEs by configuring the underlying algorithm. Thus, we can also show that TiEE is capable to process long-term as well as high-frequent data. We also evaluated the performance of TiEE including routing and algorithmic processing of the incoming events. The average duration is around 15ms per event. The third question is implicitly proven by the usage of TiEE within two different scenarios. We were able to reuse once defined TILs within both scenarios in terms of a repository. Thus, the only effort was to model the information demand and map it to a TIL-Profile. Furthermore with TiEE a basic infrastructure for communication I/O is already given. Summarized we can point out that TiEE supports an optimization of information supply by reducing the amount of data.

## 5 Conclusion and Outlook

With TiEE, the Telemedical ILOG Event Engine, we investigated a solution to cope with one main problem in telemedical scenarios: information overload. Especially the continuous measuring of vital signs requires an intelligent reduction and processing of data in real-time. Since a medical situation changes very often over time, TiEE uses TIL-Profiles and TILs as a modular concept to facilitate the reuse of once developed rules and algorithms. Like described above, trend detection is an important class of algorithms for pattern recognition in streams of vital signs. At the moment TiEE is based upon trivial calculations using the CUSUM method to detect critical increasing, decreasing and stagnation of them. To show the technical feasibilities of TiEE we executed a first evaluation of the conceptional and implementational insights. TiEE need around 15ms per event and is capable to reduce the amount of date by generating CTPEs. In the future we'll start to evaluate also the medical evidence of TiEE.

## 6 References

1. Arasu, A. et al.: STREAM: The Stanford Data Stream Management System. Stanford InfoLab (2004).
2. Bates, J. et al.: Using Events for the Scalable Federation of Heterogeneous Components. Proceedings of the 8th ACM SIGOPS European workshop on Support for composing distributed applications. pp. 58–65 ACM, New York (1998).
3. Chandy, K.M. et al.: Towards a Theory of Events. Proceedings of the 2007 inaugural international conference on Distributed event-based systems. pp. 180–187 ACM, New York (2007).
4. Chandy, M.K. et al.: 10201 Executive Summary and Manifesto -- Event Processing. In: Chandy, K.M. et al. (eds.) Event Processing. Schloss Dagstuhl - Leibniz-Zentrum fuer Informatik, Germany, Dagstuhl, Germany (2011).
5. Charbonnier, S.: On-Line Extraction of Successive Temporal Sequences from ICU High-Frequency Data for Decision Support Information. AIME. pp. 1–10 (2003).
6. Deiters, W. et al.: Report Informationslogistik - Informationen just-in-time. Symposion Publishing (2001).
7. Deiters, W. et al.: The Information Logistics Approach toward User Demand-Driven Information Supply. Cross-media service delivery : Based on papers presented at the Conference on Cross-Media Service Delivery - CMSD-2003. Kluwer Academic Publishers (2003).
8. Geesen, D. et al.: Data Stream Management in the AAL - Universal and Flexible Preprocessing of Continuous Sensor Data. In: Wichert, R. and Eberhardt, B. (eds.) Ambient Assisted Living. pp. 213–228 (2012).
9. Haftor, D.M., Kajtazi, M.: What is Information Logistics? An Explorative Study of the Research Frontiers of Information Logistics, (2009).
10. Haimowitz, I.: Knowledge-based trend detection and diagnosis. Massachusetts Institute of Technology (1994).
11. Hinchley, A.: Understanding Version 3: A primer on the HL7 Version 3 Healthcare Interoperability Standard. Münch, A (2007).
12. Hunt, R.E., Newman, R.G.: Medical Knowledge Overload: A Disturbing Trend for Physicians. Health Care Manage Rev. 22, 70–75 (1997).

13. IEEE: ISO/IEEE 11073-10201:2004: Health informatics Point-of-care medical device communication Part 10201: Domain information model. (2004).
14. Koch, O.: Process-based and context-sensitive information supply in medical care. Proceedings of the CIR'07 Workshop on Context-Based Information Retrieval in conjunction with CONTEXT-07. (2007).
15. Koch, O., Rotaru, E.: Using Context to Improve Information Supply in the Medical Sector. In: Abramowicz, W. et al. (eds.) Business Information Systems Workshops - BIS 2010 International Workshops. pp. 192–203 Springer (2010).
16. Liu, D. et al.: University of Queensland Vital Signs Dataset: Development of an Accessible Repository of Anesthesia Patient Monitoring Data for Research. Anesthesia & Analgesia. (2011).
17. Lowe, H.J. et al.: STRIDE--An Integrated Standards-based Translational Research Informatics Platform. AMIA. 2009, 391–395 (2009).
18. Luckham, D.C.: The Power of Events: An Introduction to Complex Event Processing in Distributed Enterprise Systems. Addison-Wesley, Boston (2002).
19. Meister, S. et al.: FitPit – Fitness cockpit: information system to optimize training Schedules in Rehabilitation. ECIS 2013 Proceedings. 106, (2013).
20. Meister, S.: Telemedical Events: Intelligent Delivery of Telemedical Values Using CEP and HL7. In: Niedrite, L. et al. (eds.) BIR Workshops. Springer (2012).
21. Meister, S.: Usage of CEP and HL7 to Solve Information Logistics Problems in Telemedicine. 4th International Workshop on Information Logistics and Knowledge Supply for Viable Enterprises (ILOG 2011). (2011).
22. Meister, S., Koch, O.: Using Complex Event Processing and Context for Intelligent Information Filtering and Supply. In: Schilling Klaus and Leutert, F. (eds.) 1st IFAC Conference on Embedded Systems, Computational Intelligence and Telematics in Control CESCIT 2012. (2012).
23. Meister, S., Stahlmann, V.: Telemedical ILOG Listeners: Information Logistics Processing of Telemedical Values Using CEP and HL7. In: Wichert, R. and Eberhardt, B. (eds.) Ambient Assisted Living - Advanced Technologies and Societal Change. Springer (2012).
24. Society for Medical Decision Making: Proposal for Clinical Algorithm Standards: Society for Medical Decision Making Committee on Standardization of Clinical Algorithms. Medical Decision Making. 12, 2, 149–154 (1992).
25. Weber, S. et al.: Implementing a Real-time Complex Event Stream Processing System to Help Identify Potential Participants in Clinical and Translational Research Studies. AMIA. 2010, 472–476 (2010).
26. Willems, A. et al.: Information Logistics Research Report - Frameworks in Healthcare Industry. (2009).
27. Wilson, T.D.: Information Overload: Implications for Healthcare Services. Health Informatics Journal. 7, 112–117 (2001).
28. Wilson, T.D.: Models in information behaviour research. Journal of Documentation. 55, 3, 249–270 (1999).

# The Time Dimension in Information Logistics

Andrejs Gaidukovs and Marite Kirikova

Institute of Applied Computer Systems, Riga Technical University, Latvia
{andrejs.gaidukovs; marite.kirikova}@rtu.lv

**Abstract.** The purpose of information logistics is to ensure that the right information, which is necessary in accomplishment of business tasks, is available in the right location, in the right time, and in the right quality. Different combinations of models have been suggested for supporting this purpose. However, none of them includes the model that corresponds to the time dimension of information logistics. Taking into consideration that the right time is an essential goal that corresponds to the purpose of information logistics, this paper takes a closer look at the time dimension and suggests extending the set of related models of information logistics with the time dimension model.

**Keywords:** information logistics, time dimension, information demand model

## 1 Introduction

Information Logistics is a branch of research that addresses concepts, methods, technologies, and solutions that help to ensure situation sensitive availability of high quality information for individuals or groups with respect to their information needs, time, location, and user-friendly form of representation [1]. As revealed in [2] the notion of Information Logistics was coined in 1978. According to data available in scientific repository SCOPUS, the number of papers on information logistics has gradually increased from one paper in 1982 [3] to more than 40 papers per year during 2009-2012.

Information logistics usually is organized around the following four main dimensions [4]:

- *Personalization*: each person or group has particular information needs that depend on their knowledge and experience
- *Time*: information has to be available or delivered in a particular time when it is actually needed
- *Communication*: information has to be available (or represented) in the form that is convenient for its users
- *Context*: information has to be deliverable regarding the location and situation, in which it is used in a particular moment of time

One more dimension – the *quality* dimension of information logistics is suggested in [1].

Most of the research in the information logistics has been devoted to the context dimension, e.g., [5], [6], [7], and [8]. Usually the time dimension is considered from the point of view of sequence of time moments or intervals in which particular items of information are needed. The intervals can be relative, e.g., "four weeks before a particular time moment", "six weeks before a particular time moment" and so forth. The moments and start and end points of the intervals are represented on a scalar time axis. Thus, while such issues as roles, tasks, information items, and quality parameters are addressed in specific interrelated models, the time dimension has not been considered as a separate information logistics model so far.

The goal of this paper is to analyze the time dimension in depth and propose the model of the time dimension that would enhance possibility to represent and analyze information needs and information availability and delivery patterns more accurately.

The paper is organized as follows: In Section 2 features of the time dimension are discussed. In Section 3 conceptual modeling of time dimension is considered and the model of the time dimension is presented. In Section 4 the time dimension is put in the context of information demand modeling. In Section 5 the practical applicability of the time dimension model is discussed. Brief conclusions are stated in Section 5.

## 2 Features of Time Dimension

In this section the discussion of the time dimension is pragmatically oriented; it is mainly based on research reflected in [9] and [10] and does not concern philosophical considerations of time dimension.

The features of time can be structured as reflected in Fig 1.

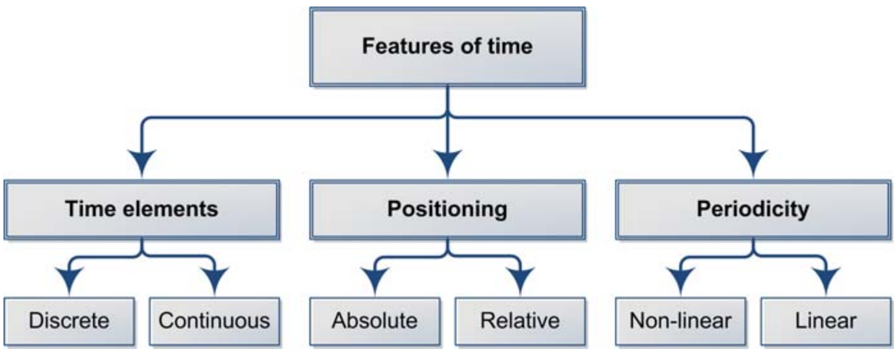


**Fig. 1** Features of time

Time can be considered as discrete items or as a continuous phenomenon. In information technology context it is mainly considered as discrete items. It can be positioned as absolute time referring to a chosen "clock" or as a relative time (e.g., "one year ago"). An essential feature of time is "periodicity" or frequency of intervals that can be reflected using linear or non-linear scales. It is essential that users are free to define time periods by themselves. In related work one can find the following types or aspects of the time dimension:

- Calendar granularity (N years, Year, Half-year, Quarter, Month, Week, Day, Hour, Minute, Second)
- Monthly (January, February, March, April, May, June, July, August, September, October, November, December)
- Daily (Monday, Tuesday, Wednesday, Thursday, Friday, Saturday, Sunday)
- Part of the day (Day time, Night time)
- Seasonal (Spring, Summer, Autumn, Winter)
- Arbitrary time intervals (periods)
- Arbitrary time moments
- Time zone
- Process time (Start time, Duration, End time, Deadline)
- Relative pointers (Before, After, Now, Every)
- Time types (systems time, real time, see e.g. ISO SQL:2011 standard [11])

The model of the time dimension, which is proposed in the next section, includes most of above-listed time dimension types.

## 3 The Model of the Time Dimension

While there is rarely specific time model available in the related work, time has been included in different conceptual models. Ten of such models are surveyed in [9]:

- *Infologistic data model* by Langefors, presented in 1973 [12]. In this model the time (moment or period) is an attribute of the elementary fact, belonging to its built-in context. E.g., $O(p)$ is a set of objects with property p, $O_t(p)$ is a set of objects which have property *p* during time period *t*.
- *Conceptual Information Model* [13] distinguishes between extrinsic time automatically set by the software system and intrinsic time (a part of the fact that refers to real time when the fact is true). Here the time is considered as a component of the relationship. In its later modifications the distinction between time moments and time intervals is taken into consideration.
- *The time extended Entity Relationship Model – TERM* [14] focuses on historical structure of data, where each attribute, role, or relationship has historical property. This is achieved by specific basic, derived, or inductive historical operations.
- *Logic based model* proposed by Lundberg [15], which argues that, despite of continues nature of time, in information systems it is to be represented as a discrete phenomenon.
- *INFOLOG* [16] model maintains time dimension as a set of temporal operators. This model ignores future and focuses only on past and present issues of facts.
- *DMILT* [17] model is based on specific temporal logic that addresses process network, where processes exchange messages via the temporal database of the information system.

- *The Entity, Relationship, Attribute, Event (ERAE)* [18] model distinguishes between past, present and future of individual data entities or groups of entities. Time is represented as a specific data type.
- *The Conceptual Modeling Language (CML)* [19]. This language has object-oriented structure. It considers discrete time reflected as time intervals using such predicates as *meets*, *equals*, *during/over*, *before/after*, *startsbefore/startsafter*, *endsbefore/endsafter*, *overlaps*, *costarts*, and *coends*. This language includes also two constants, namely, *Alltime* and *Now*.
- *TEMPORA* [20] consists of two types of models such as Entity Relationship Time (ERT) model and Conceptual Rule Language (CRL). Time is reflected as time stamps of entities and relationships in ERT. CRL includes different predicates that allow reflecting various time based situations. Time moments and time intervals are considered (for events – only time moments). Modeling time and real time are considered, as well as historical relationship is presented.

While above-listed methods have been developed quite a time ago and time is an essential issue in contemporary business and information systems environment, there are not many research works available that would consider conceptual modeling aspects of the time dimension. We can distinguish between the following main areas where time issues are currently discussed:

- *Neuropsychology*
- *Simulation*
- *Data Warehouses*
- *Business Intelligence*
- *Ontology Engineering*

In neuropsychology, simulation, data warehouses, and business intelligence only some aspects of the time dimension are considered [21], [22], [23], [24]. The broadest scope of the aspects is analyzed in the research on the time ontology [25], [26].

The fact that recently issued SQL:2012 standard [11] considers only some aspects (calendar granularity, real time, and systems time) of the time dimension, shows that it is necessary to further research the time dimension to better understand its features and incorporate it into models of information logistics.

In this paper we present the first version of the time dimension model that has been developed, mainly, by amalgamating various aspects of the time dimension in a single model. This was done with the purpose to obtain a generic view on different aspects of time relevant in information logistics and to have a possibility to control relationships between these aspects in the time dimension model. The simplified version of the proposed model is presented in Fig. 2. The model presented in Fig. 2 has been obtained independently of the time ontology presented in [26]. While the proposed model has many similarities with the time ontology, such as inclusion of parts of the year, time zones, etc; still there are some essential differences how periodicity and time moments are handled. The model in Fig. 2 distinguishes between periods and intervals as separate entities since both the length of transactions (intervals) and the periods of time showing repetitive execution of transactions are important in information logistics. More detailed comparison of the time ontology and proposed time dimension model is beyond of the scope of this paper.

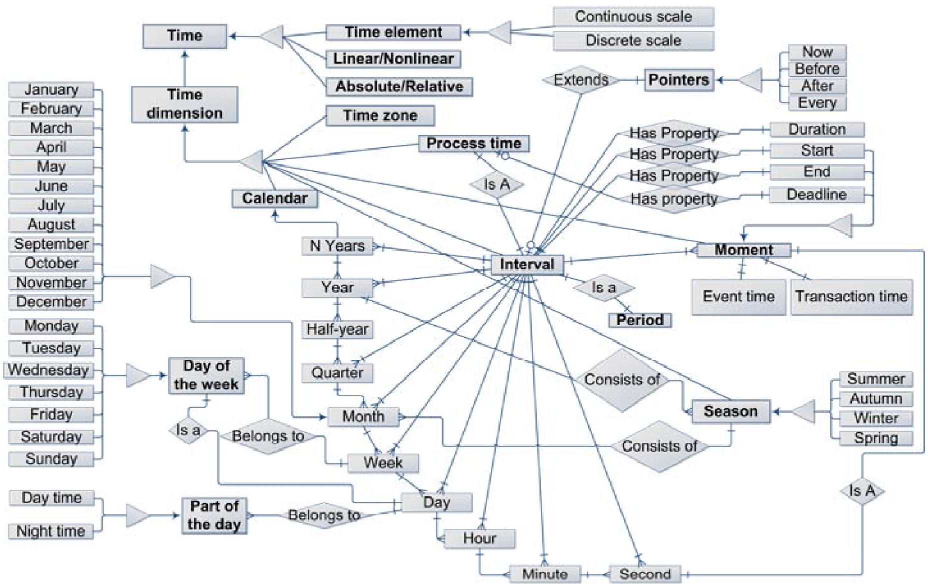


**Fig. 2** Time dimension model (simplified)

Fig. 2 reflects different aspects of time that can be relevant in various tasks of information logistics (See Section 2). The positioning of the time dimension model in the context of other models of information logistics is discussed in the next section.

## 4 Time Dimension and Information Demand Model

For illustrating the role of the time dimension in information logistics we use Information Demand Pattern discussed in [2]. From the modeling perspective this pattern prescribes the model that consists of the following sub-models:

- Information Model that represents the items of information used in performance of tasks
- Effects Model that reflects different issues of quality of information logistics, such as economical effect, motivation, experience etc.
- Organizational Chart showing relationships between the roles in organization
- Task Model showing the architecture of the tasks.

The elements of above-listed models should be iner-related in order to show, which information is needed for which role when performing a particular task, and what effects the availability/unavailability of this information may cause. Fig. 3 illustrates how these models can be related to the time dimension model.

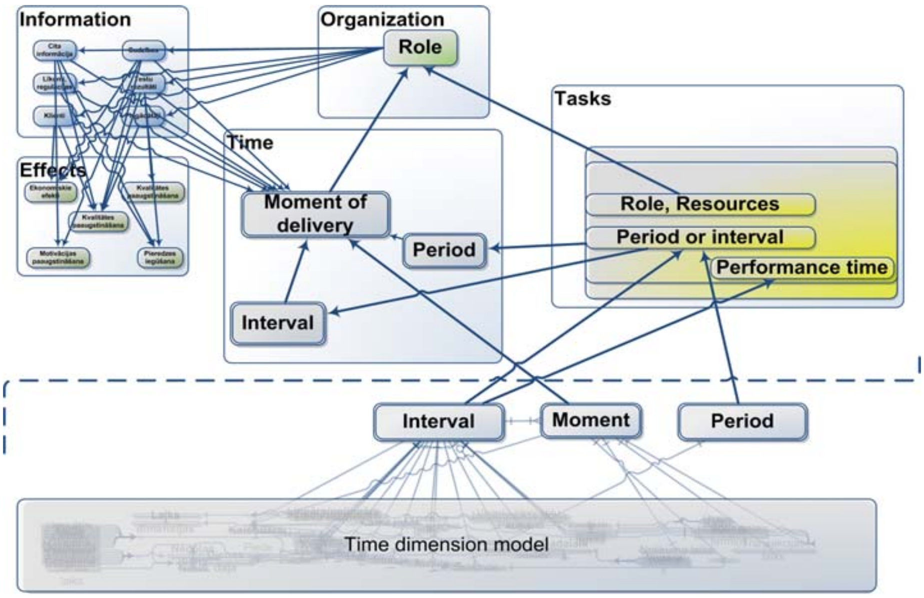


**Fig. 3** Information Demand Model related to the time dimension model

At high level of abstraction Fig. 3 shows how elements of Information Demand Pattern can be related to the time dimension model. However, any element of the time dimension model reflected in Fig. 2 can be indirectly related to Information Demand Pattern. This is illustrated by the time dimension model bellow the dotted line in Fig. 3. The usage of the time dimension affects the Task Model. Each Task Model's element has to have attributes that help to characterize their duration (performance time or interval) and periodicity (period or interval of repetition of the task).

We can distinguish between the following types of tasks:

- Meetings: project meetings, conferences, seminars
- Ordinary Tasks: performing a particular transformation

On one hand, the type of the task does not change the way how the time dimension is incorporated in Information Demand Pattern. On another hand, the experimentally obtained models, which represented particular Information Demand Pattern's (extended by the time dimension model) instances, revealed that the extended Information Demand Pattern models have slightly different outlooks of above-mentioned types of tasks. Nevertheless, in both cases there is a time interval $\Delta t$ that is related to start time point $t_s$ and end time point $t_e$ of the task

$$\Delta t = t_e - t_s \tag{1}$$

Moment of Delivery $M_d$ is defined as a moment in which a particular information unit is sent to the role. In most cases this moment should not occur later than the beginning of the task.

Both types of tasks can occur periodically. In this case it is necessary to define period $P$ as the delta between two sequential starting points of tasks $t_{s.n}$ and $t_{s.n+1}$,

$$P = t_{s.n+1} - t_{s.n} \tag{2}$$

The periods can be constant and arbitrary. In case they are constant ones (e.g. handing in monthly reports), the $M_d$ can be calculated as follows:

$$M_d = P - \Delta t \tag{3}$$

Often in information logistics the moment, when information is received, can be considered being practically equal to the moment, when information is delivered. However, there are cases, when the time of delivery has to be taken into account.

## 5 Experimental Application of Time Dimension Model

The time dimension model proposed in Section 2 related to Information Demand Pattern (Section 3) was applied for information demand modeling in the project proposal preparation in a public organization. Effects Model was not used in the experiment. The models illustrating the obtained information demand pattern are shown in Fig. 4.

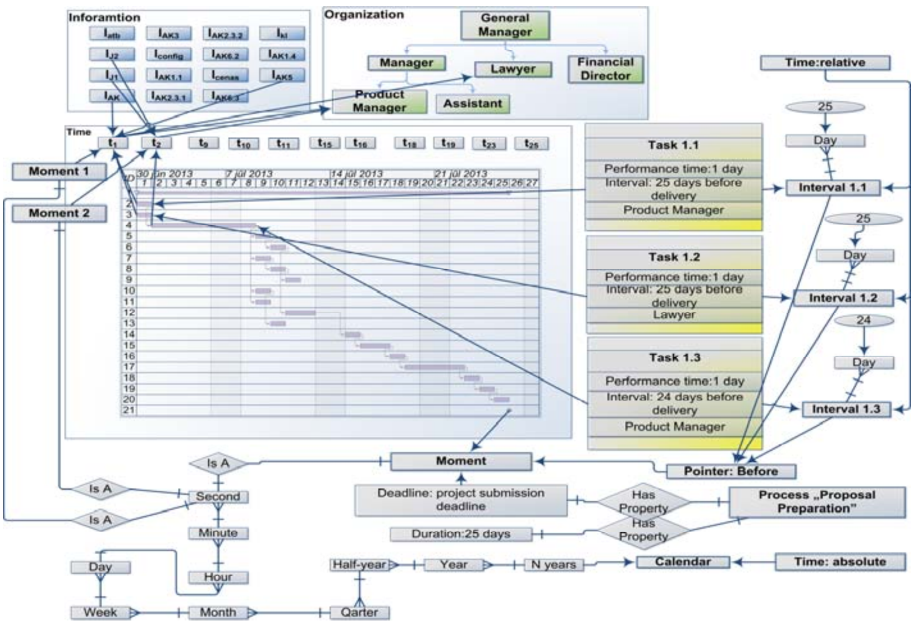


**Fig. 4** Part of Information Demand Pattern models for project proposal preparation extended by the time dimension model

There were six organizational roles, nineteen tasks, and twenty five information units defined. The maximal duration of the project proposal preparation was 25 days. On the basis of the models, the information system's infrastructure and the data model were defined for supporting the tasks represented in the Task Model. The time dimension model was helpful for identification of time issues and attributes relevant for information logistics of project proposal preparation. The model helped to represent and analyze information needs and show information availability and delivery patterns more accurately.

## 6 Conclusions

The paper focuses on the time dimension in information logistics. It contributes the first version of the conceptual time dimension model for information logistics as well as relates well-known Information Demand Pattern to the time dimension.

First experiments with inclusion of the time dimension model in the set of inter-related information logistics models show that the time dimension model helps to represent and analyze information needs and show information availability and delivery patterns accurately. The model is useful for visualizing the tasks of roles and designing information architecture that supports the performance of the tasks.

Further research should concern tuning of the time dimension model, lager scale experiments with Information Demand Pattern extended by the time dimension model, and designing transformation algorithms that produce role oriented and information technology support oriented views on the detailed sub-models of the extended Information Demand Pattern.

## References

1. Sandkuhl K.: Information Logistics in Networked Organizations: Selected Concepts and Applications. Enterprise Information Systems: 9th International Conference, ICEIS 2007, Funchal, Madeira, June 12-16, 2007, Revised Selected Papers. Springer, pp. 43-54 (2009).
2. Sandkuhl K., Information Logistics, The University of Rostock (2011).
3. Reichertz P.L.: Future developments of data processing in health care. Methods of Information in Medicine 21 (2), pp. 55-58 (1982).
4. Haseloff S.: Context awareness in Information Logistics, PhD thesis, Technischen Universität Berlin, Berlin, Germany, 247p. (2005).
5. Sandkuhl K.: Information Demand Patterns, Xpert Publishing Services, Rome (2011).
6. Sandkuhl K., Smirnov A., Shilov N.: Information Logistics in Engineering Change Management: Integrating Demand Patterns and Recommendation Systems. Workshops on Business Informatics Research: BIR 2011 International Workshops and Doctoral Consortium, Riga, Latvia, October 6, 2011 Revised Selected Papers, Springer Berlin Heidelberg, pp. 14-25 (2012).
7. Michelberger B., Mutschler B., Reichert M.: A Context Framework for Process-Oriented Information Logistics. Business Information Systems. 15th International Conference, BIS 2012, Vilnius, Lithuania, May 21-23, 2012. Proceedings., Springer Berlin Heidelberg, pp. 260-271 (2012).

8. Sandkuhl K.: Material Specification Responsible, Jonkoping University (2009).
9. Theodoulidis C.I., Loucopoulos P.: The time dimension in conceptual modelling", Information Systems, vol. 16, pp. 273-300 (1991).
10. Chendroyaperumal C., Time Dimensions for Managerial Excellence, SSRN:1544363 (2010).
11. Kulkarni K., Michels J.-E.: Temporal features in SQL:2011. SIGMOD Rec.,vol. 41, pp. 34-43, (2012).
12. Langefors B.: Theoretical analysis of information systems, Studentlitteratur., Lund (1973).
13. Bubenko J.: Information Modelling in the context of System Development. IFIP Information Processing '80, Amsterdam, pp. 395-411 (1980).
14. Klopprogge M.R.: TERM: An Approach to Include Time Dimension in the Entity-Relationship Model. Proceedings of the Second International Conference on the Entity-Relationship Approach to Information Modeling and Analysis, North-Holland Publishing Co., Amsterdam, pp. 473-508 (1981).
15. Lundberg B.: An axiomatization of events. BIT Numerical Mathematics, vol 22. (1982).
16. Sernadas A.: Temporal aspects of logical procedure definition. Information Systems, vol. 5, pp. 167-187 (1980).
17. Carmo J, Sernadas A.: A Temporal Logic framework for a fayered approach to systems specification and verification. Proceedings of the IFIP 8/WG8.1 Working Conference on Temporal Aspects in Information Systems, pp. 31-46 (1987).
18. Dubois E., Hagelstein J., Lahou E., Ponsaert F., Rifaut A.: The ERAE Model: A Case Study. Proc. of the IFIP WG 8.1 Working Conference on Information Systems Design Methodologies: Improving the Practice, Amsterdam, North-Holland Publishing Co., pp. 87-105 (1986).
19. IConceptual Modelling Language : An Informal Description Report, Institute of Computer Science, Crete (1987).
20. Theodoulidis C., Wangler B., Loucopoulos P.: Requirements Specification in TEMPORA. 2nd Nordic Conference on Advanced Information Systems Engineering, Kista (1990).
21. Malinowski E., Zimányi E.: A conceptual model for temporal data warehouses and its transformation to the ER and the object-relational models. Elsevier. Data & Knowledge Engineering, vol. 64, issue 1, January 2008, pp. 101–133 (2008).
22. Kelly Th. J.: Dimensional Data Modeling, Solutions, Sybase, http://www.olap.it/Articoli/DimensionalDataModeling.pdf accessed August, 2013 (2013).
23. Teki S., Grube M., Griffiths T.D.: A unified model of time perception accounts for duration-based and beat-based timing mechanisms. Frontiers in Integrative Neuroscience, January 2012, vol. 5, article 90, pp. 1-7 (2012).
24. Lobato R.S., Ulson R.S., Santana M.J.: A Revised Taxonomy for Time Wrap based Distributed Synchronization Protocols. Proceedings of the Eighth IEEE International Symposium on Distributed Simulation and Real-Time Applications (DS-RT 2004), October 21-23, 2004, pp. 226-229 (2004).
25. Hobbs J. R., Pan F.: An Ontology of Time for the semantic Web. ACM Transactions on Asian Language Information Processing, vol. 3(1), March 2004, pp. 66-85 (2004).
26. Hobbs J.R., Pan F.: Time Ontology in OWL, W3C Working Draft 27 September 2006, available at http://www.w3.org/TR/owl-time/ (2006).

# A Description Model for Policy Conflicts for Managing Access to Health Information

Raik Kuhlisch

Fraunhofer Institute for Open Communication Systems FOKUS,
Kaiserin-Augusta-Allee 31,
10589 Berlin, Germany
raik.kuhlisch@fokus.fraunhofer.de

**Abstract.** For a better business and IT alignment in hospitals an ontology-based description model for policy conflicts is introduced. Such a model is a necessary prerequisite for the subsequently domain-specific policy conflict handling as a hospital information management related activity.

Keywords: E-health, hospital intra-enterprise policy conflict, policy compliance verification, information security, conflict model

## 1 Introduction

One of the key aspects with regard to the hospital information management is the fulfillment of legal and regulatory requirements. This covers both patients' privacy instructions on the use and/or disclose of his data and the operational need by the (medical) staff to access classified information in order to treat them well. As a consequence the legal processing of health information needs to be directed accordingly. Policies are a feasible means to describe and regulate reasonable actions that underpin the compliance of a hospital and its staff. A recent literature review emerges that the modeling, measurement and evolution phases of business and information technology (IT) alignment is still an issue [1]. One of the main technical challenges in an alignment approach is how to define and apply such policies as means to translate organizational requirements into guidelines and rules in IT management.

There are multiple sources (e.g., legal provisions, regulations, patient consents) given that limit the processing of medical data and might be expressed with various kinds of policies. Briefly, major policy concerns for healthcare information exchange that can be applied in hospitals are (1) patient privacy consents, (2) purpose of use, and (3) compliance [2]. The patient explicitly identifies authorized identities that are allowed to use his data. The consent might have additional prerequisites such as an objection to the use of pre-treatment data. Derived constraints from the intended use of a certain hospital information subsystem that mediates access to protected information are subsumed under the term purpose of use. This describes limitations of the overall functionality of a subsystem under specified conditions (e.g., certain medical treatment step or workflow). It provides information on which tasks can be carried out using the specific application systems and what actions on the underlying data are permitted. The information management compliance concludes the policy concerns. Since the hospital is liable for the lawful processing of the patient data, compliance allocates responsibilities (roles, permissions, and obligations) for internal and external communication of data.

In support of the above mentioned monitoring activities it should be investigated to what extent different levels of policies during regular operation are compliant to each other. This paper addresses conflicts between these policy concerns as well as concrete policy types. As a first contribution, an ontology based description model for policy conflicts that might melt the protection of patient privacy is presented.

The remainder of this paper is structured as follows: Section 2 introduces specific policy types derived from generic policy concerns as a background for policy conflicts. In Section 3, our conflict model is defined. Section 4 briefly reflects a methodology for detecting policy conflicts. Section 5 summarizes this paper and future research directions are addressed.

## 2 Policies for Accessing Patient Data

The general notion of a policy is to regulate the system behavior: Instead of re-implementing parts of the system solely the respective policy is applied [3]. Taking for granted the ideas of a definition of a security policy in [4], *organizational security policies* and *automated security policies* can be differentiated in general. An organizational security policy is understood as a set of rules and practices that determine how an organization manages and protects its resources to achieve defined security policy objectives. Quality criteria, such as accuracy, reliability and robustness of a system are excluded from this definition since this cannot be directly protected. Restrictions and properties that determine how an IT system prevents access to information and its resources are regarded as automated security policies [4].

These security policies can be applied to the information management in hospitals, too. Obviously, policy hierarchies with the range of corporate policies, task oriented policies, functional policies, and low-level policies can be found [3]. In addition, there are other factors which are relevant for protecting health information or patient data.

The shell model in Figure 1 depicts influencing policy factors for medical data access which are previously investigated in [2]: The ultimate baseline is the national law which scopes the legal disclosure of patient data by members of health care professions. This is accomplished by a patient consent to shape the processing of patient's data. Due to the contractual nature of the consent the following forms are to be distinguished: implied consent, presumed consent, and explicit consent. The first consent type may be caused by the appearance of the patient without his written consent (his declaration of intent is assumed). The second type may be caused when the patient is unable to express his consent to the disclosure of his personal data, but he would reasonably do (e.g., in an emergency case). The latter one may be made orally. However, for reasons of securing evidence this should be in writing.

Nevertheless, patient privacy consents join the policy factors for medical data access. The corporate governance strategy defines compliance requirements—even for information management. Corporate governance defines IT strategies with roles and responsibilities and makes requirements for the organization of work. Finally, application systems mediate the access to patient data that is governed by an identity and access management.

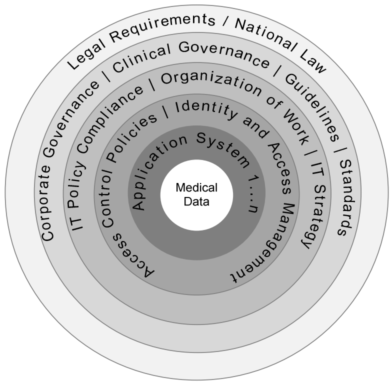


**Fig. 1.** Influencing policy factors for medical data access

### 2.1 Policy Types

Based on these factors, the security policy definitions above and definitions in [2] the following different policies types can be defined which is shown in Figure 2.

- An *information access policy* regulates who is authorized to disclose information by defining of confidentiality of at least one specific classified information object. Therefore, access demands to information result. Such policy is a special part of an overlying *information security policy* that in turn contains further regulations and information such as an overview of the corporate philosophy on security, a statement of purpose, and organization's security responsibilities that define the security organization structure [5, p. 248].
- A *resource access policy* is the special part of a *patient privacy policy* that authorizes certain individuals and organizations to use the HIS subsystem with regard to the agreed purpose of use. Technically, safeguarded entry points are authorized by this policy. A policy template might be *"I hereby authorize [roles] at [organizations] to use the 'Historical Database' Application in order to access all [Patient] [kind-of-data] for the purpose of [purpose]"* which may result in the following instance: *"I hereby authorize physicians at Clinic A to use the 'Historical Database' Application in order to access all my lab data for the purpose of medical treatment"* [2].
- A *resource behavior policy* is the counterpart of the resource access policy since it defines how certain subjects might act on certain HIS subsystems based on

their functional/structural role. Whereas the resource access policy is patient-driven, the resource behavior policy is organization-driven (i.e., it reflects the structural organization and process organization).

- An *access control policy* contains the actual access rights configured in a dedicated HIS subsystem. There are several nuances of this policy type defined in [6]. Hence, *authorization policies*, *obligation policies*, *refrain policies*, and *delegation policies* are to be distinguished additionally.

This overview reveals the versatility of policies in hospitals. To conclude, an information access policy and a resource access policy are regarded as 'pure' organizational security policies with respect to the above definitions. On the contrary, a resource behavior policy and an access control policy complement the regulated access to patient data as automated security policies.

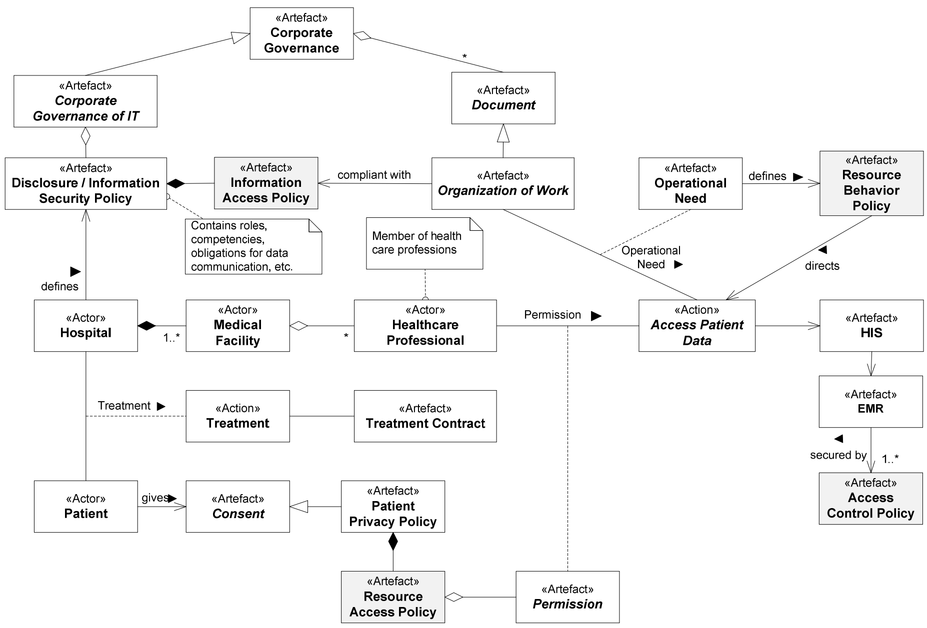


**Fig. 2.** Range of policies to control the patient's data access (UML)

## 3 Formalization of Policy Conflicts

Policy conflicts might occur between policies of the same type/concern (e.g., two rules of two access control policies may contradict each other) or crosscutting (e.g., a patient disagrees with the usage of a health care application in a resource access policy but is a necessity for the organization of work that is reflected by a resource behavior policy). The challenge is to identify conflicts between different policy concerns since more factors for data access have to be considered. This section defines policy conflicts. Afterwards, the development of the description model is introduced.

### 3.1 Policy Conflict Notion

It is obvious that similar elements (e.g., subjects, actions) must be included in the various policies to give rise to a conflict. According to [7], a policy conflict “occurs when the actions of two rules (that are both satisfied simultaneously) contradict each other”. This definition is vague since it does not look at the different policy concerns. That is why policy conflicts are either of the same policy type or crosscutting. The latter one describes contradicting contexts and intexts (i.e., policy structures and behavior).

### 3.2 Policy Conflict Types

On the basis of the conflicting concerns defined in [2] the following policy types might be in conflict:

- *Conflicts between information access policy and resource behavior policy.* Likely, a common occurrence e.g. when access to application systems is configured ad-hoc without raising the question such as “why do an attending physician has access to the patient administration system”.
- *Conflicts between information access policy and access control policy.* This can solely happen if resource behavior policies are not synced with access control policies. That is, if entry points for application systems are safeguarded correctly, unknown identities cannot be authorized in an authorization policy. However, access might be denied although someone is allowed to use a dedicated application system (e.g., due to time constraints).
- *Conflicts between resource access policy and resource behavior policy.* Another likely scenario where a patient might restrict the use of his data for a special purpose of use, but a physician is authorized to use an application system to his task (e.g., hygiene or quality management).
- *Conflicts between resource access policy and information access policy.* Such conflicts arise when patients restrict/expand their use of data which simultaneously is a compliance breach. This is the case when identities are granted to access patient data via dedicated application systems but the organization of work does not consider such accesses. However, under the assumption that the resource behavior policy correctly implements the defaults by the information management (i.e., compatible with an information access policy) no patient data breach can occur.
- *Conflicts between multiple resource access policies.* Different permissions from patient privacy policies refer to different entry points to application or application systems. This conflict occurs if more than one resource access policies with the same context (subjects, resource) are activated for patient and have different actions are authorized.

- *Conflicts between multiple access control policies.* Actions or rather access rights contradict each other. To avoid such conflicts (activated) policies must be considered in the total. However, if the same subject is granted access to data via a specific application/application system and in turn denied to access the same data via another application/application system represents no conflict. Hence, different purposes of use are implemented.

From these conflicts the following classifications are extracted:

- *Positive-negative conflict of modalities:* By analogy with the structured model of policies and its conflict analysis defined by Moffett et al. [8], modality conflicts are conflicts regarding (to be authorized) actions and their associated policy goal. That is, on the one hand a modality expresses whether specific actions must be initiated or prevented (these actions refer to obligation and refrain policies) to achieve a goal—on the other hand it occurs if the execution of specific actions is permitted or forbidden (this refers to authorization policies). For instance, if one policy grants and another one denies access to the same resource it is a modality conflict.
- *Functional dependency:* The policy execution order is important for so that a conflict occurs. Or one policy requires permissions from another one (e.g., in delegation scenarios). This conflict type indirectly represents a priority conflict of access rights.
- *Term or attribute conflicts:* Especially when role-based access control scenarios are implemented, this conflict might occur. For instance, roles or attribute groups might have different meanings in other application domains and thus might create conflicts [9].
- *Semantic mismatches between policy concerns:* This conflict refers to goal conflicts (e.g., restriction of information implemented different in policies). Since actions implement/achieve the goal there might be a contradiction. For instance, the purpose of use in a resource access policy and a resource behavior policy have nothing in common. Hence, if the semantic match (i.e., similar concepts) cannot be checked, a conflict arises.

### 3.3 Modeling

Enterprise modeling serves as a useful tool for business and IT alignment [10]. This even applies to policy conflict analysis. Ideally, valid policies (i.e., conflict-free policies) are derived from one or more information models and deployed to application systems. Thus, a model has a verification-validation-testing function [11], which is essential for policy planners. The conflict model is based on a platform-independent model as an ontology. In the terminology of the model-driven development strategy MDA this model can be described as platform-independent. The approach is to first formalize the policy types and then define the concept of conflicts.

**Choice of Language.** Semantic Web technologies provide a rich pool of techniques for representing structured knowledge from weak to strong using ontologies.

Moreover, its logical reasoning and treatment of new contexts seem to fit very well to handle different policy types. Promising candidates are the Resource Description Framework (RDF) Schema[1] and the Web Ontology Language[2] (OWL) for structuring RDF resources. RDF Schema and OWL share the same syntax RDF/XML. With both vocabularies the possibility to define formal design models is given. When comparing the expressivity of RDF Schema to OWL, it is obvious that OWL is much more powerful than RDF Schema since its vocabulary is more comprehensive. Thus, for example, no disjoint classes are supported by RDF Schema or the necessary expressivity to define cardinalities does not exist. Nevertheless, with the primitives *rdfs:Class* and *rdf:Property* might be achieved any expressivity: Missing constructs can be reproduced by means of rules. The derivation of new knowledge (reasoning) for large ontologies with RDF Schema is—due to the lower computational complexity—easier. Since every OWL ontology is compatible with RDF Schema, it plays rather a subordinate role which language is used here. OWL provides unnecessary semantics (e.g., equivalent classes, complex classes, same individual) so that RDF Schema is used for modeling the policy types and their conflicts.

**Cardinality Constraints.** The expressivity of RDF Schema is sufficient. However, one handicap is that RDF Schema cannot express cardinality constraints. One workaround is to subclass *rdf:Property* named *RestrictedProperty*. This indirection gets the two *rdf:Property* elements *minCardinality* and *maxCardinality* with the range of values for integer data types. In this way, subclasses of policy objects can record cardinalities.

```
<rdfs:Class rdf:about="urn:policy-ns:basic-policy#RestrictedProperty">
   <rdfs:subClassOf rdf:resource="http://www.w3.org/1999/02/22-rdf-syntax-
   ns#Property"/>
</rdfs:Class>
<rdf:Property rdf:about="urn:policy-ns:basic-policy#minCardinality">
   <rdfs:range rdf:resource="http://www.w3.org/2001/XMLSchema#int"/>
   <rdfs:domain rdf:resource="urn:policy-ns:basic-policy#RestrictedProperty"/>
</rdf:Property>
<rdf:Property rdf:about="urn:policy-ns:basic-policy#maxCardinality">
   <rdfs:range rdf:resource="http://www.w3.org/2001/XMLSchema#int"/>
   <rdfs:domain rdf:resource="urn:policy-ns:basic-policy#RestrictedProperty"/>
</rdf:Property>
```

In order to check the integrity of cardinality with instances RDF Schema rules need to be extended (for the sake of clarity, the RDF/N3 sample solely checks whether minimum cardinality is set):

[1] See http://www.w3.org/TR/rdf-schema/.
[2] See http://www.w3.org/TR/owl2-overview/.

```
[restriction_rule_1:
     (?v rb:validation on()) -> [restriction_rule_1: (?x rb:violation
     error('ERROR', 'Cardinality Violation', ?y) ) <-
     (?x rdf:type p:RestrictedProperty), noValue(?x p:minCardinality ?y)
]]
```

**Ontology Overview.** The basic policy ontology defines the minimal objects of each policy: domain, goal, target, rule (with an associated action), and trigger. Since each domain (*information management*, *patient domain*, *application systems domain*, *application domain* etc.) has its specific requirements and additional *rdf:Property* elements, sub-policy ontologies are created. The sub-ontologies import the basic ontology and inherit all properties from the basic policy.

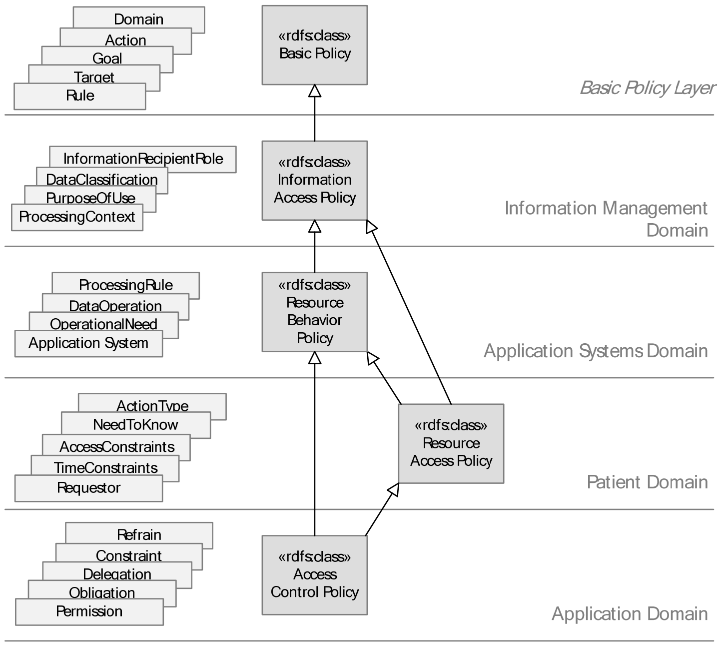


**Fig. 3.** Structure of ontologies

**Definition of Conflicts.** Each identified conflict has to be defined accordingly. For reasons of space a sample definition of a modality conflict between a resource access policy and an authorization policy is given below.

```
function modality-conflict-1 (document rap, document authzp)
   get authorized action and consent type from rap
   get configured action and rules from authp
   while authzp has more rules
      input the next rule
      get authorization decision from rule
      If consent type equals "grant access" and
         authorization decision equals "permit"
         If configured action is not a subclass of authorized action
            print "Policies do not grant access equally.";
         endif
      endif
   endwhile
end function
```

Furthermore, conflicts are defined as rules which can be applied independently to each policy type and processed by a rules engine or reasoned that ensure the proper use of the resources defined in it. Such rules can be written e.g. in RDF/N3. The sample below selects attributes from two given policies and states that the associated actions have to be in the same hierarchy.

```
{
   ?p1 a rap:ResourceAccessPolicy.
   ?p1 basic:pAction ?a1.
   ?p1 rap:consent-type "grant access".

   ?p2 a acp:AuthorizationPolicy.
   ?p2 basic:rule ?r.

   ?r acp:AuthzDecision ?d.
   ?r basic:pAction ?a2
}  => { ?a2 a ?a1 }.
```

## 4 A Methodology for Policy Conflict Detection

The detection of policy conflicts depends whether all relevant policy information is present. So the first step is to gather all information with regard to information management, need-to-know, applications/application systems and feasible patient privacy constraints. Supposing that a proper work organization with a user and identification management is established, such effort has to be done only once initially. The following steps might be useful to detect incorrect policy statements.

1. *Represent written and existing policies as ontology.* Tailor the given policy ontologies to the existing hospital environment. Ideally, the LDAP-based user directory can be imported or queried directly [12]. The description model uses the proposed privacy-friendly recommendation by the national data protection officers

for the design and operation of hospital information systems in Germany [13]. This can be used as a basis.

2. *Instantiate concrete policies of the ontologies.* Existing regulations must then be formalized by instantiating policies.
3. *Apply rules independently to concrete policy instances.* In order to indicate absence of conflicts rules are applied. Thus, emerging conflicting policies may be refined accordingly.
4. *Deploy policy instances to application systems.* Valid policy instances are transferred into policy languages that are supported by the application systems such as XACML [14]. Moreover, the representation of privacy consents as Clinical Document Architecture (CDA) Release 2 (R2) documents in accordance with the upcoming normative Health Level Seven (HL7) standard[3] is advised.

## 5 Related work

Policy conflict handling is investigated comprehensively. For instance, Kempter et al. [15] propose the use of models to support conflict handling. They map invariants (rules/dependencies from a managed system) to policy actions. Conflict definitions are derived from the invariants. Conversely, they do not look at semantic conflicts as they occur in health care related policies environments.

In addition, Aphale et al. [16] give guidance for logical and functional conflict identification and resolution. Their work is based on activities which are ranked through a prioritization model by means of heuristic mechanisms in order to achieve an individual or organizational goal. Logical conflicts refer to the before mentioned conflicts of modality whereas functional conflicts describe inconsistent pre-conditions between actions and a goal of the organization (i.e., need-to-know). However, the developed agent assistant based on OWL 2.0 focusses on goals of an organization and does not regard different policy concerns such as patient constraints.

The detection of conflicts across different policy concerns is investigated in [17]. This is a similar approach since conflicts are based on a domain description model and class-specific policy models. Conflicts are expressed as rules. The approach in this work differs because it considers the information management in a top-down manner and captures the specifics of each policy class. Conflicts in [17] are not classified but referred to possible impacts or threats which should be embedded in a security risk management point of view.

## 6 Summary and Future Work

This paper proposed the use of a semantic model for capturing the specifics of relevant policies in a health care environment (especially in hospitals). It is useful to detect policy conflicts when authoring new policy statements. Different policy concerns are represented as dedicated policy types via separate ontologies. These

[3] See http://www.hl7.org/implement/standards/product_brief.cfm?product_id=280.

types direct the overall access to information in hospitals. Conflicts between policy types are expressed as rules.

Future work will include policy conflict handling that is based on the defined ontologies. Moreover, prototypical tool support for policy authoring is intended. Ideally, (valid) policies are stored in a repository which serves as a single point of access when new organizational requirements should be translated into guidelines and rules in IT management.

## References

[1] L. Aversano, C. Grasso, and M. Tortorella, "A Literature Review of Business/IT Alignment Strategies," *Procedia Technology*, vol. 5, pp. 462–474, 2012.

[2] J. Caumanns, R. Kuhlisch, O. Pfaff, and O. Rode, "IHE IT-Infrastructure White Paper: Access Control," IHE International, Sep. 2009.

[3] R. Wies, "Policy Definition and Classification: Aspects, Criteria, and Examples," in *Proceedings of the IFIP/IEEE International Workshop on Distributed Systems: Operations & Management,*, Toulouse, France, 1994, pp. 10–12.

[4] D. F. Sterne, "On the buzzword ‘`security policy’’," in *Proc. 1991 IEEE Symposium on Research in Security and Privacy*, Oakland CA, 1991, pp. 219–230.

[5] J. R. Vacca, *Computer and Information Security Handbook*. Morgan Kaufmann, 2009.

[6] International Organization for Standardization, "Health informatics – Privilege management and access control – Part 2: Formal models," ISO/TS 22600-2:2006(E), Aug. 2006.

[7] A. Westerinen, J. Schnizlein, J. Strassner, M. Scherling, B. Quinn, S. Herzog, A. Huynh, M. Carlson, J. Perry, and S. Waldbusser, "Terminology for Policy-Based Management," Internet Engineering Task Force, RFC 3198, Nov. 2001.

[8] J. D. Moffett and M. S. Sloman, "Policy Conflict Analysis in Distributed System Management," *Journal of Organizational Computing*, vol. 4, no. 1, pp. 1–22, 1994.

[9] D. Daiqin He and J. Yang, "Authorization Control in Collaborative Healthcare Systems," *Journal of Theoretical and Applied Electronic Commerce Research*, vol. 4, no. 2, pp. 88–109, Aug. 2009.

[10] J. Krogstie, *Model-Based Development and Evolution of Information Systems - A Quality Approach*. Springer London, 2012.

[11] B. Thalheim, "The Concepton of the Model," in *Business Information Systems: 16th International Conference, BIS 2013, Poznań, Poland, June 2013, Proceedings*, W. Abramowicz, Ed. Springer, 2013, pp. 113–124.

[12] W. Yung, "Bring existing data to the Semantic Web: Expose LDAP directories to the Semantic Web with SquirrelRDF," *IBM developerWorks*, 01-May-2007. [Online]. Available: http://www.ibm.com/developerworks/xml/library/x-semweb/index.html. [Accessed: 06-Jul-2013].

[13] Conference of the Data Protection Commissioners of the Federation and the Federal Länder, "Orientierungshilfe Krankenhausinformationssysteme," in *Datenschutzkonforme Gestaltung und Nutzung von Krankenhausinformationssystemen*, Würzburg, 2011.

[14] T. Moses, “eXtensible Access Control Markup Language (XACML),” OASIS, oasis-access_control-xacml-2.0-core-spec-os, Feb. 2005.

[15] B. Kempter and V. Danciu, “Generic Policy Conflict Handling Using a priori Models,” in *Ambient Networks, 16th IFIP/IEEE International Workshop on Distributed Systems: Operations and Management, DSOM 2005, Barcelona, Spain, October 24-26, 2005, Proceedings*, vol. 3775, J. Schönwälder and J. Serrat, Eds. Springer Berlin Heidelberg, 2005, pp. 84–96.

[16] M. S. Aphale, T. J. Norman, and M. Şensoy, “Goal-Directed Policy Conflict Detection and Prioritisation,” in *Coordination, Organizations, Institutions, and Norms in Agent Systems VIII*, H. Aldewereld and J. S. Sichman, Eds. Springer Berlin Heidelberg, 2013, pp. 87–104.

[17] M. M. Casalino, H. Plate, and S. Trabelsi, “Transversal Policy Conflict Detection,” in *Engineering Secure Software and Systems*, G. Barthe, B. Livshits, and R. Scandariato, Eds. Springer Berlin Heidelberg, 2012, pp. 30–37.

# Towards Methodology for Design of Context-Aware Decision Support Systems based on Knowledge Fusion Patterns

Alexander Smirnov and Tatiana Levashova

St. Petersburg Institute for Informatics and Automation
of the Russian Academy of Sciences,
39, 14th line, St. Petersburg, 199178, Russian Federation
{smir, tatiana.levashova}iias.spb.su

**Abstract.** A pattern-based methodology for design of context-aware decision support systems is proposed. The methodology is based on knowledge fusion patterns for the knowledge fusion processes occurring at different stages of a context-aware decision support system. The methodology as it stands focuses on the context-aware stages of the system. For the knowledge fusion processes ongoing at these stages the patterns' elements relevant to the methodology are specified. Usage of the patterns to propose system functionality depending on the user needs is demonstrated.

**Keywords:** context aware decision support, ontology-based context, knowledge fusion, knowledge fusion patterns, pattern-based methodology.

## 1 Introduction

Various methodologies for information system development [1], [2], [3], [4], [5] consider information requirements as an important element of the system design. The methodologies aim is to propose the system functionality suitable to the user.

Originally, the research presented in this paper was devoted to revealing of knowledge fusion patterns for the processes occurring in a context-aware decision support system (DSS) [6], [7]. The importance of knowledge fusion in context-aware DSSs is reasoned by the intention of this technology. The objective of knowledge fusion is to integrate information and knowledge from multiple sources into some common knowledge that may be used for decision making and problem solving or may provide a better insight and understanding of the situation under consideration [8], [9], [10], i.e. knowledge fusion facilitates situational awareness.

The revealed patterns describe the knowledge fusion effects discovered in the DSS, specify states of the knowledge sources involved in knowledge fusion, and lift the effects at the ontology level. The main contribution of the present work is a pattern-based specification of the DSS' requirements to the knowledge sources involved in the processes ongoing in this DSS. The specification is intended to describe the system functionality.

Although the patterns can be used to describe the DSS' requirements ensuring the overall functionality of the system, this paper covers only context-aware phase of the DSS. It is motivated by the ongoing research; the methodology has not been developed to the full extent. It is believed that the ideas presented in this paper provide a general conception according to that the methodology will be developed.

The rest of the paper is as follows. Section 2 gives an overview of possible knowledge fusion effects and introduces the elements of the pattern description language. Section 3 describes the conceptual framework that the DSS is based upon, the stages of the DSS scenario, and presents the knowledge fusion effects produced at these stages. The pattern-based methodology is presented in Section 4. The main results and the future research are discussed in the Conclusion.

## 2 Knowledge Fusion Patterns

In this Section knowledge fusion patterns are revealed for the knowledge fusion processes occurring in the context-aware DSS. Knowledge fusion effects found in this DSS indicate presence of such processes.

### 2.1 Knowledge Fusion Effects

The main feature of knowledge fusion lies in creation of a synergic effect from the integration of information/knowledge. Based on the analysis of knowledge fusion studies, a number of knowledge fusion processes producing different effects have been distinguished:

- Intelligent fusion of huge amounts of heterogeneous data / information from a wide range of distributed sources into a form which may be used by systems and humans as the foundation for problem solving and decision making [8], [9]. The intelligence assumes consideration of the semantic contents of the sources being fused.
- Integration of knowledge from various knowledge sources resulting in a completely different type of knowledge or new idea how to solve the problem [11], [12], or integration of different types of knowledge (domain, procedural, derived, presentation, etc.) resulting in a new knowledge type [10].
- Combining knowledge from different autonomous knowledge sources in different ways in different scenarios, which results in discovery of new relations between the knowledge from different sources or/and between the entities this knowledge represents [13], [14].
- Integration of multiple knowledge sources into a new knowledge object, which is a new knowledge source [15], [16].
- Inference of explicit knowledge from information/knowledge hidden in knowledge sources being integrated or fused [17].
- Re-configuration of knowledge sources to achieve a new configuration with new capabilities or competencies or knowledge exchange to improve capabilities or competencies through learning, interactions, discussions, and practices [18].
- Involving knowledge from various sources in problem solving, which results in a solution [6].

From the analysis above it is noticed that different processes can produce the same effect, and different effects may be outcomes of the same process. The following not mutually exclusive kinds of new knowledge produced as the knowledge fusion effects are distinguished: 1) a new type of knowledge; 2) a new knowledge source; 3) a new knowledge created from data/information; 4) a new knowledge about the conceptual scheme (new relations, concepts, properties, etc.); 5) a new explicit knowledge; 6) new capabilities/competencies of a knowledge object (an object that produces or contains knowledge); 7) a new problem solving method or idea how to solve the problem; 8) a solution for the problem.

Almost all the listed effects have been found in the DSS. An exception concerns appearance of new ideas how to solve the problem. Such ideas may come as a result of conscious interactions, discussions, and practices. These issues are not considered in the research. Below, dimensions proposed to the generalization of the knowledge fusion processes and elements of a pattern specification language are discussed.

### 2.2 Pattern Language

Knowledge fusion involves multiple sources in the integration processes. Autonomies and structures of such sources have been chosen as the concepts in terms of which knowledge fusion patterns are revealed. In the context-aware systems integration of information/knowledge refers to the process of integration of their conceptual structures. Therefore, source's structure is an obligatory concept taken into account by the integration. In this research, by sources' structures the conceptual structures that represent the knowledge in the knowledge sources are meant.

Autonomy creates awareness of the reliability of the information/knowledge represented in the sources. The consideration of the knowledge source autonomy concept is reduced to the detection of relations existing between knowledge sources regardless of their structures. Autonomous knowledge source is an independent source having no relationships with other sources. Such source can get changed at any time, at that, the changes in this source produce no changes in other sources. On the contrary, non-autonomous source is linked to other (non-autonomous) sources. Changes in a non-autonomous source produce appropriate changes in the related sources.

The DSS operates in a dynamic environment. Information and knowledge represented in environmental sources that are related to the internal system sources (environmental and system sources are non-autonomous) are considered to be more reliable than information/knowledge represented in the autonomous environmental sources. An argument in favor of this is any changes in the linked (non-autonomous) environmental sources are reflected in the system sources.

The patterns measure knowledge fusion outcomes in terms of preservation/change of the structures and autonomies of the initial and target knowledge sources, and in terms of the effects the knowledge fusion processes produce in the DSS.

Initial knowledge sources are the sources that are integrated leading to the emergence of a new knowledge (producing some knowledge fusion effect). The sources resulting from the knowledge fusion or enclosing the knowledge fusion result are referred to as target knowledge sources. The environmental sources (below, resources) include sources of data/information/knowledge. In this sense the environmental resources belong to the collection of knowledge sources.

The knowledge fusion patterns are described using a pattern description language [7]. The detailed presentation of the patterns elements is as follows:

**Name**: a name to refer to the pattern
**Problem**: a problem the knowledge fusion process solves
**Solution**: a meaningful description of the knowledge fusion process
**Initial knowledge source(s)**: knowledge sources(s) that are integrated leading to producing some knowledge fusion effect
**Target knowledge source(s)**: knowledge sources(s) resulting from the knowledge fusion or enclosing the knowledge fusion result
**Related pattern** (may be omitted): an alternative pattern that can be used instead of the described one or in parallel or after termination of the described pattern
**Exception** (may be omitted)**:** a description of cases when the pattern is not applicable
**Autonomy pre-states**: the degree of autonomy of knowledge sources before the knowledge fusion process. Three degrees are provided for: autonomous, non-autonomous, and n/a (for a non-existing knowledge source)
**Effect in DSS**: the effect the knowledge fusion process produces in the DSS
**Effect in ontology terms**: ontology-based generalization of the effect produced
**Post-states**: the knowledge source autonomy and structures preservation degrees after the knowledge fusion process completes. For the knowledge source autonomies the degrees introduced in pre-state descriptions are kept on. Three degrees of knowledge object structure preservations are provided for: preserved, changed, and new (for a new knowledge object)
**Schematic representation**: the knowledge fusion process represented schematically
**DSS stage**: the stage of the DSS scenario where the knowledge fusion process occurs.

In this work one of the possible pattern applications, which is the pattern-based methodology for design of context-aware DSSs, is offered. Prior to present this methodology, the conceptual framework the DSS is based upon and the context-aware stages of the DSS scenario are described.

## 3 Context-Aware Decision Support System

The DSS is intended for support of decisions on planning emergency response actions. A two-level representation of emergency situation is used in the DSS. At the first level the situation is represented by *abstract context* that is an ontology-based intensional model of the situation. At the second level the emergency situation is represented by *operational context* that is an instantiation of the abstract context for the actual circumstances.

Environmental resources produce the operational context and solve the problem of planning emergency response actions based on this context. For this the resources organize a resource network. Nodes of this network represent the resources; network arcs signify the order of the nodes execution.

The problem solution is a set of alternative emergency response actions feasible in the current emergency situation. The decision maker chooses an alternative from the set of feasible ones. The chosen alternative is considered as the decision. According to this decision the response actions are undertaken.

Once the interactions of the decision maker with the DSS have been finished, the abstract context, the operational context, the decision, and the resources' representations are saved in a context archive. At that, the operational context and the resources' representations are saved in their states at the instant of the alternatives generation.

The DSS scenario follows two main phases: preliminary and executive. At the preliminary phase an application ontology (AO), which describes knowledge of the emergency management domain, is built. This ontology specifies knowledge to describe the emergency situations happening in this domain along with problems requiring solutions in these situations. The AO is a knowledge source fusing two types of knowledge: domain and problem-solving. The executive phase concerns support of the decision maker with alternative decisions, decision implementation, and archiving. The executive phase is the focus of this paper since at this phase context-aware functions of the DSS come into operation. Several stages are distinguished at this phase. At each of them one or more knowledge fusion effects manifest.

## 3.1 Context-Aware Stages of DSS

This Section focuses on the knowledge fusion processes going on at the context-aware stages of the DSS and the knowledge fusion effects produced at these stages. The processes are generalized in the patterns terms. In the paper only pattern elements that are relevant to the discussion of the proposed methodology are presented. They are the states for the autonomies and structures of the knowledge sources.

### 3.1.1 Abstract Context Creation

The abstract context represents knowledge relevant for decision making in the emergency situation. The represented knowledge is captured from the AO based on the type of emergency event. A (smart) sensor reads the type of event and sends it to the DSS or the user enters it in the system.

The knowledge fusion effect produced at the stage of abstract context creation is a new knowledge source of the same type as the initial knowledge source (the AO). The AO preserves its structure and autonomy; the abstract context becomes an autonomous knowledge source with a proper structure. The processes going on at the stage of abstract context refinement are generalized by the *simple fusion* pattern (Table 1).

**Table 1.** Simple fusion

| Pattern element | Initial knowledge source | Target knowledge source |
|---|---|---|
| Knowledge source | application ontology | abstract context |
| Autonomy pre-state | autonomous | n/a |
| Structure post-state | preserved | new |
| Autonomy post-state | autonomous | autonomous |

### 3.1.2 Abstract Context Refinement

In the abstract context the captured knowledge may result in discovery of new relationships between the knowledge unrelated in the AO. These relationships are the result of deductive inference. Generally, any kind of knowledge representation items can be inferred. The inferred items are considered as a knowledge fusion effect that is the new knowledge. This knowledge is introduced in the abstract context, hereby changing its structure. In the case of the abstract context refinement the abstract context plays the roles of the initial and target knowledge source at the same time. The processes taking place at the stage of the abstract context refinement are generalized by the *extension* pattern (Table 2).

**Table 2.** Extension

| Pattern element | Initial knowledge source | Target knowledge source |
|---|---|---|
| Knowledge source | abstract context | abstract context |
| Autonomy pre-state | autonomous | autonomous |
| Structure post-state | changed | changed |
| Autonomy post-state | autonomous | autonomous |

#### 3.1.3 Abstract Context Reuse

The abstract contexts are reusable components of the DSS. The reuse of an abstract context in settings when the available resources are not intended to solve the problems specified in this context may result in finding alternative resources. For instance, one unavailable method can be substituted for a sequence of methods providing by the available resources. This leads to a new configuration of the resource network.

At the stage of abstract context reuse the knowledge fusion effect is twofold: a new (alternative) problem solving method and a new configuration of the resource network. A specification of the new method(s) is introduced in the abstract context. The context structure is changed, at that this context remains autonomous. The autonomies and structures of the resources representing the new method(s) are preserved.

In the case of abstract context reuse the knowledge fusion pattern is nested. The main pattern "*configured fusion*" (Table 3) includes the *extension* pattern. The *extension* pattern corresponds to introducing the new specification in the context.

**Table 3.** Configured fusion (the main pattern)

| Pattern element | Initial knowledge source | Target knowledge source |
|---|---|---|
| Knowledge source | abstract context | resource network |
| Autonomy pre-state | autonomous | autonomous |
| Structure post-state | changed | preserved |
| Autonomy post-state | autonomous | autonomous |

#### 3.1.4 Operational Context Producing

An operational context is produced through the semantic fusion of data/information from multiple environmental resources within the ontological structure of the abstract context. Initially the operational context is a copy of the abstract context. As soon as the resources start instantiating this copy, they lose their autonomies. When the copy is fully instantiated it becomes the operational context. At that, information from the resources is constantly coming into this context. Therefore, the operational context and the environmental resources are related over the period of decision making and implementation. In practice, the operational context represents the map of the area around the emergency event where the situation dynamic is represented (the mobile responders are moving, the traffic situation is changing, etc.).

The knowledge fusion effects had at the stage of the operational context producing are 1) the operational context is *a new knowledge source*; 2) this context is *a knowledge source created from data/information*; and 3) the operational context represents *knowledge of a new* dynamic *type*. The abstract context preserves its structure and autonomy when the operational context is produced. The operational context is a new non-autonomous knowledge source. The *instantiated fusion* pattern (Table 4) generalizes the processes going on at the stage of operational context producing.

**Table 4.** Instantiated fusion

| Pattern element | Initial knowledge source | Target knowledge source |
|---|---|---|
| Knowledge source | abstract context | operational context |
| Autonomy pre-state | autonomous | n/a |
| Structure post-state | preserved | new |
| Autonomy post-state | autonomous | non-autonomous |

#### 3.1.5 Problem Solving

As it is said above, the result of problem solving is a set of feasible emergency response plans. An emergency response plan is a set of emergency responders with required helping services, schedules for the responders' activities, and transportation routes for the mobile responders. In the plans earlier independent entities become related, i.e. new relations between these entities have arisen.

The plans are represented in the picture of the operational context. In this way, the operational context and the results of problem solving are fused forming, at that, a new knowledge source. This new source represents knowledge of a new type (the instantiated knowledge fused with the solution set).

At the problem solving stage the operational context dissolves within the new knowledge source and does not preserve the structure and autonomy. At time of alternatives generation and decision making the environmental resources and the operational context are related (non-autonomous). As soon as the decision has been made the new knowledge source and the environmental resources become autonomous. The knowledge fusion effects produced at this stage are 1) new relations between entities, 2) a problem solution, and 3) a new knowledge source of a new type. Table 5 presents a fragment of the *flat fusion* pattern for the processes at the problem solving stage.

**Table 5.** Flat fusion

| Pattern element | Initial knowledge source | Target knowledge source |
|---|---|---|
| Knowledge source | operational context | knowledge source fusing the operational context and the set of alternatives |
| Autonomy pre-state | non-autonomous | n/a |
| Structure post-state | changed | n/a |
| Autonomy post-state | n/a | autonomous |

#### 3.1.6 Decision Implementation

The decision is a solution that the decision maker has chosen from the set of alternative ones. This decision is made at a certain time instant. The situation may change from the moment the decision was made to the moment of its implementation. The responders whom the decision is delivered may be unable to implement it in the changed circumstances. In some cases, the activities assigned to the responders who become unable to operate can be delegated to or redistributed between other responders participating in the decision implementation. As a result of this, the responders that are ready to take the assignments gain new capabilities / competencies.

For instance, an emergency team trained to rescue operations has failed in the course of actions because of a road destruction, ambulance blockage, etc. In certain cases these operations can be delegated to available teams. In the DSS the emergency responders are represented by their profiles. In the case of consent, the plan (the decision) is adjusted accordingly and the profiles of the teams agreed to take part in the rescue operations are extended with the new capability.

At the time of the decision implementation, the responders taking part in the response plan are not autonomous. Moreover, in the course of the response actions the structures of their profiles as well as the decision structure may change. The changed decision structure results in changing the structure of the knowledge source containing the set of solutions. This knowledge source is not autonomous until the decision is implemented. The knowledge fusion effect produced at the decision implementation stage consists in gaining new capabilities / competencies by the emergency responders. The *adaptation* pattern (Table 6) generalizes this case.

**Table 6.** Adaptation

| Pattern element | Initial knowledge sources | Target knowledge sources |
|---|---|---|
| Knowledge source | • knowledge source representing the decision<br>• profiles of the emergency responders | |
| Autonomy pre-state | non-autonomous | non-autonomous |
| Structure post-state | changed | changed |
| Autonomy post-state | non-autonomous | non-autonomous |

### 3.1.7 Archival Knowledge Management

The stage of archival context management deals with the management of knowledge contained in the archived components. The main intention of such management is inference of new knowledge based on the accumulated one.

For example, an emergency team participated in different emergency response actions. Some operational contexts in which this team appeared and then participated in corresponding actions represent the same hospital. Based on a comparative analysis of these operational contexts it can be judged that most probably the team is a part of the hospital found together with this team in different contexts. The *part-of* relation between the emergency team and hospital is the new revealed relation. The revealing of a new knowledge based on a set of observations is a kind of inductive inference.

In the archive, the operational contexts, the representations of environmental resources, the responders' profiles, and the knowledge object representing the decision are related. I.e., in the archive all the listed knowledge sources are non-autonomous. As a result of archival knowledge management, new knowledge about conceptual schemes of the archived knowledge sources can be inferred. This new knowledge gets specified in the AO. As a result of this, the structure of the AO gets changed but its autonomy is preserved. The knowledge fusion effect produced is a new knowledge about the conceptual scheme. Table 7 shows a fragment of the *historical fusion* pattern for the processes occurring at the stage of archival knowledge management.

**Table 7.** Historical fusion

| Pattern element | Initial knowledge sources | Target knowledge source |
|---|---|---|
| Knowledge source | operational contexts | application ontology |
| Autonomy pre-state | non-autonomous | autonomous |
| Structure post-state | preserved | changed |
| Autonomy post-state | non-autonomous | autonomous |

## 4 Pattern-Based Methodology

The offered methodology follows four steps: 1) specification of the system information requirements; 2) capturing the user requirements; 3) matching the system requirements against the user requirements; 4) finding available system functionality.

The proposed patterns enable to formulate the system requirements in terms of the patterns' inputs/outputs (Table 8). Further on in this paper only the patterns' inputs are taken into account to formulate the requirements. The patterns' inputs and outputs jointly are planned to be used to track the information flows across the DSS' stages and between the patterns. These flows manifest interrelationships between the stages and thereupon allow ones to specify explicitly in the methodology what output scenario components can serve as the input at what stage.

The system requirements (Table 8) are formulated for the particular DSS. The patterns' parameters are presented in the way they are used in this DSS. These parameters generalized are presented in Table 9 (the columns "System requirements"). These columns formalize possible parameters' values for general cases and formulate general conditions for the patterns applicability. In the table the following notation is used: “a” – autonomous knowledge objects, “na” – non-autonomous, “m” – modifiable, “nm” – not-modifiable, “/” – logical OR, “&” – logical AND. The resource network is considered as a single knowledge object. Modifiable resource network means that this network is reconfigurable, at that any changes inside the network nodes are not supposed (the structures of the resource organizing the network are not changed).

User requirements are demonstrated by an example. For instance, the user does not possess any AO. Though, he/she has an unalterable abstract context representing the abstract situation this user usually deals with and a set of resources authorized for his/her needs. As well, this user can manage the actors’ profiles as at he/she discretion. The example of the user requirements is presented in the column "User requirements" (Table 9). "Not defined" parameter value means that the user has no specific requirements to the input component. The system designers can manipulate such a component by their choice. Here it is supposed that the system provides the user with the components resulted in the system scenario execution with the parameters required for the overall system functionality.

The system requirements are matched against the user requirements to determine the patterns applicability for this user (the column "Pattern applicability" of Table 9). From Table 9 seen, that the fully applicable patterns are *flat fusion* and *adaptation*.

**Table 8.** DSS requirements

| Stage | Input | Output | Functionality | Pattern |
|---|---|---|---|---|
| Abstract context creation | Autonomous AO | Autonomous abstract context | Creation of a non-instantiated ontology-based model of the situation | Simple fusion |
| Abstract context refinement | Autonomous modifiable abstract context | Autonomous abstract context | Inference of new (contextual) knowledge | Extension |
| Abstract context reuse | • Autonomous modifiable abstract context<br>• Autonomous reconfigurable resource network | • Autonomous abstract context<br>• Autonomous resource network | Reconfiguration of the resource network according to the current circumstances | Configured fusion |
| Operational context producing | • Autonomous abstract context<br>• Resources able to lose their autonomies | • Non-autonomous modifiable operational context<br>• Non-autonomous resource network | Creation of a near real-time picture of the situation | Instantiated fusion |
| Problem solving | • Non-autonomous modifiable operational context<br>• Resources able to lose their autonomies | • New autonomous knowledge source<br>• Autonomous resource network | Providing the decision maker with a set of alternative decisions | Flat fusion |
| Decision implementation | • Non-autonomous modifiable knowledge source representing the decision<br>• Non-autonomous modifiable actors' profiles | | Gaining new capabilities / competencies by actors | Adaptation |
| Archival context management | • Non-autonomous operational contexts<br>• Autonomous modifiable AO | Autonomous AO | Inductive inference of new knowledge | Historical fusion |

The *configured fusion, instantiated fusion,* and *historical fusion* patterns are applicable partly. The applicability of the *configured fusion* and *instantiated fusion* patterns covers management of the resource network. The *historical fusion* pattern allows for some inductive inference. The remaining patterns are inapplicable.

The analysis of the applicable patterns results in the following system functionality. The system can reuse the existing abstract context. As introducing any knowledge into this context is not allowed, this context can be reused if the previously used set of environmental resources is available. At that, any new network configurations are senseless as they cannot be specified in the context. Having the abstract context and the resource network the system can provide the user with an operational context (the dynamic picture of the situation) and a set of feasible decisions in this situation. As the user is able to manage the actors' profiles, the system provides he/her with the ability to manage the decision implementation. The operational contexts produced based on the reusable abstract context can be archived, inductive inference over them can be supported but the inference results cannot be retained.

**Table 9.** Pattern-based requirements

| Pattern | System requirements | | User requirements | Pattern applicability |
|---|---|---|---|---|
| | Input(s) | Parameter value | Parameter value | |
| Simple fusion | AO | a/na/m/nm | unavailable | n/a |
| Extension | Abstract context | a/na&m | nm | n/a |
| Configured fusion | Abstract context | a/na&m | nm | n/a |
| | Resource network | a/na/m/nm | a/na&m | applicable |
| Instantiated fusion | Abstract context | a/na/m/nm | nm | n/a |
| | Resource network | na&m/nm | a/na&m | applicable |
| Flat fusion | Operational context | na&m | not defined = na&m | applicable |
| | Resource network | na&m/nm | a/na&m | applicable |
| Adaptation | Decision | na&m | not defined = na&m | applicable |
| | Actors' profiles | na&m | na/a&m | applicable |
| Historical fusion | Operational contexts | a/na/m | not defined = a/na/m | applicable |
| | AO | a/na&m | unavailable | n/a |

The proposed functionality means that the user can deal with the situation he/she usually deals with, solve repetitive problems with different values for the problems' variables, and manage the decision implementation. For instance, applying to the emergency management domain, the system can support decisions on the emergency situations caused by one and the same type of event. At that, such events are supposed to happen in some area where the system has access to a fixed and the same set of environmental resources.

## 5 Conclusion

The possible knowledge fusion effects were described. These effects were found in the context-aware DSS for the emergency response domain. Knowledge fusion patterns specifying these effects along with the states of knowledge sources involved in the knowledge fusion processes were revealed. One of the possible pattern applications, which is the pattern-based methodology for design of context-aware DSSs, was offered. In the methodology the patterns were used to specify the system requirements ensuring the full system functionality.

So far, the methodology covers the executive phase of the DSS. Some future research is needed to specify the overall set of the system requirements and to describe information/knowledge flows across the DSS' stages and between the patterns. Such a specification will enable to find out dependencies between the system stages and available functionalities, and to specify more accurately the applicability of the knowledge fusion patterns to different user's requirements.

**Acknowledgments.** The research was supported partly by projects funded by grants 11-07-00045, 11-07-00058, 12-07-00298 of the Russian Foundation for Basic Research, the project 213 of the research program "Information, control, and intelligent technologies & systems" of the Russian Academy of Sciences (RAS), and the project 2.2 of the Nano- & Information Technologies Branch of RAS.

## References


1. Sage, A.P.: Decision Support Systems Engineering. John Wiley & Sons, New York (1991)
2. Vessey, I., Conger, S.: Learning to Specify Information Requirements: The Relationship between Application and Methodology. JMIS, 10 (2) 177–201 (1993)
3. Elliott G., Strachan, J.: Global Business Information Technology: an Integrated Systems Approach. Addison Wesley, Harlow (2004)
4. MDA – the Architecture of Choice for a Changing World, OMG Model Driven Architecture, http://www.omg.org/mda/ (access date: 05.07.2013)
5. Shahbazian, E., Pigeon, L., Krtaft, J., Bossé, É.: Building Technology-Enabled Decision Support Applications. In: Bossé, É., Shahbazian, E., Rogova G. (eds.) Prediction and Recognition of Privacy Efforts Using Collaborative Human-Centric Information Systems. IOS Press, Amsterdam, pp. 98–108 (2013)
6. Smirnov, A., Pashkin, M., Chilov, N. Levashova T., Haritatos, F.: Knowledge Source Network Configuration Approach to Knowledge Logistics. Int. J. Gen. Syst. 32 (3) 251–269 (2003)
7. Smirnov, A., Levashova T., Shilov, N.: Context-Based Knowledge Fusion Patterns in Decision Support System for Emergency Response. In: Comes, T., Fiedrich, F., Fortier S., Geldermann J., Műller, T. (eds.) 10th International ISCRAM Conference, Baden-Baden, Germany, pp. 597–606 (2013)
8. Alun, P., Hui, K., Gray, A., Marti, P., Bench-Capon, T., Cui, Z., Jones, D.: Kraft: an Agent Architecture for Knowledge Fusion. Int. J. Coop. Inf. Syst. 10 (1–2) 171–195 (2001)
9. Scherl R., Ulery D.L.: Technologies for Army Knowledge Fusion. Final report no. ARL-TR-3279, Computer Science Department, Monmouth University, Monmouth (2004)
10. Holsapple C.W., Whinston A.B.: Building Blocks for Decision Support Systems. In: Ariav G., Clifford J. (eds.) New Directions for Database Systems. Ablex Publishing Corp., Norwood, pp. 66–86 (1986)
11. Lee, K.-R.: Patterns and Processes of Contemporary Technology Fusion: The Case of Intelligent Robots. Asian J. Technol. Innov. 15 (2) 45–65 (2007)
12. Grebla, H.A., Cenan C.O., Stanca, L.: Knowledge Fusion in Academic Networks. BRAIN: Broad Res. Artif. Intell. Neurosci. 1 (2) http://www.edusoft.ro/brain/index.php/brain/article/download/60/145 (2010)
13. Laskey K.B., Costa P., Janssen T.: Probabilistic Ontologies for Knowledge Fusion. In: 11th IEEE International Conference on Information Fusion, http://ieeexplore.ieee.org/xpl/-freeabs_all.jsp?arnumber=4632375 (2008)
14. Jonquet, C., LePendu, P., Falconer, S., Coulet, A., Noy, N.F., Musen, M.A., Shah, N.H.: NCBO Resource Index: Ontology-Based Search and Mining of Biomedical Resources. J. Web Semant. 9 (3) 316–324 (2011)
15. Kuo, T.-T., Tseng, S.-S., Lin, Y.-T.: Ontology-Based Knowledge Fusion Framework Using Graph Partitioning. In: Chung, P.W.H., Hinde, C.J., Ali M. (eds.) IEA/AIE 2003. LNAI, vol. 2718, pp. 11–20 (2003)
16. Gou J., Yang J., Chen Q.: Evolution and Evaluation in Knowledge Fusion System. In: Mira J., Alvarez, J.R. (eds.) IWINAC 2005. LNCS, vol. 3562, pp. 192–201 (2005)
17. Roemer, M.J., Kacprzynski, G.J., Orsagh, R.F.: Assessment of Data and Knowledge Fusion Strategies for Prognostics and Health Management. In: 2001 IEEE Aerospace Conference, vol. 6, pp. 2979–2988 (2001)
18. Lin, L.Y., Lo, Y.J.: Knowledge Creation and Cooperation Between Cross-Nation R&D Institutes. Int. J. Electron. Bus. Manag. 8 (1) 9–19 (2010)

# Selecting Content Ontology Design Patterns for Ontology Quality Improvement

Birger Lantow[1], Kurt Sandkuhl[1,2], and Vladimir Tarasov[2]

[1] Institute of Computer Science, University of Rostock
Albert-Einstein-Str. 22, 18059 Rostock, Germany
{birger.lantow, kurt.sandkuhl} @uni-rostock.de

[2] School of Engineering at Jönköping University
P.O. Box 1026, 551 11 Jönköping, Sweden
tavl@jth.hj.se

**Abstract.** While Ontology Design Patterns (ODPs) are intended to support the process of ontology engineering, they can also be used in order to improve existing ontologies. This paper describes strategies for the selection of candidate ODPs that qualify for the improvement of a given ontology. Starting point is the ExpertFinder ontology that allows for competency description of researchers. ODP selection strategies are performed on content ODPs from the ODP wiki-portal that was initiated by the NeOn-project. Lessons learned and problems faced are discussed, and possible future developments are mapped out. The contributions of this paper are (1) a strategy for selecting ODP suitable for improving a given ontology, (2) experiences from using this strategy for selecting ODP, (3) recommendations for a better support of ODP selection.

**Keywords:** semantic web, ontology design patterns, ontology alignment, linked data, ontology engineering

## 1 Introduction

Work presented in this paper combines techniques from ontology engineering with experiences from quality management. Quality is considered an essential factor for acceptance of technologies and solutions in many disciplines. Furthermore, quality also contributes to efficiency of work and operation processes and to robustness and usability of products. Due to the growing use of ontologies in knowledge-based systems for industrial and administrative applications, standards, procedures and practices for quality improvement of ontology construction processes and the artifacts produced during these processes gain of importance. Although considerable efforts have been spent on developing ontology assessment and evaluation approaches, including ways to measure quality and techniques to improve it (cf. Section 2.2), generally accepted practices for industrial are still missing.

The objective of this paper is to contribute to quality ontologies by focusing on the use of ontology design patterns for improving the quality of existing ontologies.

Ontology design patterns (ODP) have been proposed as encodings of best practices (cf. Section 2.1) supporting ontology construction by facilitating reuse of proven solution principles. In this paper, focus is specifically on Content ODP and on investigating feasibility and utility of incorporating them into ontologies. Our working hypothesis is that Content ODP which have proven to be useful in constructing quality ontologies also can be supportive in improving ontologies. The intention is to gather experience how to best apply Content ODP for this purpose.

The contributions of this paper are (1) a strategy for selecting ODP suitable for improving a given ontology, (2) experiences from using this strategy for selecting ODP, (3) recommendations for a better support of ODP selection.

The remaining part of the paper is structured as follows: Section 2 gives a brief overview to ontology design patterns and approaches for quality improvement. Section 3 describes a strategy for ODP selection in order to improve existing ontologies. The ExpertFinder ontology is introduced as a case study in section 4, while section 5 applies the suggested ODP selection strategy based on the ExpertFinder ontology. The final section 6 summarizes the experiences and gives recommendations for a better support of ODP selection and usability.

## 2 Background on ODPs

Relevant background for this paper includes ontology design patterns (section 2.1) and approaches for quality assurance of ontologies by use of ODPs (section 2.2).

### 2.1 Ontology Design Patterns

In a computer science context, ontologies usually are defined as explicit specifications of a shared conceptualization [10]. Due to the increasing use of ontologies in industrial applications at larger scale, ontology construction and ontology evaluation have become a major area of ontology engineering. The aim is to efficiently produce high quality ontologies as a basis for knowledge management, semantic web applications or enterprise systems. Despite quite a few well-defined ontology construction methods and a number of reusable ontologies offered on the Internet, efficient ontology development continues to be a challenge, since this still requires a lot of experience and knowledge of the underlying logical theory.

Ontology Design Patterns (ODP) are considered a promising contribution to this challenge. In 2005, the term ontology design pattern in its current interpretation was mentioned by Gangemi [2] and introduced by Blomqvist & Sandkuhl [3]. Blomqvist defines the term as “a set of ontological elements, structures or construction principles that solve a clearly defined particular modeling problem“ [4]. Ontology design patterns are described as encodings of best practice, which reduce the need for extensive experience when developing ontologies. Using ODPs, less experienced engineers can apply the well-defined solutions provided in the patterns when creating ontologies.

[5] discusses different types of ODP under investigation with their differences and the terminology used. The two types of ODP probably receiving most attention are logical and content ODP. Logical ODP focus only on the logical structure of the representation, i.e. this pattern type is targeting aspects of language expressivity, common problems and misconceptions. Content ODP offer actual modeling solutions within an application domain and are often instantiations of logical ODP. Due to the fact that these solutions contain actual classes, properties, and axioms, content ODP are considered by many researchers as tailor-made for a specific domain, even though the domain might focus on general issues like ‘events’ or ‘situations’. This paper has its focus on the use of content ODPs. Platforms offering ODP currently include the ODP wiki portal initiated by the NeOn-project[1] and the logical ODPs maintained by the University of Manchester.

### 2.2 Quality assurance of ontologies and ODP

Work in the area of quality assurance for ontologies includes different perspectives, such as the quality of the ontology as such, the quality of the process of ontology construction, and tools supporting the ontology engineer in achieving high quality. In the context of this paper, the focus is on quality of the ontology as such. Quality assessment of ontologies as such has been subject of many research activities [6], but the quality criteria vary considerably between different approaches and often address structural, logical, and computational aspects of ontologies. Furthermore, metrics originating from software quality evaluation have been investigated [7]. Many of the metrics proposed during last years lack an empirical validation in a large number of cases, i.e. what metrics value can be considered as „good“ or as „bad“ often has not been defined due to an insufficient number of reported applications.

Evaluation of the accuracy of ontology content, i.e. suitability and conformance with the domain to be represented, can be performed using a gold standard. In this context similarity metrics, as proposed for example by [8], are used to measure the deviation from the gold standard. These approaches are criticized for mainly using structural graph similarity and for not taking into account the semantics of class definitions or that different kinds of deviations should be weighted differently. Furthermore, a (single) gold standard often is difficult to develop due to a very limited number of experts available in the domain.

Furthermore, approaches were proposed for evaluating „ontologies in use“, i.e. to evaluate the fitness for a task to be performed with an ontology in a defined scenario. An ontology of high quality "helps the application in question produce good results on the given task" [9]. However, it is difficult to generalize the results from such approaches, since they can hardly capture all aspects potentially relevant.

The general consensus of our work is that ODPs as best practices have an inherent proven quality and that their use in ontologies increases for example readability and thus reusability. Additional support for reusability stems from the expectation that ODP can set quasi-standards.

[1] http://ontologydesignpatterns.org

## 3 Strategies for ODP Selection

Goal of the pattern selection strategies is to efficiently find appropriate patterns that are good candidates for ontology improvement. The final decision should be based on the expertise of the ontology engineer. Thus, the number of choices for the ontology engineer should be minimized/decreased by stepwise filtering the set of ODPs on the base of certain criteria. In the case of automated pre-selection, complexity should also be minimized. In general, this process may also lead to an empty set of ODPs

Prior to the selection of ODPs for ontology improvement, the scope of the improvement process needs to be defined. This influences the applicability of certain filter criteria, as we will see later. In our context, there is a difference between *ontology reengineering* and *ontology restructuring*.

*Ontology reengineering* covers the complete ontology engineering process. This includes the requirements definition. Thus, ODPs serve as best practices for domain specific or general requirement definition. For example, additional competency questions may be defined that result in additional ontology concepts and in additional information stored in the knowledge base.

*Ontology Restructuring* on the other hand just aims at the ontology quality by *refactoring* and does not change the informational requirements. As seen in section 2.3 ontology quality has many aspects. In the case of Ontology Restructuring we see for example computational aspects for ontologies in use (reduction of required storage) and benefits in readability and reusability of ontologies. The latter are important for the process of ontology engineering. In the context of this paper, the notion of "living" ontologies that need to be adapted to changes in the real world and to changes in requirements respectively, implies that these quality aspects are relevant. The measurement of the effects on quality themselves is out of focus of this paper.

By *Ontology Restructuring* new conceptualizations may be introduced, others may become obsolete or are going to be represented differently. However, there must exist a mapping that completely describes the newly structured ontology based on the old structure. With reference to Haslhofer and Klas [10], *Ontology Restructuring* includes at least one of the three activities that must be considered during the ontology transition:

1. Linking to ontologies that add conceptualizations and representation structures, e.g. linking to ODPs.
2. Transformation of ontology structure
3. Instance transformation

The tasks that need to be performed can be relevant for filtering, if effort is considered as a criteria for ODP selection. If restructuring steps 2 and/or 3 are performed the application logic around an application ontology needs to be changed-This can lead to considerable additional effort. Effort is not considered in the suggested approach but can be added as an additional step in the selection process.

The suggested approach for ODP selection includes several stages of filtering:

1. Filter by Domain: While ODPs of the "general" domain should always be considered for ontology improvement, also those of the same or closely related

domains compared to the ontology in focus are relevant.

2. Filter by requirements: This filter should not be applied for ontology reengineering tasks since the goal is to derive new requirements. Those should not be filtered out.
   A common method for ontology requirements definition are competency questions. A filter by requirements would check for similar competency questions in ODPs and in the ontology in focus. As stated in [11] by Noy and McGuiness, competency questions may on the one hand serve for testing the ontology but on the other hand they are a help to roughly describe the scope of a domain and do not need to be exhaustive.
   However, requirements specification may also be done in many different ways. Looking into the documentation provided with the ODPs on the ODP wiki, besides “Competency Questions” we will find “Intent”, “Solution description”, and “Scenarios” as documentation elements that are candidates to provide requirements that are fulfilled by the patterns. However, “Intent” and “Competency Questions” seem to the most appropriate fields for filtering by requirements. “Solution Description” and “Scenarios” in contrast provide help for understanding the used conceptualizations which is useful for the next 3rd step of filtering.

3. Filter by shared conceptualizations: The number of shared conceptualizations should be counted here. A first threshold would be 1. Thus, an ODPs remains in the set of candidates if it shares at least 1 conceptualization with the ontology in focus. The threshold may be increased if necessary. Shared conceptualizations can be identified automatically by comparing the IRI[2]s of the used conceptualizations. This is only possible, if the same representation has been used in ontology and ODP. Another way would be the comparison of used labels, maybe in addition with a synonym data base. However, a manual review may reveal additional conceptualizations that are identical or that overlap but are not represented identically. This filter step is close to the idea of selecting ODPs by their names as described by Hammar et al. in [12]. There ontology engineers selected patterns if the name corresponded to their modeling needs. This is assumed to generally happen on the level of conceptualizations. Therefore the filtering by name at an earlier step than this one does not seem appropriate. This also emphasizes on the need of ODPs to be small enough in order to clearly understand the respective conceptualizations behind them.

4. Filter by compatibility: It should be checked whether all conceptualizations of the remaining patterns are compatible with the ontology in focus. Since only restructuring is intended, there must a transformation rule that populates the classes of pattern based on the current ontology. Structural incompatibilities like abstraction level discrepancies (see for example [11]) may be an obstacle here.

---

[2] Internationalized Resource Identifier

After narrowing down the set of ODPs to compatible patterns that can be used in order to fulfill the ontology requirements, further selection can be done based on the evaluation of expected effort of restructuring and based on the expected ontology improvement. But this is outside the focus of this paper and will not be discussed.

## 4 Case Study: ExpertFinder

The ExpertFinder ontology is the result of an internal research project at Jönköping University. It has undergone several development steps and has also been investigated for the possibilities to foster information reuse and interoperability with ODPs and Linked Data. Results are in [13].

The ontology is the base for an ExpertFinder application that allows to find potential experts among all researchers and teachers of the university. The search is done on competence profiles that are represented in the ExpertFinder ontology. The ontology is implemented in the OWL language.

Reliable information about the researchers and teachers at Jönköping University therefore needs to be maintained and efficiently retrieved into the ontology. This process should be supported by a software system for gathering experts' competencies in different areas, and proposing suitable experts to the user.

Overall, the ontology is in the domains of *Research* and *Teaching*. Since organizational structures (position of an expert) are relevant, the Domain of *Management and Organization* may be added. Due to the scientific focus at Jönköping University, *Computer Science* and *Electrical Engineering* are further domains of the ontology.

Specification is available in the form of competency questions:

1. Finding experts in teaching
   - 1.1. Who can give a quest lecture about ontology applications in medicine in a master's course?
   - 1.2. Who can give lectures on knowledge management in a master's course?
   - 1.3. Who can supervise labs in web programming in a bachelor's course?
   - 1.4. Who can be the course coordinator for the Embedded Systems Architectures course?
   - 1.5. Who can supervise master's theses in information engineering?
2. Finding experts in research
   - 2.1. Who has been doing research projects about sensor networks?
   - 2.2. Who has participated in industrial projects in avionics engineering?
   - 2.3. Who has PhD in computer science and is involved in EU projects?
   - 2.4. Who are the authors of journal papers on distributed databases?
   - 2.5. Who is the expert in ontology engineering?
   - 2.6. What is the expertise of person X?

and in the form of requirement definitions that define the following information about a person as required:

1. Education of the expert including field(s) of education, level and year,
2. University/scientific degrees including area and year,
3. Research areas the expert works/worked in (these should be supported by publications/projects),
4. Research papers written by the expert including research field, publication type, year, length,
5. Research projects the expert participates/participated in including research field(s), project type and start/end date,
6. Courses taught by the expert including field(s) of education, course level, number of credits, type of involvement, how many times was involved, year of the last involvement,
7. Employment type, affiliations ,
8. Contact details,
9. Previous experience

Ontology engineering resulted in the ExpertFinder ontology as shown in figure 1. This is the starting point for the investigation of a possible restructuring based on ODPs.

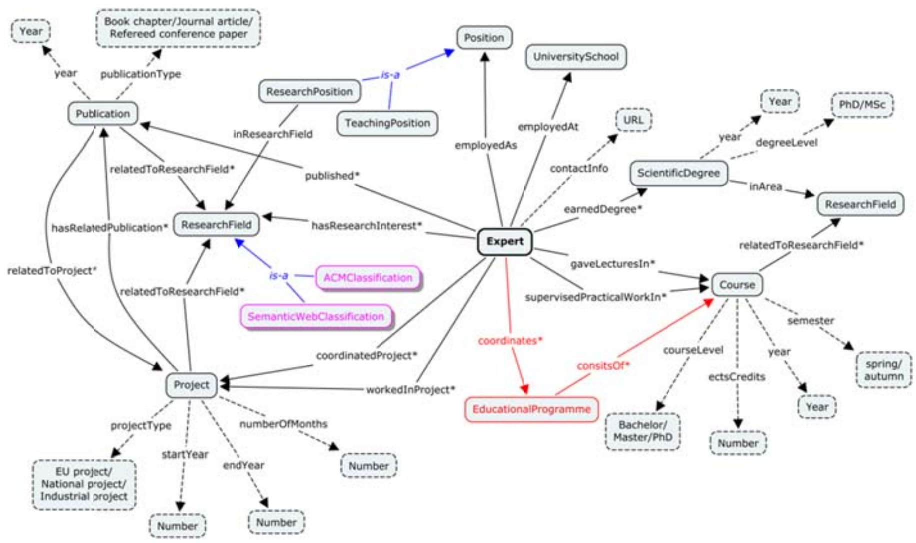


**Fig 1.** The ExpertFinder ontology

## 5 Application of the ODP Selection Strategy

In the following, we describe the filtering process of section 3 on the example of the ExpertFinder ontology that has been introduced in section 4. Occurring problems in the several steps are described. Furthermore, solutions and suggestions for future support of ODP selection are derived.

### 5.1. Filter by Domain

Domain filtering reduced the complete set of 97 Content-ODPs from the ODP wiki down to 72 candidate ODPs. Among the domains of the ExpertFinder ontology and synonymously labeled domains only the “Management” has been found. Thus, ODPs of the domains “Management”, “General”, “Parts and Collections” which could be a subset of general, and ODPs with no given domain remained in the set of candidates. Some of the ODPs with no given domain could have been singled out regarding the domain they actually represent. Therefore, adding the domains here would be an improvement for the possibility of pattern filtering. Furthermore, an automated would be possible if domains and their relations would be more formalized. A taxonomy could be an improvement.

Exemplary decisions ( - = neglected, + = passed):

| Pattern | Domain | Decision | Comment |
|---|---|---|---|
| ClimaticZone | Fishery | - | % |
| Co-Participation | General | + | % |

### 5.2. Filter by requirements

It seemed reasonable to use the high abstraction level of the “intent” description of ODPs in order to do a separate filter step based on “intent”. Filtering was done by answering the question whether or not the pattern intent fits to the purpose of the ExpertFinder ontology. The set of candidates was narrowed down to 41 patterns. No intent was given for 14 of these patterns.

Exemplary decisions ( - = neglected, + = passed):

| Pattern | Intent | Decision | Comment |
|---|---|---|---|
| Communication Event | To model communication events, such as phone calls, e-mails and meetings,…. | - | % |
| Agent Role | To represent agents and the roles they play. | + | Different Roles of Experts described |

The next step was the comparison of the competency questions. This step took more effort per pattern. Each pattern needed to be opened separately and several competency questions needed to be compared. A problem arose when comparing pattern competency questions to the competency questions of ExpertFinder ontology. There was a different abstraction level. Only three patterns qualified - all of them because they are dealing with questions of participation which fits to competency question 2.2. The main reason may be that there were almost no patterns specific to the domains of ExpertFinder ontology, but only of the "General" domain. However, also considering the rest of the ExpertFinder specification in relation to the pattern competency questions a set of 35 candidate patterns remained. No competency questions were given for 18 of them. Again, it seems that better documentation of ODPs would foster their application.

Exemplary decisions ( - = neglected, + = passed):

| Pattern | Competency Questions | Decision | Comment |
|---|---|---|---|
| Types of Entities | What kind of entity is that?<br>Is this an event or an object? | - | General types of ontology elements clear and not in question. |
| Participation | Which objects do participate in this event? | + | Competency Questions only:<br>2.2: Who (object) has participated in industrial projects (events) in avionics engineering? |
| SimpleOrAggregated | What elements are aggregated members of this object? | + | Complete specification:<br>e.g. requirement 1.6: aggregation of courses taught by expert |

**5.3. Filter by shared conceptualizations**

An automated matching has not been tested. However, there were no identical IRIs in the patterns and the ExpertFinder ontology. Even just looking into the patterns, the same conceptualizations where represented by different IRIs. Manual interpretation by reviewing the OWL representations of the patterns, the "Solution Description", and the "Scenarios" did not lead to a reduction of candidate patterns because there were at least overlaps of the conceptualizations.

An exemption forms the "Template Instance" pattern, which does not introduce new conceptualizations in addition do RDF basics. It adds an annotation "Template" and describes how to model individuals with recurring property values as templates based on that annotation. In general, it is applicable if there are such individuals in the ontology either before or after restructuring based on other implemented ODPs. The goal of the "Template Instance" pattern is a reduction of ontology size. This is different from the other ODPs. The question of the purpose of ODP implementation pops up again.

Exemplary decisions ( - = neglected, + = passed):

| Pattern | Conceptualization(s) | Decision | Comment |
|---|---|---|---|
| Collection | Collection | + | "EducationalProgramme" is a specialization of "Collection" |

### 5.4. Filter by compatibility

Manual evaluation of conceptualizations left 14 candidates. Four of them have no competency questions given. Compared to the 18 patterns without competency questions in step 2, it seems that a lot of effort could have been saved if there were competency questions for filtering.

Exemplary decisions ( - = neglected, + = passed):

| Pattern | Conceptualization(s) | Decision | Comment |
|---|---|---|---|
| Criterion | Description | - | Contains specialization of “Description” which does not fit to the purpose of the ontology |
| Collection | Collection | + | “EducationalProgramme” is a specialization of “Collection”. There are no further conceptualizations in the pattern. |

### 5.5. Final pattern selection and implementation

A closer look into the remaining patterns revealed that some were incompatible with each other. Some are specializations or inclusions of other patterns. Also there is for example a basic pattern building block “TimeIndex” that is recurring in several patterns but is not listed as a pattern itself.

Information about the relationships between the patterns can save a lot of effort at this step. A specialization of an incompatible pattern or an inclusion of it will generally be also incompatible. At the end, five patterns remained. Only one of them, namely “Nary participation”, was among the patterns, found in step 2 considering solely a comparison of competency questions. Therefore, the restriction of requirements filtering to competency questions has to be seen critically.

The “Time indexed Participation” pattern for example describes basically the same as the “Nary Participation” pattern but uses a different conceptualization and structure which is due to abstraction level discrepancies. Thus, only one of the patterns can be implemented. The “Nary Participation” pattern seemed to be less complex and had been chosen. The “Agent Role” pattern is a specialization of conceptualizations in the “Object Role” pattern. Finally, the following patterns have been selected:

- “Nary Participation”: It can be used to describe the participation of experts in projects and courses. It contains the patterns or pattern building blocks “Participation”, “Situation”, “Time interval”.
- “Collection”: It can be used to describe the courses within an educational programme.
- “Classification”: It can be used to classify “Course”, “Project”, and “Publication” by “ResearchField”
- “Persons”: It can be used to express the relation of the “Expert” to “Position” and “UniversitySchool”. It contains the patterns or pattern building blocks “AgentRole”, “Classification”, “Description”.Altkough, ”Description” is used to provide a definition to a “SocialPerson”, such information is not part of the original ExpertFinder ontology. Therefore, this conceptualization will be left out and an adaptation of “Persons” is used.
- “Topic”: It can be used to model the relations between “ResearchField” instances more comprehensive.

Steps for the actual implementation of patterns in ontologies are described for example in [14] and [13]. These steps would follow the pattern selection, but they are not in our focus.

## 6 Conclusions

In general, it has been proven that appropriate patterns for ontology quality improvement can be found in a structured way by the application of the proposed strategy. However, some problems were evident and there is a lot of space for improvements .

A first step on side of the provision of ODPs would be a better documentation of patterns. Several patterns missed information like domain, intent, and competency questions. Additionally, some standard for domain description would be helpful. Since there are close dependencies between ODPs like specialization and inclusion, an overview or a formal description of these dependencies is of interest. This idea has also been proposed quite similarly in [12]. Regarding the incompatibilities of the ODPs it should be investigated whether a harmonization is possible or some patterns may be neglected because they base on similar conceptualizations or incompatibilities should be part of pattern descriptions in conjunction with recommendations which pattern to use in what scenario. Maybe a restriction to a smaller set of basic patterns would be helpful too. This basic patterns would be generally smaller in size than patterns that combine several sub-patterns as they are present now. In consequence, less dependencies have to be considered and understandability of patterns increases. Another problem were unclear conceptualizations. For example, what exactly is an "object" in the context of a pattern, what a "concept"? A definition of such terms as part of the pattern documentation is suggested.

Looking at the process of pattern selection there is little potential for automatization. A semi-automatic process may include a search for domains and conceptualizations by IRIs or labels in conjunction with information about synonyms and related terms.

For example, Sabou et al. discussed such approaches in [15]. An approach to formalize competency questions and use them for automated pattern selection is seen critical. As shown earlier, competency questions do not need to cover the ontology scope completely. Furthermore, the level of abstraction of formulated competency questions varies depending on the ontology engineer. This phenomenon can also be observed looking at the competency questions of ODPs. Some steps of the pattern selection will always have to be done manually.

A question that is still open and cannot be answered in this paper is, what actual quality improvement is achieved by use of patterns. Some patterns like "Template instance" aim at computational aspects, namely space requirements, but what do the others aim for? What if ODPs are incompatible with standard ontologies as described in [13]. Is the ODP a bad pattern in this case or not? What are the preferences of the ontology engineer – being compatible with standards or achieve whatever quality improvement by the pattern?

## References

1. Gruber, T. (1993) A translation approach to portable ontology specifications. In Knowledge Acquisition, vol. 5, pages 199–220, 1993.
2. Gangemi, A. (2005) Ontology design patterns for semantic web content. In The Semantic Web ISWC 2005, Vol. 3729, Lecture Notes in Computer Science. Springer, 2005.
3. Blomqvist E., Sandkuhl K. (2005) Patterns in Ontology Engineering – Classification of Ontology Patterns. Proc. 7th International Conference on Enterprise Information Systems, Miami, USA, May 2005.
4. Blomqvist, E. (2009) Semi-automatic Ontology Construction based on Patterns. PhD thesis, Linköping University, Department of Computer and Information Science, 2009.
5. Gangemi, A. and V. Presutti (2009) Ontology design patterns. In Handbook on Ontologies, 2nd Ed., International Handbooks on Information Systems. Springer, 2009.
6. Vrandečić, D., Sure, Y., 2007. How to Design Better Ontology Metrics, in: Franconi, E., Kifer, M., May, W. Eds., The Semantic Web: Research and Applications. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 311–325.
7. Duque-Ramos, A., Fernandez-Breis, J.T., Stevens, R., Aussenac-Gilles, N., 2011. OQuaRE: A SQuaRE-based approach for evaluating the quality of ontologies. Journal of Research and Practice in Information Technology 43, 159.
8. Maedche, A., Staab, S., 2002. Measuring Similarity between Ontologies, in: Gómez-Pérez, A., Benjamins, V.R. Eds., Knowledge Engineering and Knowledge Management: Ontologies and the Semantic Web. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 251–263.
9. Brank, J., Grobelnik, M., Mladenić, D., 2005. A survey of ontology evaluation techniques, in: Proceedings of the Conference on Data Mining and Data Warehouses (SIGKDD 2005). Ljubljana, Slovenia.
10. Haslhofer, Bernhard; Klas, Wolfgang (2010): A survey of techniques for achieving metadata interoperability. In *ACM Comput. Surv* 42 (2).
11. Natalya F. Noy and Deborah L. McGuinness (2001): Ontology Development 101: A Guide to Creating Your First Ontology. In *Stanford Knowledge Systems Laboratory Technical Report KSL-01-05 and Stanford Medical Informatics Technical Report SMI-2001-0880*.2001
12. Hammar, Karl (2012): Ontology Design Patterns in Use - Lessons Learnt from an Ontology Engineering Case. In Eva Blomqvist, Aldo Gangemi, Karl Hammar, Mar\'ıa del Carmen Suárez-Figueroa (Eds.): Proceedings of the 3rd Workshop on Ontology Patterns, Boston, USA, November 12, 2012: CEUR-WS.org (CEUR Workshop Proceedings, 929).
13. Hammar, K., Tarasov, V. & Lin, F. (2010). Information Reuse and Interoperability with Ontology Patterns and Linked Data: First Experiences from the ExpertFinder Project. In: Witold Abramowicz, Robert Tolksdorf, Krzysztof Wecel (Ed.), Business information system workshops: BIS 2010 International Workshops, 3rd Workshop on Information Logistics and Knowledge Supply in conjunction with 13th International Conference on Business Information Systems (pp. 168-179). Berlin: Springer, LNBIP 57
14. Presutti, Valentina; Gangemi, Aldo (2008): Content Ontology Design Patterns as Practical Building Blocks for Web Ontologies. In Qing Li, Stefano Spaccapietra, Eric S. K. Yu, Antoni Olivé (Eds.): Conceptual Modeling - ER 2008, 27th International Conference on Conceptual Modeling, Barcelona, Spain, October 20-24, 2008. Proceedings: Springer (Lecture Notes in Computer Science, 5231), pp. 128–141.
15. Sabou, Marta; Lopez, Vanessa; Motta, Enrico (2006): Ontology Selection for the Real Semantic Web: How to Cover the Queen's Birthday Dinner? In Steffen Staab, Vojtech Svátek (Eds.): Managing Knowledge in a World of Networks, 15th International Conference, EKAW 2006, Podebrady, Czech Republic, October 2-6, 2006, Proceedings: Springer (Lecture Notes in Computer Science, 4248), pp. 96–111.