

Towards Systematic Reflection of Data, Information, and Knowledge

Ligita Businska, *Riga Technical University*, Inese Supulniece, *Riga Technical University*

Abstract – Currently available business process modeling notations do not provide an opportunity to clearly distinguish between data, information and knowledge. The problem lies mainly in a still not fully agreed-upon understanding of relationship between the notions *data*, *information*, and *knowledge*. Therefore in this paper we focus on the analysis of this relationship by investigating intersection of modern information theory assumptions and knowledge management definitions of information and knowledge. The results are expressed as new graphical notation for knowledge, information, and data flow systematic reflection in business process model.

Keywords – data, information, knowledge, business process model.

I. INTRODUCTION

The concepts of data, information, and knowledge are widely used for a long time and have demonstrated diverse nature. Concepts related to these terms have been important enough to become widely utilized. However such wide and extensive distribution has made the boundary between the mentioned concepts indistinct. As a result, many researchers consider data, information, and knowledge as quite intuitive and strictly not definable. Certain confusion exists also in modeling of the mentioned concepts using business process modeling notations (e.g., IDEF0 [1], IDEF3 [2], eEPC diagrams in ARIS tool [3], GRAPES BM in GRADE tool [4], UML 2.0 activity diagram [5], BPMN 2.0 [6]) and knowledge modeling languages (e.g., KMDL [7], PROMOTE [8], RAD[9]). Often data, information, and material flow are represented by the same symbols and without unambiguous definitions of these concepts. At the same time knowledge has poor or no modeling capabilities in business process modeling notations, but knowledge oriented languages have limited process perspective representation. This leads to ambiguity and misunderstanding of the developed models, especially in modeling knowledge intensive processes [10].

Separation of the notion of information from knowledge and data started in the middle of the 20th century, when the significance of information and communication technology increased in business. Earlier it was not essential (originally computers were used for data processing; however, since 1970-ties information processing systems were introduced). Since that time concepts „data” and „information” were used in various ways, however they were often mutually replaceable. Today „information” is often used as a synonym for terms the news, knowledge, and notice. In cybernetics „information” characterizes control signal, which is transferred via the communication channel. From philosophy

perspective „information” is related to interaction and cognition. In probability theory „information” is established as uncertainty prevention measure and ensures possibility to measure information quantitatively.

The knowledge concept includes various interpretations. For more than two centuries a number of philosophers and scientists attempted to understand the nature of knowledge, highlighting different knowledge properties and regularities. However in the business society active interest in knowledge started only in the 50-60-ties of the 20th century, when it was detected as one of essential management objects. Today value of knowledge is recognized in almost all aspects of social and professional operations.

This paper proposes to introduce a more explicit border between data, information, and knowledge concepts in business process models based on modern approaches of information theory [11, 12], which examine information interaction processes from signal theory perspective. Since the central research object of the information theory is information, it is attempted to integrate it with knowledge definition from knowledge management and with data definition from information technology. In the paper mentioned concept definitions are proposed and essential requirements for knowledge, information and data flow systematic reflection in business process model are listed. Acquired results are illustrated and analyzed with the help of an example.

In Section 2 we consider the terms data, information, and knowledge and come to the conclusion that the use of information codes as a supplementary notion helps to clarify relationship between these three terms. In Section 2.B we analyze information interaction in the context of the business process model. In Section 3 practical interpretation of definition of data, information, and knowledge is given and essential requirements for their systematic reflection in business process models are listed. In Section 4 the research results are illustrated and analyzed by example pointing out benefits and further improvements. Section 5 consists of brief conclusions and outlines further research work to be done.

II. DATA, INFORMATION, AND KNOWLEDGE: THEORETICAL ISSUES

Present achievements in the information theory have shown that in information exchange a substance called information codes is involved [11,12], i.e., information exchange is accomplished via information codes.

Suppose a simple scheme that consists of three concepts: Object, Environment, and Interaction. Object is anything

consistent in time and limited in space that is perceptible as a single whole. Environment is a set of all potential objects that exist around and are capable to influence the given object, and, v.v., the given object may have impact on those objects in the environment. Interaction is a continuous process of interdependent change of parameters of object and environment state. This scheme represents a closed system in the sense that the environment includes all potential objects capable to influence the state of the given object.

In the nature two fundamental kinds of interaction exist: an exchange of matter and energy. All other types of interactions occur only by means of those fundamental interactions. These fundamental kinds of interaction conform to the conservation law. As much substance and energy one object has transferred to another, that is as much another object has received (in this sense the fundamental interaction is symmetric). The losses of energy and matter occurring during interaction aren't considered, because we assume that in this case another interaction with other objects of the same environment occurs.

Let us assume that a combination of fundamental interactions may be the basis for higher order interactions such as information interaction. In contrast to fundamental interaction the information interaction is asymmetric, e.g., during this kind of interaction one object receives the same amount of substance (namely, information), but another object, that is the holder of giving substance, does not lose it. The form of energy and material with the help of which the substance or information is transferred during the information interaction is called information codes.

Thus, we could conclude that information interaction has such basic properties as: a) information cannot exist externally out of interaction between objects, and b) information does not lose any object during interaction. In the next subsection we will define basic stages of information interaction and specify how data, information and knowledge occur and are transformed during information interaction.

A. Stages of information interaction and definitions of data, information, and knowledge

According to the information theory [11] a single cycle of information interaction between an object and its environment is divided into three sequential phases: (1) object receives information codes from its environment, (2) obtained codes are interpreted, and finally (3) information is realized (reflected upon, absorbed, put into action). The single cycle of information interaction between two objects is illustrated in Fig. 1.

A. Information codes perception and detection. In the first phase, when object *O1* starts the information interaction with object *O2*, a certain amount of information codes is transferred. Let us consider that during the information interaction object *O2* receives the set of information codes Ic_1 . To perceive those codes object *O2* needs a particular "linguistic" device that can recognize the codes. (E.g., if the codes are information in English, it can be recognized if there is a "device" that can handle English). The received codes change internal state parameters of object *O2* and are

transformed into a particular set of data Δd . Meanwhile the obtained data have no meaning for object *O2* and it's not clear what part of them and to what extend they are useful for object *O2*. Thus, *data are functional values of information codes which correspond to new parameters of object state obtained in interaction with another object* [10].

B. Information codes interpretation. In the next stage object *O2* decides upon the meaning of the obtained data set Δd comparing it with the current set of goals *M1*, and identifying those goals that could be realized with a given set of data [11]. The goal in this context is the overall vector of object actions that allow satisfying the particular needs of the object. The needs of the object may change and thus result in modification of goals structure and selection of another direction of object actions. The data without meaning does not have any information for the object and returns it in the initial state. The data useful for object *O2* are interpreted as a result of processing information Δi . The interpretation is done inside the object by current knowledge *K1* and in this sense it is subjective. According to [13] *a structured and processed data is information that is time dependent (relevant only in a given point of time) and correct with respect to the processed data set.*

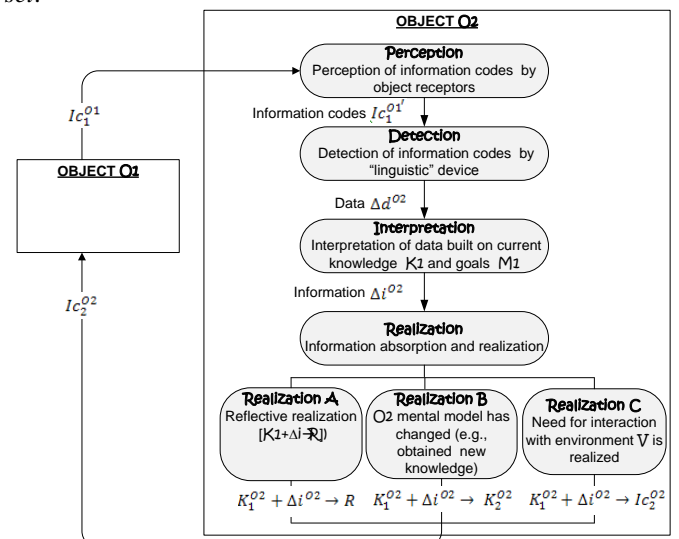


Fig. 1. Information interaction between object *O1* and object *O2*

In general, the amount of received information can be calculated as difference between knowledge after data interpretation and knowledge before interaction with object *O1*: $\Delta i = K1 - K2$. It can be regarded as a measure of reduction of uncertainty for choosing actions in order to achieve particular goals *M1* [14]. Information exists from the moment data is interpreted till the moment when the information has been absorbed or included in the mental model of the object. As a result of information absorption the content or structure of the mental model (including procedural and declarative knowledge which is stored in it) can be changed.

C. Information codes realization. In the final phase realization of obtained information Δi can lead to changes of internal state parameters of object *O2* or/and to the next cycle of interaction with environment. There can be several

overlapping option of realization, for example: (1) a reflective action: $K1 + \Delta i \rightarrow R$; (2) if the object starts the next cycle of iteration with object(-s) from its environment, object $O2$ delivers appropriate set of information codes: $K1 + \Delta i \rightarrow Ic2$; (3) if object changes its internal state, its mental model can change, in certain conditions obtaining new knowledge: $K1 + \Delta i \rightarrow K2$. According to [15] knowledge is reasoning about data that is stored in the “mental model” of the object in order to promote action, problem solving, decision making, learning, and teaching.

Knowledge is a higher organizational level of data that allows their specific interpretation. Requirements to data organization level can differ from a simple grouping of the data to complicated data hyper-structures.

Based on the above-described information interaction process we can assume that during communication people exchange objective information codes in a general sense. However, data, information, and knowledge are obtained from perceived information codes and belong to their owner. In this sense they have a subjective nature.

B. Information interaction in business process

Scientific literature offers a variety of business process definitions that highlight several characteristics of a business process. Our purpose is not to investigate peculiarities of the available definitions; we focus on information interactions that occur during process execution. In general sense a business process is a network of connected activities and buffers with well defined boundaries and precedence relationships, which utilize resources to transform inputs into outputs for the purpose of satisfying customer requirements [16]. Information interaction is part of a business process, i.e., business process activities are performed by a person(-s) or computer system(-s) that communicate with each other using several documents, graphics, tables, and other artificial objects of data, information, and knowledge as inputs, outputs, or resources.

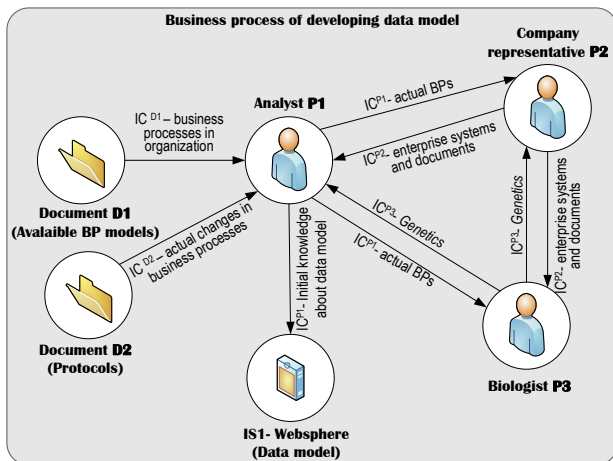


Fig. 2. Information interaction network between participants (Process performer is noted by circle and received information codes - by arrows)

Thus, during the business process several natural and artificial objects interact with each other in order to exchange data, information or knowledge. Graphically this could be modeled as a network shown in Figure 2. If we additionally

define what activities and in what sequence are carried out, we could obtain process view represented in Figure 3.

Information interaction occurs when process performer(-s) fulfill(-s) activity, e.g., in Figure 1 there is an example of two consecutive activities that are performed while the conceptual data model of Bioinformatics Company is created. At first, the analyst studies documents available in the company and during this activity he/her perceives information codes that are embodied in a particular document. Then the analyst together with a biologist and a company representative discuss knowledge obtained in the previous activity. As a result they exchange information codes and each of them processes perceived codes inside their brains according to personal mental model and goals. As a result, they could extend their mental model with new data or under certain conditions obtain new knowledge.

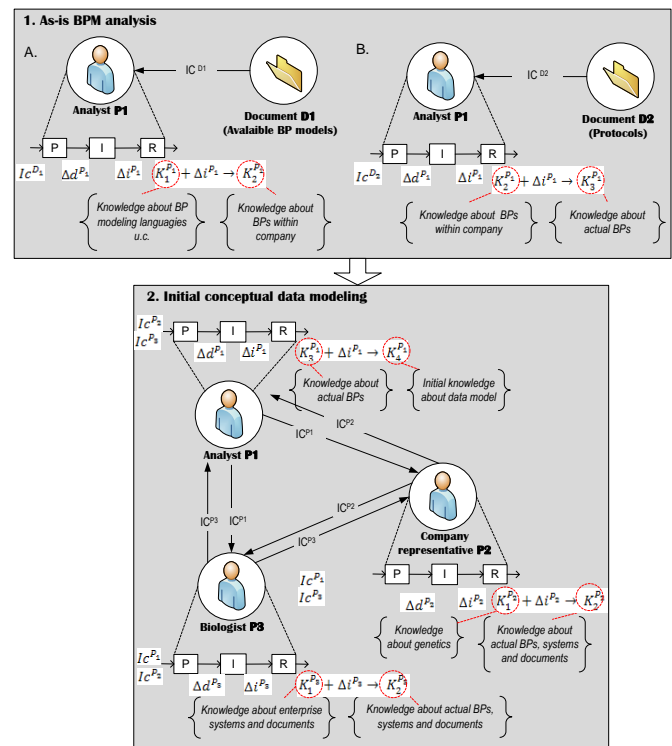


Fig. 3. Business process view of information interaction between participants. (P – perception, I – interpretation, R – realization)

III. SYSTEMATIC REPRESENTATIONS OF DATA, INFORMATION, AND KNOWLEDGE IN BUSINESS PROCESS MODEL

In our previous research [17] we tried to extend existing business process notation (namely, BPMN) with knowledge flow, but it did not bring expected results. Firstly, existing notations (BPMN as well as ARIS eEPC, GRADE, IDEF0, IDEF3 and UML activity diagram) do not strictly separate data from information; secondly, the owner of data, information, and knowledge is not indicated and; thirdly, these languages and notations are suitable for process sequence representation, but poorly present data and information flows. Additionally, process oriented notations do not include knowledge flow, and knowledge oriented notations (e.g.,

KMDL, RAD) have weak process logic representation and have no representation of data.

In this paper we modify the business process model in order to avoid the mentioned incompleteness and to meet the definitions obtained in Section 2.2. Thus, we try to represent separately data, information, and knowledge flows that are exchanged between actors who fulfill a particular activity. Table 1 illustrates elements of the proposed graphical representation, and Table 2 shows possible connections between these elements.

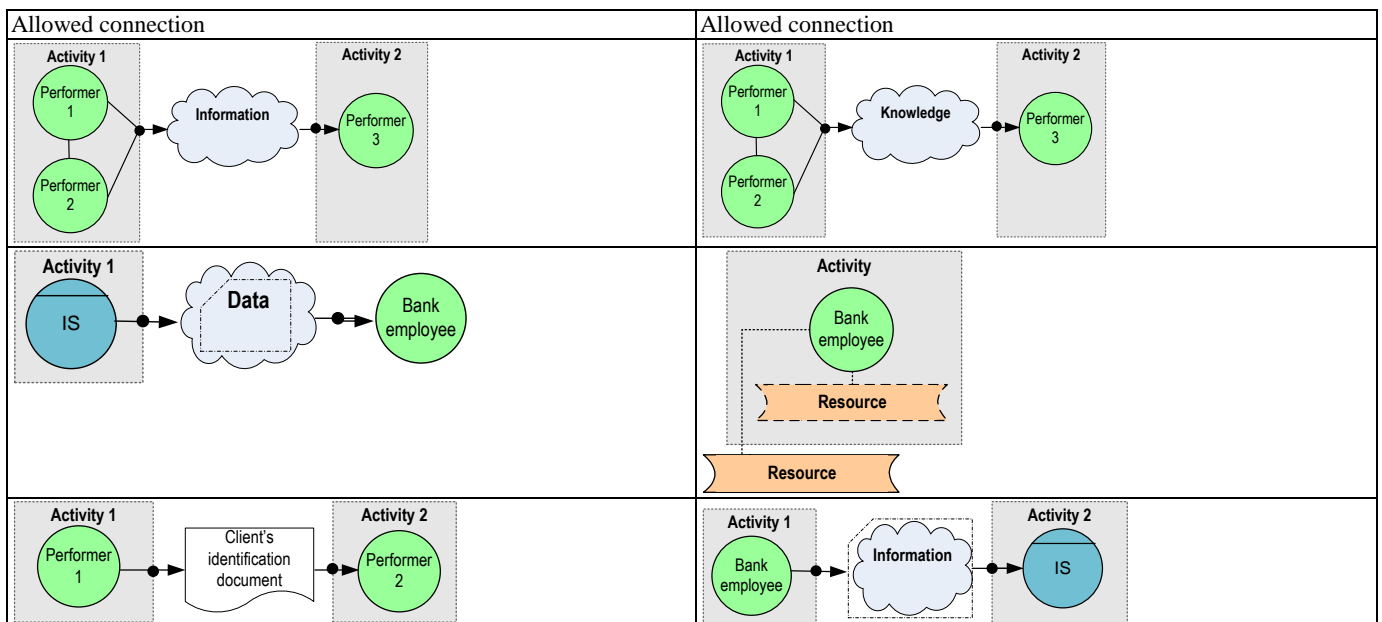
Our aim is not to establish a new business process notation, but rather to reflect the proposed ideas graphically. More

studies and experiments are required to determine the best way of reflection. The main difference from the existing notations is pointing out the owner of data, information, and knowledge, which can be a human or artificial object, e.g., information system or computer system. Processing of information codes occurs inside the actor, but in the model it is shown just as the result of the processing: obtained knowledge or received interpretation of data (information). Additionally, actors have a link to materialized or non-materialized knowledge which is used as a resource for processing information codes.

TABLE I
VISUAL REPRESENTATION OF ELEMENTS

Elements	Description	Elements	Description
	Role that is carried out by a human being performing a certain activity and is represented inside the activity		A computer system (information system or intellectual agent) that is a performer of an activity and is represented inside the activity
	Document in hard (non-electronic) format that is the snapshot of performer's data, information, or knowledge		Document in electronic format, which is the snapshot of performer's data, information or knowledge
	Materialized knowledge that is used as a resource, thus this knowledge is used to interpret information codes, but is not changed during the activity		Non-materialized knowledge that is used as a resource: it is just potential knowledge that should have a performer to positively influence performance of the activity
	Non-materialized information that is input or output of an activity. Information is always connected to its owner and is short term, e.g., it is not important to save it for a long period of time		Non-materialized knowledge that is input or output of an activity. Knowledge is always connected to its owner and is long term, e.g. it is important to save it for a long period of time
	Non-materialized information that belongs to human and is an input to the activity which is performed by an information system or an intelligent agent		Data that is output of an information system or intelligent agent and is used as input to human performed activity
	Business process activity		Data that is input or output of an activity. Data is always connected to its owner, e.g., information system or intelligent agent

TABLE II
VISUAL REPRESENTATION OF POSSIBLE CONNECTIONS



IV. ILLUSTRATIVE EXAMPLE AND ANALYSIS OF THE PROPOSED GRAPHICAL REPRESENTATION

To illustrate the described differences among data, information, and knowledge representation, a simplified business process example from the banking industry is used. The presented business process does not reflect all details and flow alternatives as this is not the aim of this paper. Business process starts, when client contacts the bank with the aim to open a deposit account. The bank employee presents to the client available products/services and their conditions. Based on this information, the client chooses a fixed-term deposit (also known as a term deposit, a bond or fixed deposit in other countries). Selection of other deposit products is not modeled. Then the bank employee presents possible “terms” (period of time) and respective interest rate offers, and the client selects one. The bank employee asks if the client has the bank account. The process branch for the client without the account is not modeled to keep the example as small as possible. If the client has the account, then the employee starts identification process to perform manipulations with the client's account. For identification the employee asks the client to present some document, e.g., passport, ID, or a similar document. Then he/his enters some client data in the information system to check validity of the document, whether the client has rights to manipulate with the account and other information according to internal regulations. When the client is successfully identified, the bank employee prepares the contract in the information system and asks the client to sign it. Simplified business process example ends when the client signs the contract.

Analyzing the developed business process model we make several conclusions that distinguish the proposed graphical representation from other business process modeling languages:

- Unlike in other modeling languages in the proposed model two flows are represented: process execution flow and information, knowledge, and data flow. This is obtained with the link between information, knowledge, data and their owner and the flows of the data, information, and knowledge.
- Proposed graphical representation allows better structuring of the process by outlining the information processes as much as possible. In other words, if the aim is to model information, knowledge, and data flow, more accurate modeling is required, which is not essential while modeling only process execution logics. For example, in Figure 4 two activities (“present products and conditions” and “select appropriate products”) could be the same activity in process view, however, in information process view they should be separated not to break the flow.

- Three types of flows may exist between the processes: data, information, and knowledge flow. Data flow might appear between artificial systems. Information and knowledge flow might appear between human-human or human-artificial, or human-document systems. In addition, knowledge is represented as a flow if it is essential for executor to keep it permanently, because it affects the successful process completion. Information flow appears when the executor remembers it temporarily and does not transform it into knowledge. For example, in Figure 4 a bank employee remembers what products were offered according to client goals which increases personal competence and speeds up client servicing. At the same time information about goals of a particular client becomes insignificant when the next client comes to the desk.
- It is vital to understand the interaction between a human and artificial computer system. The interchange and knowledge development can proceed differently depending on the level of intelligence of the artificial system (e.g., without data interpretation; only with data interpretation, and with ability to create new knowledge). In the graphical notation a special symbol is proposed for the cases when computer system perceives just data without interpretation, e.g., when the bank employee enters selected maturity period in the information system.
- In the model non-materialized and materialized information and knowledge are displayed. Non-materialized knowledge and information belong to the process executor, but materialized knowledge and information are snapshots of it. Materialized knowledge and information are stored in a document (hard or electronic form).
- While modeling knowledge, information, and data, it is impossible to declare that they belong to the executor in reality, but it is possible to model only potential or a prospective knowledge, information, and data.
- In the proposed representation a concept “knowledge” as resource which might be a requirement for the role is introduced. Knowledge as resource unlike other types of resources (materials, technology) does not have amortization, however training curve needs to be taken into account (time is essential for executor to collect all the required knowledge and skills).
- Resources, which are materialized and available in the knowledge repository, are separated from the resources that are not materialized, but can encourage process completion. For example, in Figure 4 in the first activity the bank employee needs to learn conversation models which are presented in the bank's repository. However, the bank employee still needs to improve personal skills to lead the conversation with the client.

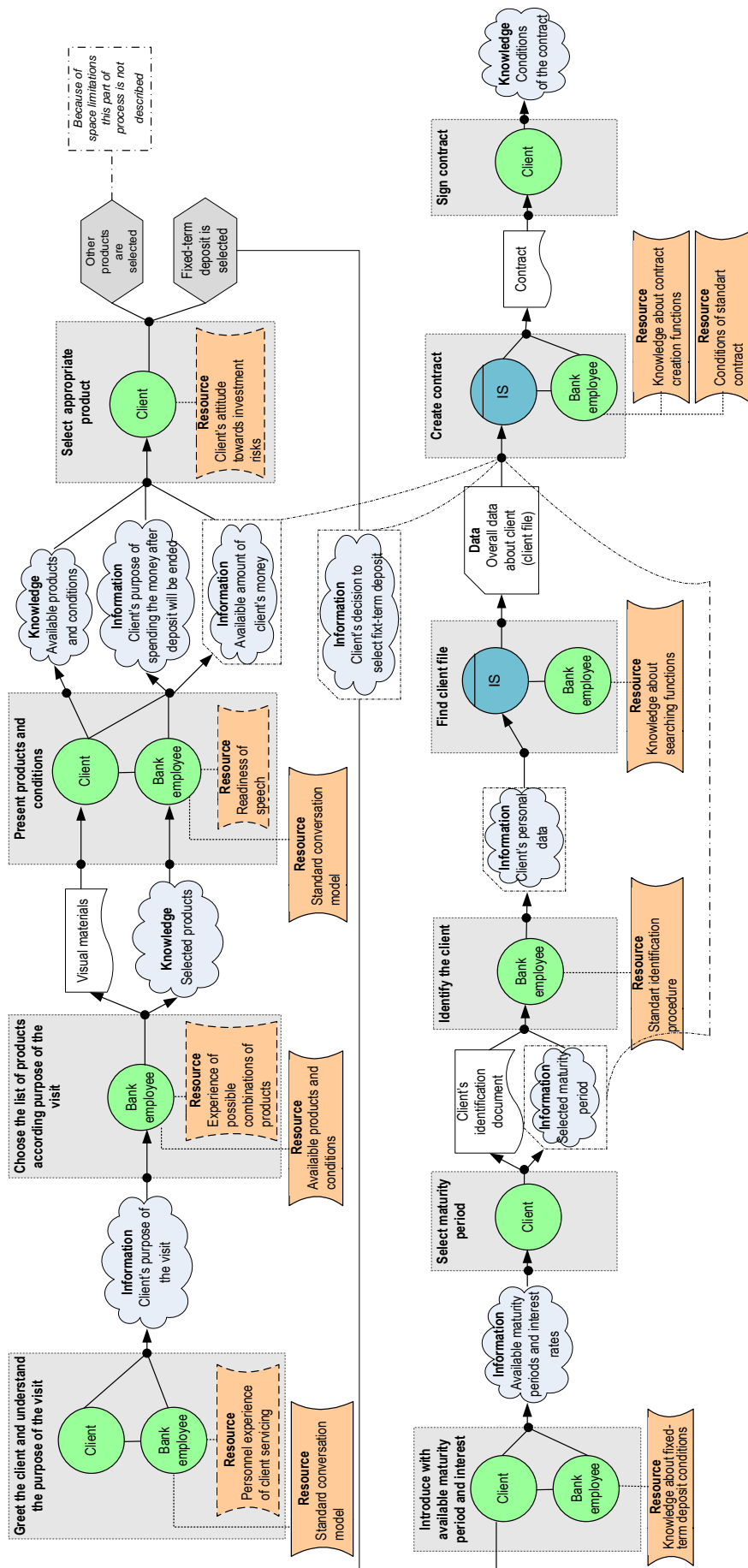


Fig. 4. Example of proposed notation.

In Figure 4 the developed model should be bonded with other model types, e.g., knowledge model for the role (role-oriented knowledge management) or for a particular person (human-oriented knowledge management), organization structure, competence maps, and conversation models. In addition the possibility to model two separate views would be required: a process view and an information process view. In the process view emphasis would be on process execution logics, as in ARIS EPC and BPMN, but the information process view could be similar to proposed representation.

V. CONCLUSIONS AND FUTURE WORK

Three different phenomena: data, information, and knowledge are relevant in business process modeling. However, business process modeling notations currently do not provide an opportunity to clearly distinguish between them. The problem mainly resides in a still not fully agreed-upon understanding of relationship between such notions as data, information, and knowledge [10]. Therefore in this paper we focus on the analysis of this relationship by investigating intersection of modern information theory assumptions and knowledge management definitions of information and knowledge. The obtained results, as well as a proposed graphical representation are presented and discussed in this paper. The graphical representation is in its experimental stage. Analysis of possibility to introduce it into different business process modeling languages is the next step of the research presented in the paper. Proposed graphical representation of business process model should be connected with other model types, e.g. knowledge model for the role or for a particular person, organizational structure, competence maps, and conversation models. In addition the possibility to model two separate views: the process view and the information process view would be required.

ACKNOWLEDGMENTS

This research has been supported by the European Social Fund within the National Program "Support for carrying out doctoral study programs".

REFERENCES

1. *IDEF0*, [Accessed 02.08.2011]. Available: <http://www.idef.com/IDEF0.html>
2. *IDEF3*, [Accessed 102.08.2011]. Available: <http://www.idef.com/IDEF3.html>
3. ARIS Expert Paper, Business Process Design as the Basis for Compliance Management, *Enterprise Architecture and Business Rules*, March 2007

4. GRADE Business Modeling, *Language Reference*, Infologistik GmbH, 1998
5. *UML*, [Accessed 102.08.2011]. Available: <http://www.uml.org/>
6. *BPMN*, [Accessed 102.08.2011]. Available: <http://www.omg.org/spec/BPMN/2.0/PDF>
7. Arbeitsbericht (umfangreiche Beschreibung) - *KMDL® v2.2*, 2009, [Accessed 102.08.2011]. Available: <http://www.kmdl.de>
8. **Woitsch, R., Karagiannis, D.** Process Oriented Knowledge Management: A Service Based Approach. *Journal of Universal Computer Science*, vol. 11, no. 4, 205, p. 565-588
9. **Weidong, Z., Weihui, D.** Integrated Modeling of Business Processes and Knowledge Flow Based on RAD. *In: International Symposium on Knowledge Acquisition and Modeling*, Wuhan, China, 2008, p. 49-53
10. **Businska, L., Supulniece, I., Kirikova, M.** On data, information, and knowledge representation in business process models. *In: The 20th International Conference on Information Systems Development (ISD 2011)*, Edinburgh, Scotland, August 24-26, 2011 – to be published by Springer
11. **Янковский, С.Я.** Концепции общей теории информации. НиТ. Текущие публикации, 2001, [Accessed 102.08.2011]. Available: <http://n-t.ru/tp/ng/oti03.htm>
12. **Романенко, В.Н., Никитина Г.В.** Многозначность понятия информации. *Философия науки*, 2010, №4 (47), стр. 75-99
13. **Maier, R.** *Knowledge management systems. Information and communication Technologies for knowledge management*. Springer-Verlag Berlin Heidelberg, third edition, 2007.
14. **Corning, P.A.** Control Information Theory: 'The missing link' in the science of cybernetics. *Systems research and behavioral science*, Syst.Res. 24, 2007, p. 297-311
15. **Beckman, T.** A methodology for knowledge management. *In: The IASTED International Conference on Artificial Intelligence and Soft Computing (ASC'97)*, Banff, Canada, 1997, p. 29-32.
16. **Laguna, M., Marklund, J.** *Business Process Modeling, Simulation, and Design*. Pearson Education, Inc., New Jersey, 2005
17. **Supulniece, I., Businska, L., Kirikova, M.** Towards extending BPMN with the knowledge dimension. *In: The Enterprise, Business-Process and Information Systems Modeling*, Tunisia, Hammamet, 2010, p. 69-81

Ligita Businska is a PhD student of Computer Systems at Riga Technical University in Latvia. She obtained master degree in Computer Systems from the mentioned institution in 2002. Her main research fields are business process modeling, knowledge oriented business process management and aspect-oriented requirements engineering. She has experience in working as a system analyst in some development projects. Her professional activities are related to the analysis of complex geographical information systems in the State Land Service of the Republic of Latvia. Ligita Businska works also as an assistant at the Department of Systems Theory and Design of Riga Technical University. She has participated in several research projects related to the research fields such as agile information systems, multifractal systems and knowledge management system for future oriented research.

Inese Supulniece is a PhD student at the Institute of Information Technology of Riga Technical University in Latvia. She obtained master degree in Information Technology from the mentioned institution in 2009. Her main research fields are user-adaptive systems, user modeling, business process modeling and process individualization in business applications. She has participated in several research projects related to adaptive systems. Her professional activities are related to business and system analysis.