Proceedings of the Federated Conference on
Computer Science and Information Systems pp. 1105–1110
ISBN 978-83-60810-51-4

Applying Linked Data Concepts in BPM

Silva Robak
Uniwersytet Zielonogórski, ul. prof. Z. Szafrana 4a, 65-516 Zielona Góra, Poland
Email: s.robak@wmie.uz.zgora.pl

Bogdan Franczyk
Uniwersytet Ekonomiczny we Wrocławiu, ul. Komandorska 118/120, 53-345 Wroclaw, Universität Leipzig, Germany
Email: franczyk@wifa.uni-leipzig.de, bogdan.franczyk@ue.wroc.pl

Marcin Robak
XLogics Sp. z o.o., ul. Kostrzyńska 4, 65-127 Zielona Góra, Poland
Email: m.robak@xlogics.eu

Abstract—**One of the contemporary problems in business networks of supply chains are the information integration issues. They are either related to incompatible information interchange cross independently designed data systems or lack of common semantic model in the domain. The networked supply chains need mechanisms to describe the choreographies of the cooperating business units. In this paper we analyze the possibilities of an application of the (some) Linked Data concepts into the interaction models for the choreographies in business process management BPM. We present our approach on a 4PL (Fourth-party Logistics) integrator example.**

I. Introduction

IN the contemporary world the business companies have to face unprecedented challenges. As a result of globalization the competition is becoming fiercer and the customers often expect integrated services, what requires a close cooperation between several involved organizations. Therefore, companies have to adapt to new, such as networked business models, and further rethink their role and position in their value chain regarding the potential possibilities to add value for their customer and suppliers. This requires some changes from companies in their organizational view, but at the same time in their information technology view. The second result comes mainly from the appropriate technology environment needed to support their interoperable business cooperation.

The problem of the appropriate information technology environments for collaborative processes between business participants is not new and it has been approached with several solutions like common IT platforms consisting of (possibly common) components based on established standards, standard enterprise information systems, and standard business protocols. Nevertheless, the IT environment platforms still contain proprietary applications like ERP (enterprise resource planning), CRM (customer relationship management), etc. The foreseen scale of collaboration between business partners may require to undertake further steps for the integration of the IT environment, such as one of the known enterprise application integration solutions or usage of the loosely coupled e-business solutions (Web services).

The inter-company networks are defined in [1] as complex arrays of relationships between companies, which establish these relationships by interacting with each other. Whilst widening the markets toward inter-company networks (webs) with the collaborating organizations, mentioned above IT integration solution approaches doesn't seem to be sufficing and satisfactory. This level of an organizational form of the market participants requires mutual adjustment in information sharing and data management, and at the same time coordination of collaborative processes of the business participants.

In the paper we will try to tackle both problem statements by recommending the usage of a common (open stated) data format as to be offered by Linked Data [2], and incorporating this idea into business collaborations in business process management BPM that can be described in form of choreographies that further could be denoted in a modeling notation, such as Business Process Model and Notation BPMN 2.0 [3]. We will consider the network from a supply chain perspective with emphasis on the value-adding partnerships.

The supply chains define the network that embraces all the organizations and activities associated with the flow and transformation of goods from the raw material stage, through to the end user, as well as the associated information flow [4]. For the aims of our paper we will concentrate on the networked supply chain activities and flow of information.

In the inter-organizational information systems, which link the companies to their suppliers, distributors and customers, a movement of information through electronic links (e.g. XML/EDI - Extensible Markup Language/ Electronic Data Interchange) takes place across organizational boundaries, between separately owned organizations. It requires not only the electronic linkage in form of basic electronic data interchange systems (as for purchase orders, delivery notes, cash flows, etc.), but also interactions in complex cash management systems or by access to shared technical databases. So the problems with sharing and the exchange of information are still viable in supply chains contexts.

The existing EDI standard, as message-centric solution has limited possibilities in enabling the value chain interactions. The representation of business processes and vocabularies in a domain to possibly automate the trading partners interactions is missing.

Another important aspect in networks of supply chains bears integrating the needed additional data from semantic Web applications into logistic systems.

A. Business Process Management

There are several known business process definitions in the literature and they vary from the perspective of their underlying business models. The business models may primarily consider the participants in a joint business venture (Elliot, Meargetta), view the processes and structure of a business organization (Timmers) or regard the perspective of a market place (Chaffey) [4]. According to M. Papazoglou, a business process consists of one or more than one related activities that together respond to a business requirement for an action [4]. He states further, that the processing steps in a workflow may undertake numerous transformations of data (geographic, technological, linguistic, syntactical and semantic transformations), communication is an important part of the process and that (e-)business processes exist within certain environments.

Another definition formulated by F. Soliman in [5], states that the business processes may be considered as a complex network of activities.

In the dynamic business environment, such as networks of venture participants involved in value chains in logistics, the key roles take the coordination problems in the business process management. Therefore, for the aims of this paper both business process definitions will be used and further examined if needed.

In our paper we will consider communication of the business processes denoted in the BPMN 2.0 in a form of the choreographies, and we will analyze the applicability of the Linked Data principles and concepts into the business process choreographies for supply chains in business networks. For this aim the rest of the paper is organized as follows.

In Section 2 we briefly characterize the business process management and choreographies concepts, and also the BPMN 2.0 elements destined for choreographies. In Section 3 we behold the Linked Data principles and concepts, in Section 4 we provide an example scenario for 4PL (Fourth-party Logistics) integrator and then we examine how the Linked Data concepts may be applied in conjunction with the BPM choreographies. We will also try to show (in Section 5) how the integration of the Linked Data into choreography tasks may be used to achieve the leverage of the choreography concepts in the semantic data environments. In the last Section we conclude our work.

II. Business Process Choreographies and BPMN 2.0

A. BPM – Orchestration and Choreography

A business process management (BPM) view involves explicit planning, organizing, execution and control of business processes [4]. In addition to other process characteristics (like environment, customer(s), data, inventories, decision points and the final process product) communication is an important part of any business process.

The processes can interact with each other in form of orchestration or choreographies. While the orchestrations are undertaken under control of a single endpoint, the choreographies incorporate the interactions of the independent agents: "*a choreography defines the sequence and conditions under which multiple cooperating independent agents exchange messages in order to perform a task to achieve a goal state*." [6].

The Web Services choreographies can be specified by means of Choreography Description Language (WS-CDL) [7].

In particular, there are two approaches to modeling the choreographies: the interaction models and the interconnection models, (i.e.) interconnected interface behavior model [3]. The interactions models consider the basic interactions in form of message exchanges, and the interconnections models concentrate on the control flow of each participant of the choreography.

For the aims of this paper we consider only the choreography interactions models. For this form of the message interchanges among collaborating partners representative approaches are WS-CDL, Let's Dance [8], iBPMN [9] and recently BPMN 2.0. We concentrate us below on the last possibility, which is the industrial international standard recommended by the Object Management Group consortium OMG, BPMN 2.0.

B. Modeling Choreographies in BPMN 2.0

The Business Process Model and Notation (BPMN) version 2.0 was released in January 2011. BPMN possesses a metamodel, own notation and the exchange format. It has a rich expressiveness for modeling business processes including orchestrations and choreographies. The modeling elements for choreographies are Choreography Tasks representing the Interaction (Message Exchange) between two Participants. The choreography task is a rectangle: its top and bottom sections denote Participants; the color of the section indicates whether the participant is initiating the Task (white rectangle section) or receiving it (shadowed rectangle section).

A refined choreography with several interactions may be contained in a Sub-Choreography. In BPMN 2.0 there is also the Call Choreography that is a wrapper for a globally defined Choreography and the plus-symbol denotes a call to a Sub-Choreography.

In the Choreography Diagram the Initiating and the Response Messages attached to the Choreography Tasks are also included. Multiple participants of the same type can be denoted as a set of participants of the same kind (denoted as three vertical slashes-symbol).

Other BPMN 2.0 elements like activities, events, timers, exceptions and compensation can supplement the choreography descriptions.

III. Linked (open) data

The Linked Data principles were introduced by Tim Berners-Lee on his presentation at the TED 2009 conference [10]. In this presentation he outlined four rules for making human or machine-readable links for the exploration of web of data. The first rule refers to the usage of Uniform Resource Identifier URI [11] for identification of items (called "things"). The second given rule specifies that only HTTP

URIs are meant, so that people can look at them and these can be found by the standard established Domain Name Space (DNS) system. The third rule was formulated for the purpose of providing additional useful information for the items defined by URI. The information should be denoted in a standard format, such as Resource Description Framework RDF* [12], (i.e.) in form of the RDF/XML or an alternative serialization (N3, Turtle). The last, fourth rule concerns providing the linkage of such described items with other related items (data), so that the related information on the Web can be discovered more easily.

The recent development in this topic is the notion of Linked *Open* Data LOD, the concept recommended by the World Wide Web Consortium W3C [2]. It is a star rating system of linked data that allows for proving to which extent the linked data can be regarded as open. The star rating is formulated as a five principles scheme, where each next scheme principle extends (for the next star added) the former one by integrating an additional feature. The first star as principle states that the data should be available on the Web, no matter in which format, but it should be one with an open license. The second principle (two stars) adds that it should be machine-readable structured data. The third principle (three stars) adds that the format of the data should be a non-proprietary format. The fourth principle assumes achievement of the former three principles and additionally presumes the usage of an open W3C standard for identification of items, like RDF (RDF/XML, N3 or Turtle) or SPARQL [2] formats (SPARQL Protocol and RDF Query Language) for larger data amounts of data sets. The last principle, required for getting the five star grading assumes contextual linkage of rated data to other resources described in the same way.

IV. Example Choreography scenario–4PL integrator

The development in contemporary logistic networks leans toward outsourcing the various logistic functions or services. The further trends include possible integration of outsourced functions/services or even the outsourcing of the whole business processes. At present the dominant are called 3PL (Third-party Logistics) solutions [13], where the external organizations support the organizations with some of the logistic activities, like the transportation or technical activities like printing of the shipping labels and manifest lists for particular carriers, and so on. At the next developmental stage the concept of the 4PL (Fourth-party Logistics) emerged. The concept was defined by Andersen Consulting (now Accenture) and encompasses the functions offered by a 4PL logistic provider, which is acting as an integrator, assembling the resources, capabilities and technology needed for design, building and running of the comprehensive supply chain solutions [14].

As an example of the ideas presented in previous sections we consider an example of a 4PL-logistic provider managing the shipment of commodities bought from the Webshops in Asia to customers residing in Europe. Our 4PL logistic integrator operates (utilizes an outsourced or provides own) software platform that integrates the requests for quotations incoming from the diverse shops and then processes the communication with the shipping software to settle the orders. The needed software platform may be owned by the 4PL or they utilize services offered by another parties. In the second case the 4PL plays the role of the "pure broker" [15] and governs the settlements of the agreements.

One of the important services offered by the 4PL and realized from the software platforms (as their important functions) consists of checking the availabilities and the economic aspects of the possible transport routes. The international 4PL does not need to have its own transport [15]. They may work directly with companies offering transport, or directly with the 3PL providers, what includes different kinds of carriers, consolidators and forwarders (ocean carriers, airfreight forwarder, local carriers).

After comparison and selection from different carrier alternatives, the software platform enables the transport of a given Webshop's parcel with the best possible (i.e., fastest, cheapest, trackable) available (offered by 4PL) transport mode.

Thus, the purpose of 4PL-software as the carrier integrator is one of its most important roles of the 4PL. It includes taking into account the up-to-dateness of solutions for the European market. The particular solutions will be configured in accordance with the data provided by the contracted fulfillment company. As the 3PL they may have own warehouses and established contracts with the carriers, which are delivering palletized shipments to Europe (i.e. by sea or air). Their carriers (e.g. DHL, UPS) pick the individual parcels, which are inbound from Asia, from the fulfillment warehouses in Europe. With the physical set of parcels the carrier receives the printed listing with information about the parcels. Every single parcel has a label, which is printed by the Webshop in Asia, and fully complies with particular European carrier labeling specification.

A. Exchanged Data Flows

In the above scenario the needed data is transferred between various software systems, where possibly each of the system (or majority of them) maintains own data formats and makes use of the customized standards for communication aids (e.g. Electronic Data Interchange EDI [16]).

Firstly, the Webshop in Asia stores up in proprietary back-end databases or other back-end storage systems the order data obtained from the Web forms.

Secondly, the collected order data is forwarded to a communication platform, which contains software-offering services for decisions support regarding the choice of the carrier and preparing and creating the shipments (parcels) for the particular carrier system. The communication between the platform and the back-end storage system of the Webshop can be event-driven (taken place for each shipment), occur periodically, or is foreseen only at a specified time(s) of the day.

The communication platform then passes all data needed for the transport of the shipments; as a result it receives the graphical data for the parcels labels, to be prepared accordingly to the standards used by the specific (last stage) carriers.

In the next step, when the palettes with the labeled parcels arrive from Asia, the fulfillment company also uses the software platform for the preparation of printed lists (manifests) of readily available parcels for particular carriers and sending the information to the carriers that the particular parcels are ready to a pick-up.

This part of interaction between the shipment software and the carrier firm often takes place by using the EDI standard; nevertheless the carriers use the EDI standard fitted to the carriers' specifications.

The decision of the carrier choice will be taken by the software platform operated by the 4PL, or it can be taken by the multi-carrier shipping software application integrated with the platform. At this stage various information can be taken into account for choosing the most appropriate carrier. Not only costs determine the choice, also other factors like: type of the packet (e.g. standard or express), its weight and dimensions (size), the destination country, the delivery on specific days (daytimes), paying options, possibility of tracking and tracing, and so on.

Additionally, in case of using different carriers, e.g. the ocean carriers, the supplementary inputs like the weather data provided from Web applications could be needed to support the transportation decisions. Often also the geo data considering the road traffic needs to be integrated into the data of the 4PL software systems.

As described by the above scenario, we can see that there is a need to transform the data or customize it to different data formats expected by diverse systems. Therefore in the next Section we propose the integration of logistic data by using a common standard.

V. Value chain data integration with Linked Data

As stated at the beginning of the paper, the process steps in a workflow could undertake numerous transformations of data. In our example given in the last Section there are the geographic, technological, linguistic, syntactical and semantic transformations needed. The common format could alleviate the communication between the participants collaborating in the process environments.

Thus, it appears that a common solution for the data integration is appealing and desirable. In our example the source data attained from a Webshop until delivery of the commodities to the customer is administered in the subsequent stages, changing format and being adjusted and enriched through numerous additional transformations needed to accomplishing of their activities (process steps) by the participants in a joint business venture.

The contemporary evolution in Web comes toward the Semantic Web. The Semantic Web concept supports the basic idea of the Web considered as an open community sharing the information around the world. Nowadays problems comprise of integration and consistency of information, what is a result from the fact that the Web at present is concentrated on the data presentation. What we need in the context of BPM is the integrated description of objects and their relationships. The data model for the Semantic Web infrastructure promoted by the W3C consortium is the Resource Description Framework (RDF).

As pointed out from our hypothetic Webshop parcels example, one part of the data integrated into the flow of data exchange in business networks could be the data supplied into the 4PL integrator software platform from Web applications like the Geo, metrological or traffic data, partially enriched with semantic information described with RDF triples.

Consequently, we analyze the possibility of the integration of data exchanged between the business venture participants on the base of semantic data. There are several approaches to implement the usage of the ontology [17] for the communication in Web, but there are still no known established business solutions working successfully on the base of ontologies.

For that reason we consider the application of Linked Data concepts as more of a lightweight solution to the semantic description of data exchanged in networks connected through Internet and enriched with data from web applications like OpenStreetMap [18] or DBpedia [19]. We focus on the further development of linked data supported by the W3C consortium as Liked Open Data with the goal to provide a distributable, not only human-readable but machine-readable description of the data.

As described in Section 3, the Linked Data principles encompass the following: HTTP URIs as names for entities, usage of standards RDF* (SPARQL) and links to other URIs (of linked data).

The information assumed is to be presented as Linked Open Data, this presumes also that the data is available on the Web with an open license, it is machine-readable and not in a proprietary format.

Therefore it is important that the communication software (e.g. of 3PL) will support open data formats, such as CSV (Comma-separated values) [20] or transformation to such an open format.

The data exchanged between the choreography's participants may be enriched with semantical information by means of the RDF graphs. At present time information is stored mostly in relational databases. For transformation of the data some solutions are partially available, i.e. Triplify for transforming the data stored in transactional databases into RDF representations [21]. Other possibilities like object serialization, or hierarchical representation should be mapped into the graph data models. Meanwhile there are multiple semantic databases implementation known, such as Triplestores, a purpose-built database type dedicated for the storage and retrieval of triples (subject-predicate-object), e.g. Virtuoso [22]. In addition to queries the triples can be imported or exported using RDF and other formats.

The mapping between the customized IT solutions and other data formats into the triple representation could be undertaken by means of dedicated software.

The technical aspects of data integration are in fact beyond the focus of this paper, but the new software trends in the information technology apply numerous reuse techniques and approaches to extensive automation and adjustment of software with generative techniques [23], and variability management known from the software product line approach, as described in [24].

The usage of open formats with RDF-defined semantic could support easier entries in the digital value chains. The enrichment of data with the semantic information can help communicate and mediate between multiple viewpoints. The semantic enriched data stored in an open format can be made widely available for the participants of the value chains if it could be further managed by using of cloud computing – the web-based, dynamical IT services. The cloud computing solutions warrant moreover the security on the infrastructure and on the data level, and also would eliminate the need of initial investments in IT infrastructures.

The usage of open formats may considerably contribute to raise flexibility and content transfer of business process choreographies in supply chains organized as webs and simplify the transformation of data into the diverse e-business standards and enterprise applications data.

The drawback of the given approach is the need of its integration into IT solutions. To be useful it should be supported by the numerous diverse E-business standards coming from non-profit organizations like Electronic Data Interchange EDI, the United Nations Electronic Data Interchange for Administration, Commerce and Transport UN/EDIFACT, Business Message Standards from the GS1-organization especially GS1 XML, Electronic Data Interchange Forwarding EDIFOR, web service business protocols, other standards specific for specialized manufacturing branches, and also the enterprise applications.

VI. Conclusion

The concepts of the Semantic Web and further the Linked Data are time and again regarded as the Web 3.0 extensions to the existing Web technology. It is expected that their application may substantially support the automated extraction of information published on the Web by using open standards and additionally describing with semantic meaning and the contextual relationships of the data. The contemporary environment of the inter-organizational business processes working together in fulfillment of the distributed business scenarios do not incorporate the above mentioned extensions.

In the paper we have considered incorporating the semantic concepts associated with the Linked Data into the business process management BPM, especially for the aims of building the interaction models of the business process choreographies in value chains.

In the paper we have shown on the 4PL integrator example a need for a common standard for data exchange between the various IT applications interacting in the logistics value chains.

We suggest the usage of the Linked Data to contribute in speeding up of the data integration and data transfer. The saving of costs (previously needed for the transformations of data from and to diverse formats) will create the added value for the network participants.

The suggested solution would additionally allow the describing semantics of the data to be stored and processed. It also raises new possibilities (solutions) for the scalability issues of the value webs in logistics. The network effect causes that with increased number of participations the added value for the participants of the network grows.

There are different possibilities for modeling the choreographies in business value networks. As a notation for the business processes collaborations we suggest the usage of the OMG industrial standard in business process modeling - Business Process Model and Notation (BPMN) Version 2.0.

In the future the integration of the Linked Data concepts into the BPMN choreographies may be investigated.

References

[1] J. C. Jarillo, "On strategic networks", in *Strategic Management Journal*, 9, 1988.

[2] W3C LinkedData, 2011, www.w3.org/wiki/LinkedData

[3] OMG Consortium, *Documents Associated with Business Process Model and Notation (BPMN) Version 2.0*, Release date: January 2011, ww.omg.org/spec/BPMN/2.0

[4] M. P. Papazoglou and P. M. A Ribbes, *E-business: organizational and technical foundations, John Wiley and sons*. London 2006, pp.88-90.

[5] F. Soliman, "Optimal level of process mapping and least cost business process reengineering," *International Journal of Operations & Production Management*, 18(9): pp. 810-816.

[6] Web Services Glossary, W3C Working Group Note 11 February 2004, http://www.w3.org/TR/ws-gloss

[7] N. Kavantzas, D. Burdett, G. Ritzinger, T. Fletcher, and Y. Lafon, *Web Services Choreography Description Language Version 1.0*, W3C Candidate Recommendation, November 2005.

[8] J. M. Zaha, M. Dumas, M. Hofstede, A. Barros, and G. Decker, "Service interaction modeling: bridging global and local views", In *Proceedings of 10th IEEE International EDOC Conference* (EDOC 2006), Hong Kong, 2006.

[9] G. Decker, and F. Puhlmann, "Extending BPMN for modeling complex choreographies", In R. Meersman, and Z. Tari et al. (Eds.): OTM 2007, Part I, *Lecture Notes in Computer Science LNCS* 4803, pp. 24-40, Springer, 2007.

[10] T. Berners-Lee, "On the next Web", Talk on the TED Conference 2009, www.ted.com/talks/tim_berners_lee_on_the_next_web.html

[11] W3C Architecture domain, Naming and Addressing: URIs, URLs, …, http://www.w3.org/Addressing/

[12] W3C Semantic Web, RDF Working Group, *Resource Description Framework (RDF)*, 2004, www.w3.org/RDF

[13] H. Baumgarten, *Das beste der Logistik*. Springer Verlag, Berlin 2008.

[14] S. Chopra, and P. Meindl, "*Supply chain management: strategy planning, and operation*", 3rd ed., Prentice Hall, 2007, pp.427.

[15] A. Matopoulos, and E. -M. Papadopoulou, "The evolution of logistics service providers and the role of Internet-based applications in facilitating global operations" in *Enterprise Networks and Logistics for Agile Manufacturing*, L. Wang, and S.C.L. Koh, Eds., Springer, 2007, pp. 298-304.

[16] M. Kantor, and J. Burrows, *Electronic Data Interchange (EDI)*, National Institute of Standards and Technology, 1996.

[17] D. Allemang, and J. Hendler, *Semantic Web for the working ontologist modeling in RDF, RDFS and OWL*. Morgan Kaufman, 2008.

[18] OpenStreetMap Project, available at www.openStreetMap.org

[19] DBpedia Project, Free University of Berlin, and University of Leipzig, OpenLink Software, http://wiki.dbpedia.org/About

[20] MastPoint, *CSV-1203, CSV File Format Specification*, Best practice for business-to-business operations, available at http://mastpoint.curzonnassau.com/csv-1203/index.html

[21] Triplify expose semantics!, http://www.triplify.org/Overview

[22] Virtuoso Universal Server, Universal Server Platform for Enterprise Data Integration, Web Services, & Process Orchestration, openLink software, available at http://virtuoso.openlinksw.com/

[23] K. Czarnecki, and U. W. Eisenecker, *Generative Programming: methods, tools, and applications*". Addison Wesley, 2000.

[24] S. Robak, *Contribution to the Improvement of the Software Development Process for Product Families*. Oficyna Wydawnicza Uniwersytetu Zielonogórskiego, Zielona Góra, Poland 2006.