BUSTECH 2017

The Seventh International Conference on Business Intelligence and Technology

ISBN: 978-1-61208-536-4

February 19 - 23, 2017

Athens, Greece

BUSTECH 2017 Editors

Malgorzata Pankowska, University of Economics in Katowice, Poland

BUSTECH 2017

Forward

The Seventh International Conference on Business Intelligence and Technology (BUSTECH 2017), held between February 19-23, 2017 in Athens, Greece, continues a series of events covering topics related to business process management and intelligence, integration and interoperability of different approaches, technology-oriented business solutions and specific features to be considered in business/technology development.

We take here the opportunity to warmly thank all the members of the BUSTECH 2017 technical program committee, as well as all the reviewers. We also kindly thank all the authors that dedicated much of their time and effort to contribute to BUSTECH 2017. We truly believe that, thanks to all these efforts, the final conference program consisted of top quality contributions.

We also gratefully thank the members of the BUSTECH 2017 organizing committee for their help in handling the logistics and for their work that made this professional meeting a success.

We hope that BUSTECH 2017 was a successful international forum for the exchange of ideas and results between academia and industry and to promote further progress in the area of business intelligence and technology. We also hope that Athens, Greece provided a pleasant environment during the conference and everyone saved some time to enjoy the charm of the city.

BUSTECH 2017 Committee

BUSTECH 2017 Steering Committee

Malgorzata Pankowska, University of Economics in Katowice, Poland
Hermann Kaindl, Vienna University of Technology, Austria
Jürgen Sauer, Universität Oldenburg, Germany
Pierre Hadaya, ESG-UQAM, Canada

BUSTECH 2017 Industry/Research Advisory Committee

Stefanos Vrochidis, Information Technologies Institute - Centre for Research and Technology Hellas, Greece
Anbang Xu, IBM Research - Almaden, USA
Hans-Friedrich Witschel, University of Applied Sciences and Arts, Northwestern Switzerland
Oscar Ferrandez-Escamez, Nuance Communications Inc., USA
Silvia Biasotti, CNR - IMATI, Italy

BUSTECH 2017

Committee

BUSTECH Steering Committee
Malgorzata Pankowska, University of Economics in Katowice, Poland
Hermann Kaindl, Vienna University of Technology, Austria
Jürgen Sauer, Universität Oldenburg, Germany
Pierre Hadaya, ESG-UQAM, Canada

BUSTECH 2017 Industry/Research Advisory Committee
Stefanos Vrochidis, Information Technologies Institute - Centre for Research and Technology Hellas, Greece
Anbang Xu, IBM Research - Almaden, USA
Hans-Friedrich Witschel, University of Applied Sciences and Arts, Northwestern Switzerland
Oscar Ferrandez-Escamez, Nuance Communications Inc., USA
Silvia Biasotti, CNR - IMATI, Italy

BUSTECH 2017 Technical Program Committee

Bernardo Almada-Lobo, INESC-TEC | Porto University, Portugal
Faten Atigui, Conservatoire National des Arts et Métiers, France
Seyed-Mehdi-Reza Beheshti, University of New South Wales, Sydney, Australia
Silvia Biasotti, CNR – IMATI, Italy
Peter Bollen, Maastricht University, Netherlands
Albertas Caplinskas, Vilnius University, Lithuania
Giuseppe A. Di Lucca, University of Sannio - RCOST (Research Center on Software Technology), Italy
Johannes Edler, University of Applied Sciences Upper Austria Campus Hagenberg, Austria
Oscar Ferrandez-Escamez, Nuance Communications Inc., USA
Foteini Grivokostopoulou, University of Patras, Greece
Pierre Hadaya, ESG-UQAM, Canada
Ioannis Hatzilygeroudis, University of Patras, Greece
Hermann Kaindl, Vienna University of Technology, Austria
Franck Le Gall, Easy Global Market, France
Haim Levkowitz, UMass Lowell, USA
Wenbin Li, Easy Global Market, France
Daniel Lübke, innoQ, Switzerland / Leibniz Universität Hannover/FG, Germany
Malgorzata Pankowska, University of Economics in Katowice, Poland
Andreas Pashalidis, BSI, Belgium
Isidoros Perikos, University of Patras, Greece
Erwin Pesch, University in Siegen, Germany

Table of Contents

Non-formal Education Architecture Modeling

Malgorzata Pankowska
Department of Informatics
University of Economics in Katowice
Katowice, Poland
email: pank@ue.katowice.pl

***Abstract*—Nowadays, professional life, education and leisure time are characterized by increasing change and diversity. People with different skills, knowledge, cultural background, cognitive and psychological abilities prefer learning in an unique way, because Information Communication Technology (ICT) environment provides new opportunities for them. The paper covers discussion on formal, informal and non-formal education forms, which are the lifelong learning notions as well as ideas of the learning society. The main goal of the paper is to explain the non-formal learning concept and its system architecture model as an additional educational option at university level.**

Keywords-enterprise architecture; non-formal education; ArchiMate; university learning; system modeling.

I. Introduction

The lifelong learning approach focuses on fostering interchange of knowledge among educational institutions and on constant modernization of institutional solutions in the areas of management and financing higher education. The approach supports innovativeness of education and reinforces social cohesion of different university governance systems in different countries. It is oriented towards openness, peering, sharing and acting globally. Openness of education is associated with transparency, flexibility, expensiveness, engagement and free accessibility. Peering means that people are involved in the process of self-development. They are self-organizing to learn, design goods and services, create knowledge and produce shared experiences [18]. The peer activities rely entirely on self-support, egalitarian communities of individuals, who come together voluntarily to produce a shared outcome. Traditionally, people argue that when they are acting globally, they should control and protect proprietary resources and innovations through patents, copyright and trade marks [8]. However, digital products are easy to share, mix and replicate, therefore, new models of intellectual property governance are needed [6]. Internet permits acting globally to monitor e-learning opportunities offered by different institutions and it enables tapping into global knowledge resources [15]. The paper covers analysis of non-formal education. Although the most common way is formal education supported by Web 2.0 education, in this paper formal, non-formal and informal education forms are considered as complementary one to another [19]. The paper is not oriented towards comparison of these forms, although they all are provided and managed by university staff. The next part of the paper covers discussion on andragogy, heutagogy, formal, informal and non-formal education. To present the non-formal education system architecture model, the third part includes a short presentation of the enterprise architecture definition and standard. The last part of the paper contains an architectural model of non-formal education. The model is visualized in ArchiMate 4.0 beta version language.

II. Formal, Informal and Non-formal Education

In the learning process, participants obtain recommendations for learning goals from the system in three ways:

- formal learning, where learning goals are generated from the domain knowledge,
- informal learning, where the recommendations are provided by the Web 2.0 community of learners,
- non-formal education including random suggestions on learning goals, loosely related to formal education goals, but enabling the knowledge development.

The distinction among that three forms is not only an administrative point of view. Formal education is linked with schools and training institutions, non-formal with community groups and other organizations, and informal covers interaction with friends and work colleagues. All these forms accompany human beings from birth to death; however, the perception of their value is changing.

Particularly important in this paper, adult learners are coming in the educational process with concrete and immediate goals, e.g., professional, social and personal development. Learners have specific expectations from the learning process and when this process meets their expectations, then their motivation for learning increases. They have a wide spectrum of prior knowledge and specific life experiences, so they prefer the educational institutions to be related to these experiences. In the context of informal and non-formal education for adults, the heutagogy development seems to be useful. According to Blaschke [2], heutagogy is defined as self-determined learning rooted in andragogy. In that approach, learners are highly autonomous and self-organized. There is a focus on development of student capacity and capability with the goal of teaching students who are well-prepared to work in a complex knowledge environment. The heutagogical approach is considered as a theoretical background to be applied to emergent technologies in distance education and for steering distance education practice. Heutagogy is characterized by learner control and self-responsibility in learning, so students are defining their objectives of learning, they have intrinsic motivation, and they are able to incorporate their experiences.

In formal learning processes, in European Union (EU) countries, schools and universities are required to respect the Bologna Process requirements concerning the university education. Therefore, each university ought to implement European Credit Transfer System (ECTS), European Qualification Framework (EQF) and National Qualifications Framework (NQF). NQF is an instrument for the classification of qualifications according to a set of criteria for specified levels of learning. The Bologna Process requirements were implemented to improve the transparency, access, progress and quality of qualification in relation to the demand on the labor market [7]. The basic concept in formal education, i.e., competence is defined as knowledge, skills and attitudes. In the context of EQF, competence should be described in terms of responsibility and autonomy. In the Bologna Process learning environment, the student learning outcomes (SLOs) are also the important drivers of the educational process and as such they require evidence. The SLOs describe what a student is expected to learn as a result of participating in academic activities [4]. Beyond SLOs, student progress outcomes (SPOs) are implemented to reflect student progress in course sequence and in degree programs.

In contrast to formal learning, informal learning is organized by students. It has no objectives in terms of learning outcomes or acquisition of any competences. It includes socialization, support, gathering opinions, consultancy, and self-directed learning. In contrast with the traditional view of teacher-centered learning via knowledge acquisition, informal learning is peer-to-peer learning. So, students read self-selected books and e-books, participate in self-study programs, watch YouTube videos, navigate Internet support materials, seeking advices from peers, participate in virtual communities of practice. Informal learning occurs in community, even if participants only observe, play or take part in social events. In informal learning process, students do not receive grades nor certificates of completion. There are other important opportunities, i.e., opportunity to listen the lectures provided by a famous professor or expert. Informal learning is a way to globalization of education, because of the open access to the same course materials and e-books in all the world.

Non-formal learning at university covers various less-structured learning events, such as night university visiting, open lectures, community sport events, conferences, seminars, summer schools, and company visits. That forms of education do not have curricula, accreditation or certification as it is in formal learning, but they are more structured than informal learning approaches. Non-formal learning is a method of communicating with people, of motivating them to participate, and of helping them to acquire the necessary skills. In non-formal education process, knowledge is developed by practice rather than by lecturing. It is strongly based on volunteer works and voluntary participation. It is costly, but the costs can be considered as promotion and marketing costs, funded by sponsors.

The non-formal education covers seminars for vocational and technical skills development, individual and group instruction in functional literacy, community assemblies, forums, work conferences, for example on taxation, fire prevention, drug abuse prevention, or on security. That education focuses on propagation of desirable values, customs, and traditions. The goals of non-formal education is to increase formal education participation rate, exchange of views and experiences, using alternative learning schemes, like street schools, involving street children in the education process.

The Danish non-formal education system is one of the oldest in Europe, because it is known since the 19th century. It is based on the concept of "*Folkeoplysning*" introduced by the Danish educational philosopher, N.F.S. Grundtvig. The concept means sociocultural activities, youth and adult learning, and folk education. Another challenge for non-formal education was the use of the new technologies in group work. The ICT provided many opportunities for individual learning, as well as for group work. The Open University in the United Kingdom, the Centre National d'Enseignement a Distance in France and the Universidad Nacional de Educacion a Distancia in Spain produce a large quantity of multimedia educational material. Video conferencing and video seminars on the Internet are examples of the new technologies applied in non-formal education. The new methods of communication allow organizations from different countries to establish closer contacts with the aim of exchanging ideas and materials.

Non-formal education includes activities developed under Corporate Social Responsibility (CSR), which refers to business practices involving initiatives that benefit society [14]. Therefore, an organization's internal practices can influence their employees, customers, partners and environment in a positive manner. Business organizations can take that approach to deliver better education services and in this way they improve their positive image among customers and for society. For example, University Social Responsibility (USR) can increase graduate involvement in university problem solving. The USR is to enable access to joint event funding opportunities, to enhance the university influence in the industry and in the community, and to differentiate the university from its competitors, i.e., other universities.

III. ENTERPRISE ARCHITECTURE OF NON-FORMAL EDUCATION

The term "enterprise" can be considered as an overall concept to identify a company, business organization, university or governmental institution [10]. An enterprise is defined as a social entity, which is going to achieve a certain goal [12]. For an enterprise, architectural framework as a conceptual structure related to a certain system type consists of areas of concern and a necessary and sufficient set of design domains.

The ISO/IEC 42010:2007 shows that architecture is the fundamental organization of a system, embodied in its components, their relationships to each other and the environment, and the principles governing its design and evolution [12]. The goal of enterprise architecture (EA) is to create a unified ICT environment across the firm or all of the firm's business units with links to the business side of the organization, to promote alignment, standardization, reuse of existing ICT assets, and the sharing of common methods for project management and software development across the organization [17]. The EA provides a holistic expression of the enterprise's strategies and their impact on business functions and processes, taking the firm's sourcing goals into explicit consideration [9]. The EA gives the user an opportunity of faster delivery of new functionalities and modifications of applications, as well as it enables an easier access to higher quality, more consistent and more reliable information [16]. The EA identifies opportunities for integration and reuse of IT resources and prevents the development of inconsistent processes and information [11]. Especially important to users is the capability of integrating the information among applications and across data warehouses and data marts [13].

As it was mentioned at the beginning, the goal of the paper is to visualize the non-formal education architecture. According to Dumitrescu, non-formal education is a partner in the lifelong learning process [5]. It is strongly based on external sponsoring by the local, national and international institutions. The non-formal education activities can be realized outside, as well as inside the formal education institutions. In the last case, the educational activities are governed by the school managers. The non-formal education can be also carried out within companies, by professional associations, or by self-motivated individuals. In the architecture model presented in Figure 1, non-formal education is assumed to be realized within university organization, so university staff and students are involved in the educational processes as organizers as well as a beneficiary of the educational process results. Presented in the paper, non-formal education model is visualized in ArchiMate 4.0 beta version language, therefore some additional explanations must be added. The non-formal education is assumed to be realized as a system of projects and the project management methods can be applied for them. The ArchiMate language allows only for a very general outline of business processes, therefore, the non-formal education business process model is visualized in BPMN language Bizagi tool (see Fig. 2). The ArchiMate language is an open, independent, and general modeling language for enterprise architecture. The primary focus of ArchiMate language is to support stakeholders to address concerns regarding their business and the ICT systems. The ArchiMate metamodel consists of three layers: the Business layer, the Application layer and the Technology layer. According to the EA model, the technology supports usage of applications, which in turn support the business. In this model, non-formal education is considered as a system of events, which are realized for local community. The system of events covers seminars, workshops, performance, exhibitions, excursions, conferences, meetings, community assemblies, forums, demonstration classes, and sport competitions. The audience of the event is selected according to the event project goals. Some events are organized for children, other for older people or for teenagers. The events are organized by university staff and students, who are working as volunteers. If it is necessary, the events are financially supported by external institutions. Anyway, for each event the coordinator is selected and responsible for the event tasks realization for the final success. Each event can be managed as a project, therefore the canvas model can be specified for the non-formal education events (see Fig. 1).

<table>
<tr>
<td rowspan="2">Key Partnerships
Sponsors;
ICT solution providers;
Web service developers;
Event facilitators;
Academic & Administrative Staff;
Web portal administration;</td>
<td>Key Activities
Event management;
Web portal requirement engineering;
Event realization, control & evaluation; Event scheduling</td>
<td rowspan="2">Value Propositions
Values in public services process;
Life style changes;
Long-life learning habits; Learner satisfaction; Non-formal education acceptance, efficiency & effectiveness</td>
<td>Customer Relationships
Analytics of educational services; Relations among learners & academic staff</td>
<td rowspan="2">Customer Segments
Learners; University Academic Staff; Students; Learner associations & assemblies; Political parties; Governmental institutions;</td>
</tr>
<tr>
<td>Key Resources
Software & hardware; Donation support; Non-formal education event documentation archives;</td>
<td>Channels
Non-formal education promotion portal; Chat room, forums & blogs; Websites of mentors & associations;
Printing publicity & emailing;</td>
</tr>
<tr>
<td colspan="2">Cost Structure
Web portal development, implementation, administration, & maintenance; Catering, transport & hotel costs; Event promotion costs;</td>
<td colspan="3">Revenue Streams
Learning time reduction; Learning process simplification; Participant satisfaction; Social relation development; Life status improvement; lower effort & lower costs of formal education; University Social Responsibility development; University - Business Alignment;</td>
</tr>
</table>

Figure 1. Business Canvas Model for Non-Formal Education Development.

In general, the business canvas model can be the first step in the discussion on non-formal education strategy development as well as on the management of educational events. There are many different ways of categorizing or

 ISBN: 978-1-61208-536-4

grouping events, including by size, form and content [3]. The events may have characteristics similar to project dimensions. i.e., learning objectives, time, financial and human resources, location, risk, benefits, long-term impact, audience, publicity, promoting. The event management process is modelled in Figure 2. Similarly, as for projects, the event life cycle consists of the following activities: organizing, analysis, identification, evaluation, design, promotion, realization, shutdown.

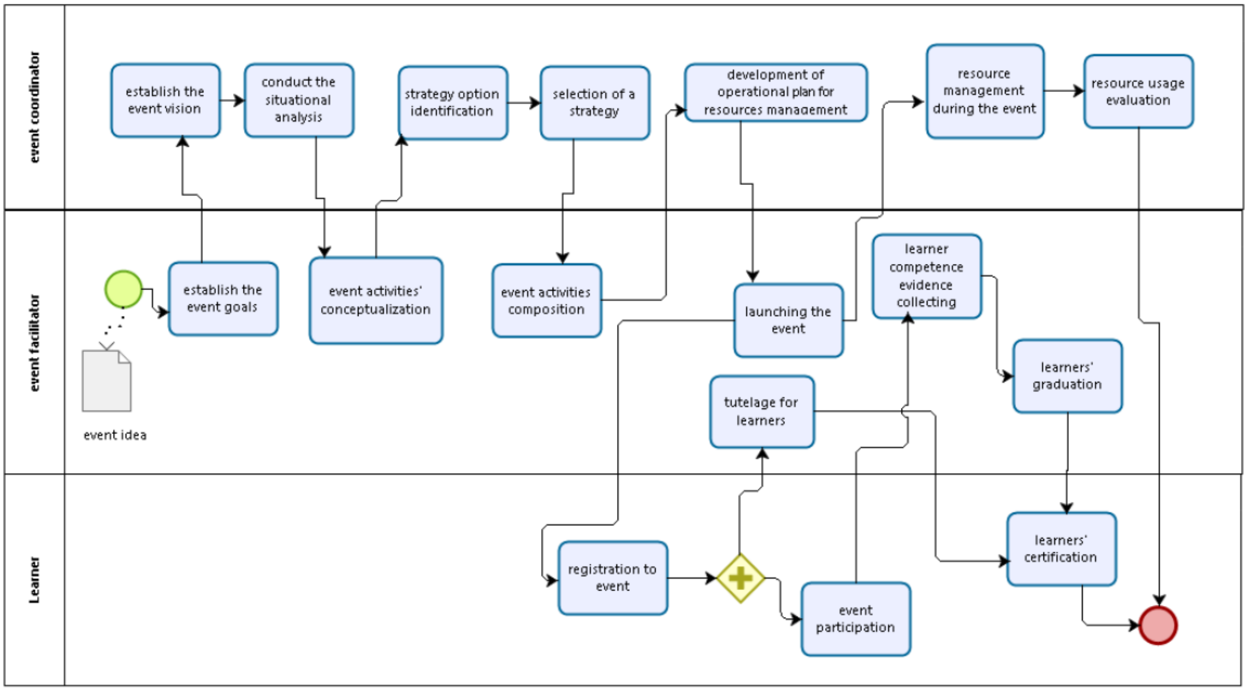

Figure 2. University Non-Formal Education Process Model.

In the proposed EA model, the exceptional role of learning facilitator is specified. The role covers:

- Giving information and opinions.
- Seeking information and opinions.
- Proposing goals and suggesting ways of initiating actions.
- Giving directions and developing plans on how to proceed.
- Summarizing related ideas, suggestions and major points discussing.
- Linking ideas and activities by relating them to each other.
- Examining the feasibility and workability of ideas, evaluation of alternative solutions, and applying them to real situations to see how they will work.
- Encouraging everyone to participate, and helping in communication among event team members.
- Monitoring and supporting the non-formal education process by which the group is working.
- Helping the team members be aware of standards and goals of non-formal education.
- Building trust, reducing the risk aversion, and encouraging individuality.
- Persuading and supporting people to reconcile disagreements.

In the communication processes, facilitators are not oriented towards argumentation, but rather on creation of mutual understanding. However, for successful event realization the coordinator is needed. That role is similar to the role of project manager [1].

Beyond the process model in Figue 2 , in Figure 3 the non-formal education architecture model is available. The ArchiMate model covers the following layers:

- BUSINESS containing the following elements:
  - Business actor (i.e., Participant, Event Coordinator).
  - Business role (i.e., Patron of Plans & Programs of Events).
  - Business service (i.e., Program & Events Planning, Event Cost Management, Non-formal Education Management, Events' Collecting in Programs, SLOs Specification for Event, SLOs Specification for Programs, Participant Enrolment Controlling, Event Outcomes Controlling, Event Management Support Service).
  - Business object (i.e., Student Evaluation Protocols, Event Description Card, Library Report, Event Evaluation).
- APPLICATION including the following elements:
  - Application components: University Event Politics and Regulations, Participant

Enrolment System, Participant Evaluation's Registration System, SLOs Event Registration System, Event Controlling System, Student's Portfolio Registration System, Information Technology Support, Library Management System.

- TECHNOLOGY covering as follows:
  - Node (i.e., Application Server, Data Server).
  - System Software: Participant Enrollment and Event Base, University Library, Learning Event Base, Student Management System, Learning Course Controlling System, Information Technology (IT) Management System.
- MOTIVATION including the elements:
  - Stakeholder (i.e., Event Participant, Sponsor, Facilitator, Innovator, Employer).
  - Driver (i.e., Event Participation, Knowledge Management and Dissemination).
  - Principle (i.e., Guides for Event Organizers).
  - Assessment (i.e., University Accreditation Commission Assessment).
  - Requirement (i.e., Programs, Plans and Event Proposals).
  - Constraint (i.e., Competition among events, National Legal Acts).
  - Goals (i.e., Participant Satisfaction, Appropriate Competencies for the market position).
  - Deliverable (i.e., Event Guidelines and Description, Student Portfolio, Participant Enrollment Report).

Taking into account the above specification, some comments should be added. Student Learning Outcomes (SLOs) in non-formal education concern mostly social competencies. Students participating as volunteers in events receive certificates as evidence of the engagement. The general non-formal education process includes a sequence of sub-processes (see Fig. 2). All the sub-processes are realized to ensure satisfaction of participants of the events. The non-formal education model in Fig. 3 includes a general vision of the ICT systems for the education management support. The model permits on a holistic view of problems and it can be utilized in an iterative approach for non-formal education improvement. The model can be useful to emphasize the alignment issues among the non-formal education staff requirements and the ICT providers' solutions. For making the good investments, the EA developers can use the model to discover opportunities, identify options and compare solutions of non-formal education organization. They can identify, which research works will be the best support the non-formal education services.

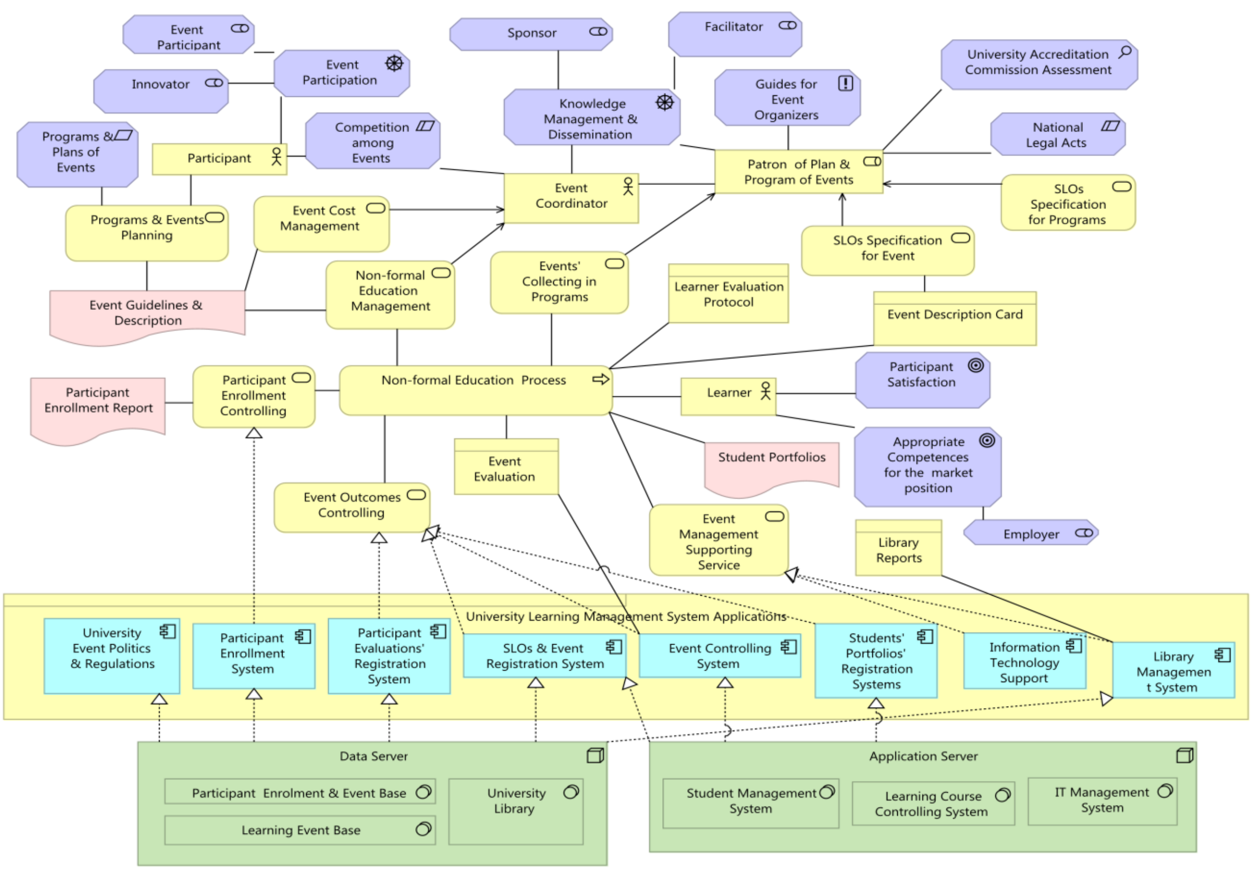

Figure 3. Non-formal Education Architecture Model

ISBN: 978-1-61208-536-4

IV. CONCLUSION AND FUTURE WORKS

Generally, the non-formal education events should be helpful to tackle problems associated with Europe's ageing population. They allow to reject the opinions that separate strategies of knowledge creation and sharing should be developed for students and for adults, although university is promoted as open for all generations of people. Finally, the following conclusions can be specified:

- Formal, informal and non-formal education forms can be developed simultaneously by one educational organization, e.g., university.
- They are not competitive, but rather complementary to one another.
- They all are supported by ICT, however, their development depends on the university priorities:
  - Basic formal education relies on the Bologna Process principles, which are considered as mandatory for university boards.
  - Social networking on the Internet is developed as informal learning.
  - University Social Responsibility is supported by non-formal education.
- Non-formal education is financially supported by third parties and can be considered as a set of events.
- Non-formal education process modelling and its system architecture development can be useful for the events management and auditing.
- Analysis of activities included in the non-formal process model allows for the formalization of that activities
- Non-formal education architecture modelling is to increase a successful repeatability of that events.

The future works will focus on the following topics:

- Profiling of events of non-formal education.
- Planning non-formal education events offline and online.
- Management of events by ICT support.
- Event management methodological approach analysis and improvement opportunities.

REFERENCES

[1] I. F. Alexander, "A Taxonomy of Stakeholders: Human Roles," in: System Development End-User Computing: Concepts, Methodologies, Tools and Applications, S. Clarke, Ed. Hershey, PA: Information Science Reference, pp. 317-350, 2008.

[2] L. M. Blaschke, "Heutagogy and Lifelong Learning: A Review of Heutagogical Practice and Self-Determined Learning," The International Review of Research in Open and Distance Learning, Oldenburg University and University of Maryland University College, vol. 13, no.1, 56-71, 2012.

[3] G. Bowdin, J. Allen, W. O'Toole, R. Harris, and I. McDonnell, Events Management, Amsterdam: Elsevier, 2006.

[4] H. Czarnecki, "KASH - Knowledge, Attitude, Skills & Habits, Auburn Works. Auburn Technical Assistance Center (ATAC), March 2013. [Online]. Available from: http://virtual.auburnworks.org/ profiles/blogs/ kash-knowledge-attitude-skills-habits. 2013.12.20

[5] C. Dumitrescu, "Non-formal education" 15 Dec 1999. [Online]. Available from: http://www.assembly. coe.int/nw/xml/XRef/X2H-Xref-ViewHTML.asp? FileID=8807&lang=en. 2013.12.20

[6] G. Elia and A. Poce, "Future Trends for "i-Learning" Experiences," in Open Networked "i-Learning", Models and Cases of "Next-Gen" Learning, G. Elia and A. Poce, Eds. New York: Springer, pp. 133-176, 2010.

[7] First Glossary, Knowledge Skills And Attitudes (Ksa) Effects, 2010. [Online]. Available from: http:// www.firstebtn.eu/files/FIRST_Glossary.pdf. 2010.01.02

[8] D. L. Gardner, Supply Chain Vector-Methods for Linking the Execution of Global Business Models with Financial Performance, Boca Raton: J.Ross Publishing Inc. 2004.

[9] J. Holt and S. Perry, Modelling Enterprise Architectures. London: The Institution of Engineering and Technology, 2010.

[10] J. A. P. Hoogervorst, Enterprise Governance and Enterprise Engineering, Berlin: Springer, 2009.

[11] J. Horkoff and E. Yu, "Evaluating Goal Achievement in Enterprise Modeling - An Interactive Procedure and Experiences," in The Practice of Enterprise Modeling, A. Persson and J. Stirna, Eds. Laxenburg: Springer, pp. 145-160, 2009.

[12] ISO/IEC 42010, 2011. System and software engineering - Architecture description. International Standard, Geneva.

[13] R. F. van der Lans, Data Virtualization for Business Intelligence Systems, Revolutionizing Data Integration for Data Warehouses, Amsterdam: Elsevier, 2012.

[14] J. D. Martin, J. W. Petty, and J. S. Wallace, Value-Based Management with Corporate Social Responsibility, Oxford: Oxford University Press, 2009.

[15] T. Pfeffer, Virtualization of Universities, Digital Media and the Organization of Higher Education Institutions, New York: Springer, 2012.

[16] Ch. Reynolds, Introduction to Business Architecture Boston: Course Technology, 2010.

[17] J. W. Ross, P. Weill, and D. C. Robertson, Enterprise Architecture as strategy. Creating a foundation for business execution. Boston: Harvard Business School Press, 2006.

[18] M. J. Thannhuber, The Intelligent Enterprise, Theoretical Concepts and Practical Implications. Heidelberg: Physica-Verlag, A Springer Company, 2005.

[19] J. Winterton, F. Delamare-Le Deist, and E. Stringfellow, Typology of knowledge, skills and competences: clarification of the concept and prototype. Office for Official Publications of the European Communities, CEDEFOP, Luxembourg. 2013. [Online]. Available from: http://europa.eu/. 2015.12.22

 ISBN: 978-1-61208-536-4

SOA*dapt*: A Framework for Developing Service-Oriented Multi-Tenant Applications

Sascha Alda, Rüdiger Buck-Emden
Department of Computer Science
Bonn-Rhine-Sieg University of Applied Sciences
Sankt Augustin, Germany
e-mail: sascha.alda@brsu.de, ruediger.buck-emden@brsu.de

***Abstract*— A plethora of architectural patterns and elements for developing service-oriented applications can be gathered from the state-of-the-art. Most of these approaches are merely applicable for single-tenant applications. However, less methodical support is provided for scenarios, in which multiple different tenants with varying requirements access the same application stack concurrently. In order to fill this gap, both novel and existing architectural patterns, architectural elements, as well as fundamental design decisions must be considered and integrated into a framework that leverages the development of multi-tenant application. This paper addresses this demand and presents the SOA*dapt* framework. It promotes the development of adaptable multi-tenant applications based on a service-oriented architecture that is capable of incorporating specific requirements of new tenants in a flexible manner.**

Keywords: Service-Oriented Architecture; Architectural Patterns; Multi-Tenant Application; Adaptation of Software.

I. INTRODUCTION

Service-oriented architectures [1] are nowadays used as a way to encapsulate and to integrate databases and applications being part of an enterprise's software landscape in terms of semantically enriched and re-usable *business services*. These business services can be orchestrated to more meaningful *workflows* that serve as executable software parts of business processes. On top of that, a *user interface* layer enables the involvement of human stakeholders during the execution of a workflow, for instance, to request initial or intermediate user inputs, as well as to represent final outputs.

The success of a service-oriented application depends on a number of factors. One important factor is the accurate modeling of the workflows including the regular flow of the activities and the potential alternate control flows. Another factor is the accurate design of the business services. These factors become even more relevant in application scenarios, in which different independent organizational units, hereafter called *tenants* (e.g., other organizations, subsidiaries of a company, faculties of a university), intend to share the same service-oriented application. The result is a *multi-tenant application* [2], in which different tenants access the same instance of an application's service stack concurrently. This concept is the foundation for latest software consumption models like software as a service (SaaS) [3] leading to high cost-effectiveness for each tenant. However, tenants often require control flows different to standardized, already deployed workflows in order to respect individual requirements. Providing a new workflow model for each tenant would be straightforward, but it breaks the idea of multi-tenancy. In addition, tenants often demand for alternate functionality to already deployed business services or even require completely new services that need to be flexibly deployed. The common approach of extending the original interfaces for each new tenant coming into play results in expanded interfaces and the risk of violating existing dependencies to other tenants.

Although a number of architectural approaches for service-oriented architectures are available [1] [4], none of these offer sufficient support for adapting both workflows and business services according to the needs of a multi-tenant application. Existing patterns and methods [5] [6] for adaptable service-oriented architectures are applicable to single-tenant applications, but are less appropriate to support the adaptation and management of multi-tenant applications.

This paper features the SOA*dapt* framework serving as a guideline for constructing a multi-layered service-oriented architecture that defines the structural decomposition of multi-tenant applications. With respect to this framework, the resulting software architecture offers a layered application stack, which is shared by multiple tenants at the same time. On the business process layer, the framework features a minimal set of *basic workflow patterns* that is suitable to model the various requirements of the tenants' workflows. On a business service layer, three types of business services can be deployed: shared business services suitable to all tenants, business services with dedicated tenant-specific service extensions, as well as fully self-contained services. Service extensions add both additional interfaces and internal behavior to the original service component that have not been anticipated and integrated in its original design. The framework contains further *architectural elements,* such as a business rule engine and a tenant context data registry that completes the architecture. Important *architectural design decisions,* such as a workflow instance model are made. Furthermore, *recommendations* for the *implementation* of the architecture are provided based on modern technologies.

This paper is structured as follows: Section 2 summarizes the related work. Section 3 describes the structure and the principles of the SOA*dapt* framework. Section 4 describes the prototype of the framework and outlines pieces of future work. Section 5 concludes this paper.

 ISBN: 978-1-61208-536-4

II. RELATED WORK

The often cited SPOSAD architecture style by Koziolek [7] provides an abstract perspective on existing multi-tenant applications, such as Force.com, merely discussing the design decisions, as well as the architectural trade-offs related to this style. The resulting multi-tier architectural style is similar to our framework. It features a context-data manager that is responsible for adapting the application logic for tenant-specific business logic and computations. Unlike in the SOA*dapt* framework, no further details are provided how the context-data for a specific tenant is organized and in what way the business logic can actually be adapted for a given tenant. A clear distinction between business service and workflow is not handled in the SPOSAD architectural style.

The work by Mietzner et al. [8] provides fundamental research on instance management for multi-tenancy and describes a set of so-called service tenancy patterns. The pattern catalogue features basic architectural elements, such as the invocation of a service or a process *under tenant context*. The tenant context is actually a piece of context-data provided by a runtime environment that describes the current tenant accessing the application. Tenant-specific customizations of business processes, however, result in the deployment of a new workflow model. Dedicated workflow patterns are not outlined. A similar approach for identifying the current tenant ("tenant context object") can be found in [9]. The SOA*dapt* framework adopts this concept for identifying a tenant and explains precisely, where it should be used.

Further studies on multi-tenant Web applications are presented by Jansen et al. [10] and Bien and Thu [9]. Both papers rely on the MVC architectural pattern for describing the global structure of a multi-tenant application. Further fine-grade models (e.g., class models) can be found. Although service-oriented applications are typically Web-based, these studies can hardly by mapped to the demands of a multi-layered architecture. A workflow perspective is ignored in both studies.

Kabbelijk [11] discusses the adoption of various combinations of multi tenancy patterns to a multi-layered architecture. These patterns are based on the number of instances e.g., of an application server or of a database necessary to serve the tenants. Owing to the rather technical perspective of the architecture, a business-driven workflow layer is omitted. So, no dedicated workflow patterns can be found in his work. However, our workflow instance model can be compared with his work and further properties and constraints of it might be extracted from his contribution.

A multi-tenant approach for business process execution can be found in the article by Pathirage et al. [12]. The authors describe an architecture based on Axis2 runtime environment and the Apache ODE workflow engine. The authors mainly discuss how the runtime environment can guarantee isolation of the running processes. Workflow patterns are not described in the paper either.

Fundamental methods for adaptable software architectures have been elaborated by Svahnberg et al. [13]. He proposes five categories of adaptation methods. Our approach corresponds to the second category "variant component specialization": additional behavior is introduced in the same component or workflow models for different tenants.

A number of academic approaches for adapting (service-oriented architectures) can be found from the state-of-the-art (e.g., [5] [6]). In the majority of cases, these approaches rely on replacing entire components on-the-fly (category one w.r.t. [13]). This approach is straightforward, but leads to dependency issues as elaborated in the introduction section.

In our approach, a (business) rule engine is used to alter the control flow at gateway elements with respect to the demands of the involved tenants. Rule engines are primarily used for evaluating complex rules that cannot be incorporated in a workflow model in a manageable way. We adopt the idea of Doehring et al. [14] to use a rule engine also for control flow management. Doehring's article, however, is not based on the multi-tenancy approach.

III. THE SOA*DAPT* FRAMEWORK

The SOA*dapt* framework introduces architectural elements and decisions, as well as architectural patterns in order to implement adaptable multi-tenant applications. An architectural overview of the framework is given in Fig. 1. The framework is merely based on a layered architecture model. A vertical enterprise service bus (ESB) component provides for a loosely coupled interaction among the components of these layers. In the following, the layers are described in detail. Special attention is drawn to both the business process and the business service layer as both layers consists of the main concepts of the framework. The user interface, as well as the application and data layer are introduced briefly. Implications for an implementation of the concepts are provided as well.

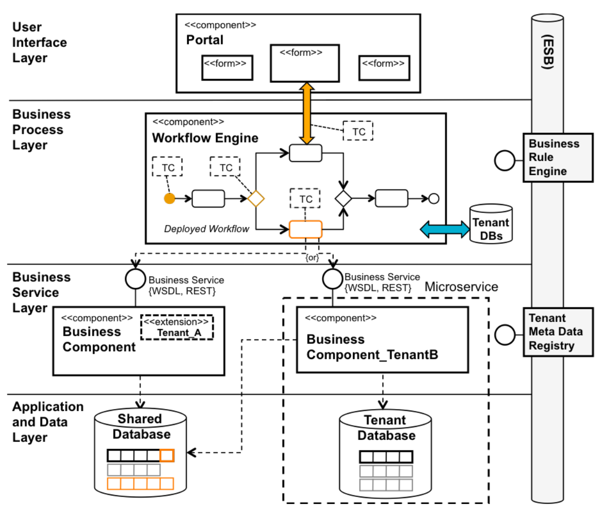

Figure 1. Architecture of the SOA*dapt* framework

A. Business Process Layer

This layer features a workflow engine, where an executable workflow model can be deployed. Although an object-oriented language can implement such a workflow, we assume a modeling language like BPMN 2.0 [15] as the preferred way to implement such workflows. In contrast to a business service, the functional behavior of a workflow model is said to be more comprehensive and may integrate user interactions through an associated portal component (see User Interface layer in Section III C). Workflow models are potentially stateful, that is, they can maintain a state across many workflow steps. Next, further important architectural decisions and elements of that layer will be outlined: the instance model, tenant context object, as well as the minimal set of workflow patterns.

1) Workflow Instance Model

A fundamental architectural property and design decision to be made for a multi-tenant application is the underlying workflow instance model. An overview is given in Fig. 2.

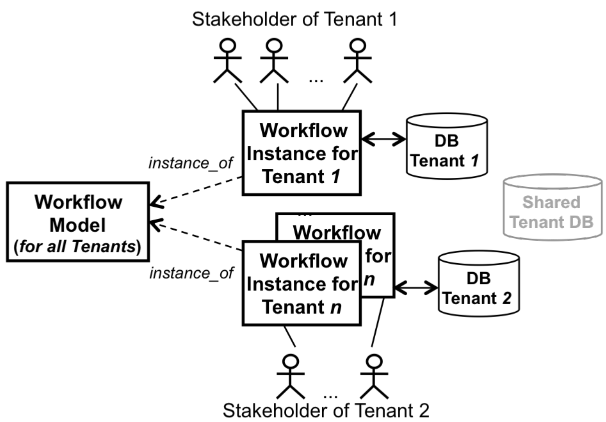

Figure. 2: Workflow instance model

It is assumed that both the structure and the course of actions of a workflow are described as a template that is referred as the *workflow model*. This could for instance be a graphical BPMN 2.0 model of a complaint management process. In a multi-tenancy application, a workflow model is said to be available for *all* tenants. Tenant-specific differences in the course of actions within a workflow model, such as different control flows or additional tasks for handling a complaint can be integrated by means of dedicated workflow patterns (see below). Given a predefined workflow model, an infinite number of *workflow instances* can be generated after the starting events has been fulfilled (e.g., triggered by the request of a user belonging to a tenant).

It is conceivable that in a strict multi-tenant application varying tenants can even share a single workflow instance. However, this approach might lead to typical technical issues as often faced on shared resources, that is, isolation problems of tenant-sensitive data, concurrency issues, or reduced scalability properties. Apart of that, such a shared instance model could hardly be appreciated from a tenant perspective. Henceforth, a single workflow instance is assigned to exactly one tenant. To further reduce the number of workflow instances, stakeholders of the same tenant could share a tenant's workflow instance (upper half of Fig. 2) – assumed that this accords to given tenant governance rules. For more restrictive scenarios, an instance could be generated for each stakeholder belonging to a tenant (lower half of Fig. 2).

The state of a workflow instance (i.e., variables, current execution state) can be stored temporarily in a corresponding tenant database, thus, guaranteeing rigidity and recoverability of the whole application. For statistical analytics, tenant data could also be stored permanently. Again, depending on given governance rules, tenant data could be stored in isolated databases (see Fig. 2). For less sensitive data or data that can trustworthily be shared among tenants (e.g., postal codes or standardized product numbers), shared databases can be integrated (see Section III.D for further details).

A workflow engine should be able to support all possible variations for workflow instance management. For the sake of scalability, separate instances of a given workflow engine could be installed on varying nodes (e.g., in a Cloud or on-premise in a local architecture). Each workflow engine instance could accommodate a dedicated workflow model or a cluster of coherent workflow models. The provision of a new workflow engine instance for each new workflow instance is a theoretical model achieving maximum scalability. However, this must be clearly contemplated from both a management and an economic perspective.

2) The Tenant Context Object

The *tenant context object* is an architectural data element representing the associated tenant of a current user accessing the workflow instance. Likewise to other context objects commonly found in Web frameworks (e.g., HttpSessionContext in the Servlet API), this object must be implemented (or rather: understood) as a global variable that is accessible in all areas of the workflow instance and corresponding objects, such as the service delegate object (see later on). The mapping of a user to a tenant is interpreted as a function that uniquely maps a given user (ID) to a tenant (ID). It is assumed that a user is derived by credentials that are passed in the beginning of the workflow execution, which is then stored e.g., in a session context object. The tenant context object tackles various parts and actors within the architecture as illustrated in Fig. 3:

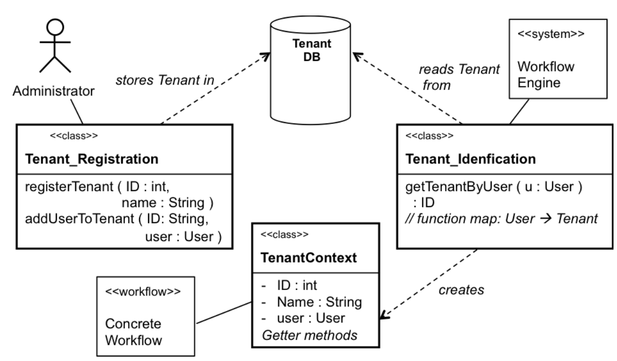

Figure. 3: Structural model of the tenant context

An administrator is able to initially register a tenant in the workflow engine. Depending on the chosen instance model, the tenant's data is stored within a single table or even in a

 ISBN: 978-1-61208-536-4

separate database that guarantees isolation of the tenant's sensitive data. Given a unique tenant ID, the administrator is then able to add users to a tenant. For the purpose of flexibility, it is assumed that users can be added to a tenant even at runtime of the workflow engine. The workflow engine itself can request a tenant ID from a user by accessing class Tenant_Identification. This class will produce the tenant context object. The current workflow instance can access and read the attributes of that object accordingly. Attributes of the tenant context object might be used for debugging purposes, for identifying tenant-specific context information from the tenant meta data registry, or for evaluating the control flow for a tenant (see patterns below in Section III.A.3).

3) *Workflow Patterns*

Our *workflow patterns* describe solutions for recurring situations during the design of a workflow model, where the execution of a workflow might expose a different behavior based on the currently given tenant. In the layered architecture of Fig. 1, these situations are marked by the "tc" symbol surrounded by a dotted rectangle. This rectangle points out that at this stage of the workflow, the invocation of that element is *under tenant context* and, thus, it might vary. The patterns shown next provide an abstract solution and indicate implementation details. While the depicted solutions abstract from a concrete language, the implementation details will be based on BPMN 2.0. The set of patterns is considered as *minimal* and *complete*, that is, more complex and language-specific workflow elements (e.g., compensation, exception handling, sub processes) can be derived easily from this set.

Name of the pattern: Tenant-specific workflow invocation.

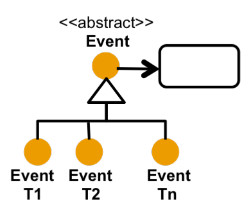

Figure. 4: Structure of the pattern "Tenant-specific workflow invocation"

Problem: The initial enactment of a workflow might vary based on the currently given tenant. That is, tenants might dictate different conditions when a workflow is to be executed. Often, additional tasks are required that have not been regarded in the standardized workflow model.
Solution: Introduce an abstract event that might be extended by concrete events that contain concrete conditions for specific tenants (see. Fig. 4). Assume a polymorphic structure: new concrete events might be introduced and bound to the abstract event at runtime (late binding).
Implementation: The introduction of an event hierarchy would result in an extension to the syntax of the BPMN 2.0 language. Consequently, a workflow engine like Activiti [16] had to be extended as well. As a trade-off solution, different starting events could be modeled and connected to the first activity in a workflow. Both BPMN 2.0 and Activiti do not support a late binding concept.

Portability: BPMN 2.0 features a bunch of different starting and intermediate event types, often leading to complex workflow models. This pattern could necessarily be ported to other events types. The usage should, however, be pondered.

Name of the pattern: Rule-based control Flow

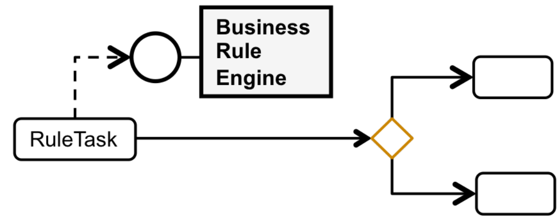

Figure. 5: Structure of pattern "Rule-based control flow"

Problem: Gateways may be used for altering the control flow. Typically, Boolean expressions are used to express conditions on the different branches that must be fulfilled for an execution. In a multi-tenant application, these conditions might differ for any of the involved tenants and, consequently, might exhibit a complex structure. The expression of conditions is not fixed, but is often subject of change (e.g., rules for expressing the credit rating of a client). The adaptation of conditions would result in the re-deployment of the workflow model.
Solution: Prior to the gateway element, a rule task is placed that uses an external rule engine to evaluate the conditions for a given tenant (see. Fig. 5). Conditions can be expressed by means of more descriptive models, such as decision tables [14]. These expressions can be adapted without re-deploying the workflow model, since both rule engine and workflow engine are completely decoupled.
Implementation: BPMN 2.0 already features rule tasks that can be used to evaluate rules from a rule engine. In the Activiti engine, the business rule engine Drools Expert can be used [16]. In order to execute the deployed rules, input variable (the so-called facts) and the result variables need to be specified in the context of this rule task. For evaluating the rules with respect to a given tenant, the tenant context object must be passed as a fact, too. The output variable will contain a list of objects that can then be evaluated at the branches of the corresponding gateway element. Depending on the state of the output variable, the control can be altered according to the demands of a respective tenant. The Drools platform offers tools for the *simple* modification of the rules e.g., within a decision table, which could even be carried out by business units with minor IT-background.
Portability: This concept cannot only be applied to exclusive, but also to inclusive gateways, in which a subset of modeled branches might be invoked concurrently. Although it is assumed that tenants share branches, it is feasible to insert branches that can exclusively be used by a dedicated tenant. Besides, additional tenant-specific activities or sub processes can be integrated into a standard workflow model.
Trade-Off: The more complex the rules and facts in a rule engine become, the more performance is needed for a thorough evaluation. So, the rule engine might result in a bottle-

 ISBN: 978-1-61208-536-4

neck within the whole architecture. Software architects should think of caches to store previous results.

Name of the pattern: Tenant-specific service call

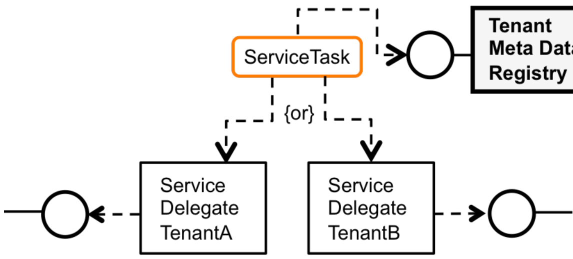

Figure. 6: Structure of pattern "Tenant-specific service call"

Problem: service call tasks implement the actual invocation of business services being placed in the business service layer. Usually, these tasks are passed input data stemming from prior tasks (e.g., a user task requesting data from a user through a portal) that serves as an input to the business service. The resulting output of that service might then be further used in subsequent tasks. Depending on the given tenant, different business services or extensions (see Section III.B) might be invoked from the business layer. Depending on the business service's interface, input and output data has to be converted (e.g., XML to JSON) or enriched (e.g., adding the postal code to a person's profile). Also, the exception handling might be different depending on the nature of the business service (e.g., transactional vs. non-transactional service call). These different service calls together with specific preliminary and subsequent tasks might be implemented by using the "Rule-based control flow" pattern. However, this would blow up the workflow model with too many technical gateway elements and different branches having no real business added value.

Solution: The invoked service tasks delegates the actual service invocation to a so-called service delegate object that is responsible of processing the whole service call including data conversion or enrichment, (remote) service invocation, and exception handling (see. Fig. 6). The correct service delegate object is instantiated by calling the tenant context data registry that stores the corresponding business services per tenant (see III.B for more details). The correct service delegate object is identified based on the current tenant context object.

Implementation: BPMN 2.0 provides service tasks as a core element of the language. However, the language itself does not support the concept of an internal service delegate element. In the Activiti engine, the invocation of a service task is handled by a service handler object, which is actually a Java object implementing a given API. The service handler object can access the input data by using a global context object, the so-called DelegateExecution object [16]. Analogously, output data can be conveyed back to the workflow instance through that context object. In this service handler object, the corresponding service delegate object can be invoked for processing the tenant-specific service call. Developer of the corresponding service handler class can fall back on the complete Java SDK, further related APIs (e.g., JAX-RS for invoking REST-based services), or frameworks (e.g., Zend for data conversion between XML and JSON). This lightweight approach is actually an improvement compared to older development models from BPEL-based engines (e.g., Apache ODE), where service calls need to be graphically bound to WSDL files of the Web Services by some proprietary and tricky development tools.

B. Business Service Layer

This layer features business components that provide business services to the upper layers. Business services provide a business value and can be reused and orchestrated in different workflow models. The corresponding interfaces of business services are described in a language-neutral format (WSDL or REST-based), from which client stubs can be generated in order to access the business service. Business components may wrap underlying applications or databases from the data and Application layer (see III.D). Business components can be implemented in an object-oriented language like Java. The deployment can be done in a conventional Java runtime environment (JRE) or in an application server, such as Axis2 or Glassfish. Business components should be stateless to ensure the scalability of the architecture and the substitutability of components within this layer. For the sake of brevity, Fig. 1 does omit execution environments on the business service layer.

Ideally, all business services can be shared among all available tenants. However, this scenario is rather seldom, since tenants are supposed to demand varying properties on the business services used in their workflows models. In order to respect different scenarios, three categories of business services are supported: (1) shared business services, (2) business services with service extensions, or (3) standalone business services deployed as Microservices. This paper mainly introduces type (2) and sketches the idea of type (3) briefly. Both variants are introduced as architectural patterns.

Name of the pattern: Service extension

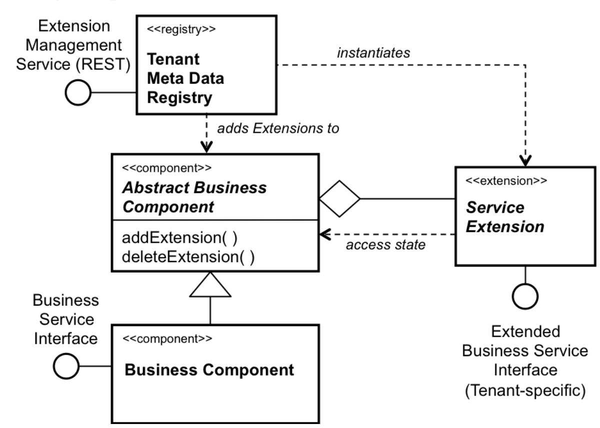

Figure. 7: Structure of pattern "Service extension"

Problem: Tenants might expose varying demands on the usage of a business service. Apparently, it seems unrealistic to anticipate the complete externally visible behavior - i.e., the interface - of a business service that tenants might use in the future. This also applies to internal implementation details within the respective business components. By doing so,

the resulting service interface might be bloated with too many service operations. Integrating new operations into the interface might cause the violation of dependencies to existing components that are coupled to the original interface.
Solution: So-called service extensions can extend the interface of a business component by further service operations that are part of an extended business service. This pattern is an adoption of the extension object pattern [17], in which objects can act as a host for object extensions that can flexibly be added and removed from that hosting object. The structure of the service extension pattern (see Fig. 7) is a slight modification of the original structure as it leaves out various abstractions and concentrates on the relevant elements. An abstract business component serves as the hosting component that provides an administrative interface for adding and removing service extensions. Again, service extensions implement tenant-specific behavior. For the proper execution of this behavior, service extensions have access to the internal state of a business component. Appropriate access rights must be granted accordingly. The business component itself inherits from the abstract business component. Note that this pattern consciously abstracts from concrete implementation techniques, the inheritance relationship just points out the different responsibilities of the involved elements. In modern component models, such as OSGi, the inheritance relationship might be dissolved, thus, resulting in a business component providing both the functional interface of the business service and the administrative interface.

The *tenant meta data registry* is in charge of managing the different service extensions per involved tenant for a dedicated business service component. Besides, further context data, such as form elements (see Section III C) can be added to a tenant. At design time, component assemblers can use the registry for equipping a business service with selected service extensions for a new tenant. During runtime, the application server can use the registry for querying tenant-specific service extensions. Having identified the necessary meta data describing the service extension, concrete service extensions for a tenant can be deployed in the business component. For new service extensions, an upload mechanism for both the meta data and the actual executable of that service extension (e.g., a JAR-file) needs to be provided.

Table. I POSSIBLE QUERIES FOR CONTEXT DATA REGISTRY

Explanation	URI (& HTTP method)
Returns all registered tenants (IDs) of a business component with the id *compID*	GET /comp/[*compID*]/tenants
Returns all registered service extensions (IDs) of a business component with the ID *compID* that are associated with a tenant(ID)	GET /comp/[*compID*]/tenants/ [*tenantID*]/ext/
Registers a new tenant with ID *tenantID* to the business component with ID *compID*.	POST /comp/[*compID*] /tenant/ [*tenantID*]

The registry features a hierarchical model to represent associations among service extensions, business service components, and tenants. Owing to the hierarchical nature of the data model, URIs can be used for identifying the meta data. Consequently, a REST-based interface could be used as the fundament of the extension management service. Table No. 1 shows some example queries that could be applied.

Name of pattern: Business service as Microservice
Problem: Tenants might insist of having self-contained business services that come with their own database and domain model, which conforms to a shared nothing solution. Isolation of data and services is an absolute must criterion.
Solution: Tenants are invited to deploy self-contained business services by means of Microservices [18] into the architecture. Microservices are closed units of deployment with no or a minimal set of dependencies to other services and infrastructure components (e.g., server, databases). Typically, a Microservice has its own domain model, a so-called bounded context. A Microservice contains its own internal application server, such as Glassfish. The deployment can be done in lean execution environments, such as Docker or in cloud-based environments, such as Spring Cloud. The usage of systems like Spring Clouds promotes the scalability of the business service layer and, thus, the entire multi-tenant application. More details on both the theory and implementation of Microservices can be obtained from [18].

C. *User Interface Layer*

This layer consists of user interface (UI) components for involving stakeholders within a workflow execution. In practice, a portal may take over this part allowing the provision of customizable forms. A single form consists of a coherent set of user interface elements, such as buttons or text fields, accomplishing a stakeholder to process data associated with a user task. This data can act as the initial input in the beginning or as an intermediate input during a workflow execution. At the end of the workflow execution, final output data can be displayed. The form rendering heavily depends on the tenant's usability requirements. So, the rendering process and the data exchange between the portal and the workflow engine must occur under tenant context. The specific form elements (e.g., HTML, CSS fragments) should be stored in the tenant context data registry. The portal accesses this registry upon the rendering process for a specific user.

D. *Application and Data Layer*

This layer consists of existing legacy applications and databases. Databases could be based on data models, such as the relational model or variants (e.g., object-relational). A database can be shared among the tenants. For respecting tenant-specific data, tables for each single tenants or schema extensions to common tables can be inserted. The definition of shared data models for multi-tenant applications is not part of this framework. Different approaches how to organize such data models can be found for instance in [2] or [7].

IV. PROTOTYPE, FUTURE WORK

A first prototype of the SOA*dapt* framework has been implemented on top of the Activiti workflow engine [16]. Version 6 of Activiti integrates the proposed workflow instance model (see Section III.A.1) and the tenant context object (see Section III.A.2), which was contributed to the project in the context of a joint master thesis project [19]. The workflow instance model follows the approach of having one instance of a workflow engine that accommodates all generated workflow instances. Each tenant has a separate database, thus, guaranteeing isolation of data.

From the workflow patterns (Section III.A.3), pattern “Rule-based control flow” has been implemented based on Drools expert and on top of the workflow engine Doxis4 BPM [20]. An evaluation in conjunction with the German IT-company SER GmbH – the vendor of Doxis4 BPM - confirmed the flexibility of the approach for having flexible control flows. However, the company criticized the complex user interface of Drools for editing business rule and the overall complexity of the rules themselves. Therefore, future work was recommended for improving the definition of rules. First tests revealed no critical performance issues. An in-depth performance analysis e.g., with stress tests has not been carried out so far, but is considered as future work.

A prototype for the service extension pattern is considered as future work. An ongoing project examines the appropriateness of the server Axis2 for implementing service extensions. The tenant context data registry has been implemented as a first prototype based on the Jersey framework.

V. Conclusion

This paper has introduced the framework SOA*dapt* that can be used for the development of adaptable multi-tenant applications that are based on a multi-layered service-oriented architecture. The framework proposes a set of architectural patterns that allow the adaptation of a multi-tenant application in order to respect varying requirements of the involved tenants. SOA*dapt* also introduces architectural elements for setting up and running a multi-tenant application in a scalable way. First prototypes have been developed.

References

[1] M. Stal, "Using architectural patterns and blueprints for service-oriented architecture", IEEE Software, vol. 23, no. 2, pp. 54-61, 2006.

[2] F. Chong and G. Carraro, "Architecture Strategies for Catching the Long Tail", Available: http://msdn.microsoft.com/en-us/library/aa479069.aspx. April 2006. [retrieved: Jan., 2017]

[3] M. Turner and D. Budgen, "Turning Software into a service", IEEE Computer, vol. 36, no. 10, pp. 38-45, 2003.

[4] T. Erl, "Service-Oriented Architecture: Concepts, Technology, and Design", Prentice Hall, 2005.

[5] R. Mirandola, P. Potena, E. Riccobene, and P. Scandurra, "A Framework for Adapting Service-Oriented Applications based on Functional / Extra-functional Requirements Tradeoffs", 6th International Conference on Software Engineering Advances, Barcelona, Spain, pp. 146-151, 2011.

[6] H. Gomaa, K. Hashimoto, M. Kim, and e. al., "Software Adaptation Patterns for Service-Oriented Architectures", 25th Symposium On Applied Computing, Sierre, pp. 462-469, March 2010.

[7] H. Koziolek, "The SPOSAD Architectural Style for Multi-tenant Software Applications", 9th Working IEEE Conference on Software Architecture, Boulder, USA, pp. 320-327, 2011.

[8] Mietzner R., Unger T., R. Titze, and F. Leymann, "Combining Different Multi-Tenancy Patterns in Service-Oriented Applications", IEEE EDOC Conference, pp. 108-117, 2009.

[9] N. H. Bien and T. D. Thu, "Multi-tenant web application framework architecture pattern", 2nd National Foundation for Science and Technology Development Conference on Information and Computer Science, Vietnam, pp. 40-48, 2015.

[10] S. Jansen, G. Houben, and S. Brinkkemper, "Customization Realization in Multi-Tenant Web Applications: Case Studies from the Library Sector", 10th International Conference on Web Engineering, Vienna, pp. 445-459, 2010.

[11] J. Kabbedijk and S. Jansen, "Multi-tenant Architecture Comparison ", 8th European Conference in Software Architecture, Vienna, pp. 202-209, 2014.

[12] M. Pathirage, S. Perera, I. Kumara, and S. Weerawarana, "A scalable Multi-tenant Architecture for Business Process Executions", IEEE International Conference on Web Services, pp. 21-41, 2011.

[13] M. Svahnberg, J. Grup, and J. Bosch, "A taxonomy of variability realization techniques", Software Practice and Experience, vol. 35, no. 8, pp. 705-754, 2005.

[14] M. Döhring, B. Zimmermann, and E. Godehardt, "Extended Workflow Flexibility using Rule-Based Adaptation Patterns with Eventing Semantics.", Informatik 2010: Service Science - Neue Perspektiven für die Informatik, Leipzig2010.

[15] OMG, "BPMN 2.0 Specification", Available: http://www.omg.org/spec/BPMN/2.0/. 2011. [retrieved: Jan., 2017]

[16] T. Rademakers, "Activiti in Action - Executable business processes in BPMN 2.0", Manning Pub., 2012.

[17] E. Gamma, "The extension objects pattern", Pattern Languages of Programs (Plop) Conference - Writers Workshop, Illinois, USA1996.

[18] E. Wolff, "Microservices: Flexible Software Architectures", CreateSpace Publishing, 2016.

[19] J. Barrez, " Multi-Tenancy with separate database schemas in Activiti", Available: http://www.jorambarrez.be/blog/2015/10/06/multi-tenancy-separate-database-schemas-in-activiti/. 2015. [retrieved: Jan., 2017]

[20] SER, "Business process management with Doxis4", Available: http://www.ser-solutions.com/products-solutions/business-processes.html. 2016. [retrieved: Jan., 2017]

Data Analysis to Better Understand Business Process Models Discovered with Process Mining

Alessandro Bettacchi, Alberto Polzonetti, Barbara Re
Computer Science Division
University of Camerino, Camerino, Italy
e-mail: {*name.surname*}@unicam.it

Knut Hinkelmann
School of Business
FHNW, Olten, Switzerland
e-mail: knut.hinkelmann@fhnw.ch

***Abstract*—Due to continuous changes in the business context, enterprises have to rapidly react to novel market scenarios. To this end, a better understanding of the actual business processes is needed. This was the real need of a manufacturing company producing coffee machines. As-is processes have been investigated to understand in detail how the production chain works. First, we applied process mining techniques which produced models fitting the expectations, but also presenting some deviations from the designed flow of production activity. In order to understand the reason behind such deviations, an in-depth data analysis using On-Line Analytical Processing has been performed. Such awareness allows the management board to re-organize the production process. We also generalized the approach by proposing a methodology that allows to define, and potentially improve, the production, by giving recommendations.**

Keywords–Smart Manufacturing; Process Mining; OLAP.

I. Introduction

In the globalized market, the continuous changes in the business context, the increasing customer demands and shorter product life-cycle determine a highly competitive environment that forces manufacturing companies to a continuous alignment of the production and the internal organization. Research in the area of smart manufacturing tries to give an answer to such emerging needs. Indeed, smart manufacturing can be defined as "the dramatically intensified and pervasive application of networked information-based technologies throughout the manufacturing and supply chain enterprise" [1]. In particular, due to the complexity of the manufacturing production processes, a deep understanding of the as-is processes is essential to be able to quickly adapt such processes to new scenarios. This also enables a continuous improvement of the production, preventing bottlenecks, avoiding unexpected behaviors, and minimizing workarounds enforced by the workers. A deep understanding of the as-is scenario was the real need of the manufacturing company, producing professional coffee machines, that motivated our study. After several meetings, we agreed with the management board that an in-depth investigation of the production process is mandatory to continuously improve the way to work. This means learning from the past to better perform in the future.

To this aim, we analyzed the as-is production process using process mining techniques [2]: we applied five algorithms that we evaluated according to quality criteria [3] and complexity metrics [4]. The Inductive Miner (IM) alghoritm [5] proved to be the most suitable for the case under study [6]. Together with the production manager, we assessed the discovered process models detecting several unexpected behaviors. This finding prompted the need to understand the issues of such behaviors.

In this work, we use data analysis to investigate the factors influencing the process context [7][8] of the discovered process models. To be more precise, we focus on the correlations between the unexpected process behaviors and the context information. We selected meaningful data from the information system and generated a Data Mart (DM). Then, by the use of On-Line Analytical Processing (OLAP) tools, we outlined several correlations that we reported to the production manager for a better evaluation. As a result, we detected additional issues to consider for enhancing the production process and we provided some recommendations to prevent exceptional behaviors.

Based on the results achieved in the case study, and inspired by the process of Knowledge Discovery in Databases (KDD) [9], we also propose a novel methodology, named Process Deviations Causes Discovery (PDCD). PDCD relies on two main pillars: (i) *Process mining* for discovering as-is process models; and (ii) *Data Warehousing* and *OLAP* for analyzing correlations between the behaviors observed in the mined process and external events. The main aim of PDCD methodology is to achieve a greater awareness of unexpected behaviors detected in discovered process models. Particularly, after mining the as-is processes, the methodology allows to investigate the external factors, namely the context, affecting unexpected behaviors and to provide recommendations for improvements.

The remainder of this paper is organized as follows. Section 2 shows a motivating case study, Section 3 reports the data analysis activity, while Section 4 details the PDCD methodology. Section 5 presents some results coming from the implementation of the methodology. Section 6 provides related works and, finally, Section 7 concludes the paper.

II. Understanding Manufacturing Process

Here, we use a case study on a manufacturing company. The company produces professional coffee machines, which are exported all over the world. The manufacturing consists of assembling components provided in most of the cases by external suppliers. The production process is spread over six production lines numbered from 1 to 6. Each production line is organized into stations, each with a specific task and identified by letters from A to F. According to the different types of coffee machines, the organization of the stations in the production lines may vary: lines 5 and 6 have only five stations because B and C are merged and their activities are performed together.

- Station A assembles the frame of a new coffee machine and activates the Radio-Frequency IDentification (RFID) tag associated to it.
- Station B handles the hydraulic system. In the production lines 1 to 4, this station only assembles a portion of the hydraulic system, while in lines 5 and 6 the entire hydraulic system is assembled.
- Station C finalizes the assembly of the hydraulic system (this station is not relevant for lines 5 and 6).
- Station D deals with the electrical circuit.
- Station E performs the testing on several coffee machines simultaneously.
- Station F completes the coffee machine production including the packaging.

The company is assisted by a customized Information Technology (IT) system for managing the production process and all the related activities, such as production planning, reorder point, warehouse management, workers' support in all production phases, etc. The IT system, named ASCCO, is implemented as Process-Aware Information System (PAIS) [10]. The system also deals with tracking all the information related to the production line (assembly steps and times, faults, fixes, etc.) that are recorded in event logs. We extracted more than 450,000 events related to six years of production of 32 different coffee machines models. We then executed process mining on such event logs using five different algorithms, and we evaluated the results according to specific metrics concluding that the IM algorithm is especially suited for the case under study. These activities have been extensively discussed in our previous work [6].

The process models discovered with process mining showed some behaviors that deviate from the standard production process: the production manager and his staff were partly able to interpret such models and to react accordingly by reorganizing some phases of the production process. Despite that, the production manager required further investigation to explore special "events" that could affect the non-standard behaviors detected. We focused on production plans, workers, fixes and faults detected, customizations and the execution times of stations activities.

III. DATA MART IMPLEMENTATION

In order to make an efficient and comprehensive analysis, we rely on a Data Warehouse (DW) [11]. We started defining the conceptual model according to the Dimensional Fact Model (DFM) notation [12], as shown in Fig. 1, then we proceeded modeling the corresponding star schema.

The investigated fact refers to any single activity (summarized with the letter of the corresponding station) for assembling a coffee machine. The only measure is the execution time, that denotes the time required to perform each single activity, because most of the analysis relies on the *COUNT* operator for counting the number of items, as performed activities or produced coffee machines, according to the considered dimensions and their combination or aggregation. The dimensions we adopted are the date of assembly, the coffee machine model, the production plan, the engaged operator, faults and fixes accomplished, and the sequence of events (trace) generated by the comprehensive assembly. The dimensions were organized into appropriate hierarchies for enabling different levels of data granularity.

We implemented the Data Mart as a single cube, then we started the analysis through OLAP tools. We executed the interactive analysis involving stakeholders to take advantage of their domain knowledge and insights. The analytical tools allowed us to infer interesting considerations.

A. *Findings from OLAP Analysis*

We performed OLAP analysis with SpagoBI [13], a Web-based open source suite for business intelligence. The user interface allows a lot of processing, but in some cases we needed to modify the MultiDimensional eXpressions (MDX) query, auto-generated by the tool, in order to insert commands not available in the interface. An outcome of OLAP analysis concerned non-standard production sequences: their trend over time for any coffee machine model is proportional to the number of coffee machines of the same model produced in the same period. This means that non-standard behaviors are not related to special periods, influenced by some specific event, but they only depend on the production progress.

1) Low Execution Time and Unusual Production Line. Analysis results highlighted two unconventional situations at a glance: a large number of stations with low execution times (even less than a minute), as shown in Table 1, and a considerable number of specific coffee machines models produced on lines 5 or 6, rather than on the lines 1 to 4 as specified in the PAIS. An example is presented in Fig. 2.

We evaluated such findings assisted by the production manager and we realized that specific models, during periods of overproduction, are also assembled in lines 5 and 6, by changing the assembly process, rather than in the lines 1 to 4 for which they are designed. Moreover, the low execution times depended on an uncommon use of the traceability system. This issue was not known so far, because the production manager prepares a report on production progress on a daily basis taking into account only the number and type of produced coffee machines regardless of assembly time.

2) Weaknesses of the Information System. The above findings revealed some weaknesses of the information system: the RFID manual reading often leads it to record non effective execution times, while the rigidity of the

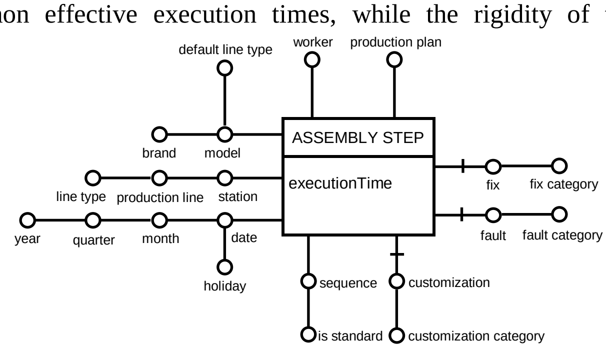

Figure 1. The ASSEMBLY STEP fact scheme.

production tracking system logs the C station activities, using fictitious times, also in the line 5 and 6, for all models designed to be assembled in lines 1 to 4. The management board decided to immediately implement improvements: (1) automating the RFID reading, in order to have absolute start and end time period of activities, (2) introducing more flexibility in the production tracking system and in particular for recording only activities really performed.

These improvements contribute to have a better event log files to be used for future process mining. In addition, tests on the new system for RFID automatic reading proved the effectiveness of such upgrade by logging inconsistent execution times for less than 0.1% of cases.

3) Customizations, Production Plans and Failures Effects. The data analysis on non-standard traces disclosed interesting connections. Many models showed an increase of customizations between 15% and 25% if compared to the global production of the same model. Similarly, the number of performed fixes and reported faults was well above the average values calculated for all the produced machines, in some cases even twice as much for the most common models as shown in Table 2. Furthermore, more than half of the traces were included in a few production plans. The above values do not seem a mere coincidence for coffee machines showing non-standard behaviors in the production process.

4) Knowledge Workers Activity. The assesment of non-standard traces revealed that a few workers seem to perform most of the activities, while in standard traces the workers are homogeneously distributed in stations. Such unexpected behavior, suggested the management board, is due to a few employees who are knowledge workers with a lot of experience, but who do not follow properly the procedures.

5) Information System Exceptions. The investigation also shows that a small part of non-standard traces comes from exceptions generated by the information system. Such traces should be marked to avoid noise in future analysis.

More generally, the observed results were thoroughly assessed by the production manager, who concluded that a good portion of non-standard behaviors was caused by operating procedures not compliant with company guidelines. These attitudes negatively affected the PAIS in recording sequences and timing of activities. The observed facts led the management board of the company to revise several aspects of the production process, and to request an upgrade of ASCCO to reflect such changes in addition to the two updates mentioned above. The planned improvements are presented next.

- New and enhanced operating procedures for the production process that will be properly fulfilled by the workers since exceptional behaviors will no longer be admitted.
- An enhanced alert system, integrated in ASCCO, for (i) reporting in real time exceptional behaviors in order to quickly react, and (ii) warning workers about previous faults encountered in the coffee machine model that they are assembling, in order to prevent the same issues.
- A new approach for performing critical customizations, consisting in specific procedures for assisting workers and different timing than the regular assembling.

IV. PDCD Methodology

The case study we run confirms how the use of OLAP analysis contributed to a better understanding of the discovered process models. This also contributed to achieve a better awareness and understanding that may be used to reorganize and improve the processes under study. It also helped to generalize the procedure we follow in a wider applicable methodology [14].

Here, inspired by the KDD process [9], we outline a methodology, named PDCD, which, starting from the selection of event logs, leads to improved process models in two steps. (i) *Process mining* for discovering as-is process models. (ii) *DW* and *OLAP* to analyze the correlations between the observed behaviors and external events.

Fig. 3 shows the basic flow of the proposed methodology. It is characterized by a high degree of interaction with the user, and it may require multiple iterations and present loops between some successive steps.

The availability of data is the beginning and the pillar of the approach. Such data may come from one or more information systems or may be gathered from many different sources such as spreadsheets, flat files, emails, etc. and then organized in a uniform and consistent manner.

TABLE I. SHORT EXECUTION TIMES

Station	Model					
	7	*8*	*12*	*15*	*18*	*19*
A	3.22%	9.08%	9.09%	2.79%	15.07%	16.98%
B	16.99%	12.63%	15.78%	8.96%	89.20%	86.67%
C	19.45%	24.80%	33.27%	38.06%	-	-
D	7.19%	13.95%	26.39%	4.47%	18.92%	21.80%
E	90.42%	86.70%	76.15%	57.83%	81.22%	85.77%

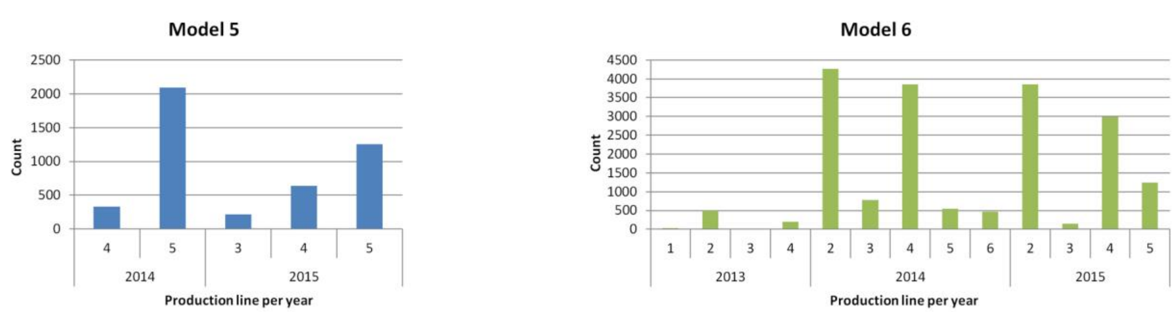

Figure 2. Two samples of coffee machine types assembled also in lines 5 and 6 even though they should be assembled on the lines from 1 to 4.

TABLE II. FIXES AND FAULTS DETECTED IN COFFEE MACHINES

Model	fixes (non-std.)	faults (non-std.)	overall fixes	overall faults
7	12.50%	10.79%	10.63%	8.93%
8	23.08%	19.23%	18.41%	8.94%
12	21.87%	21.88%	14.76%	15.24%
15	53.85%	38.46%	18.56%	16.40%
18	45.46%	50.00%	15.47%	13.87%
19	11.27%	11.26%	10.01%	8.96%

The first step, **Extraction**, consists of extracting suitable events, from the available data, as input for process mining. Events extraction means, firstly, to determine the appropriate information for the process, in order to produce an event log choosing only those data closely related to the scope of the analysis. This is necessary because, according to the adopted standpoint, it is possible to extract different event logs from the same data set. Event logs are usually stored in one of the typical formats: eXtensible Event Stream (XES) or Mining eXtensible Modeling Language (MXML).

The second step, **Process mining**, is applied to discover process models. It includes the choice of the process mining algorithm(s), the initial settings, such as parameter values, conditions or termination criteria, and the option to convert the resulting model into a different notation. The algorithms are generally chosen based on their characteristics and experiences performed in the same or similar domain. Sometimes, it is required to proceed in an empirical manner by applying several algorithms, and then determining which algorithm is the most suitable to the case under study using quality measures [3] and complexity metrics [4]. The most feasible process models will be used in the next steps.

The third step, **Evaluation**, is about evaluating the discovered process models: they are assessed and analyzed involving interested parties to understand the actual behavior of the system under study, and eventually comparing it with the desired behavior to focus on exceptions, and then making assumptions on the overall observed behavior.

The fourth step, **Data selection**, is about creating a target data set: the understanding of the actual process behavior and the assumptions made in the previous step, suggest the data subset to be selected for further investigation, among all the data initially available, in order to find probable connections to external causes related to the observed process.

The fifth step, **Preprocessing**, is data cleansing and transforming. It includes all the operations required to improve the quality of data selected in the fourth step, such as converting types and formats, removing duplicates, managing conflicts and inconsistencies, concatenating or separating relevant information, and defining methods for handling missing and unknown values. The outcome is a consistent, homogeneous and correct data set.

The sixth step, **Data modeling**, is building a Data Mart. It includes the conceptual design, for determining facts, measures and dimensions with related hierarchies, the logical modeling, for expressing the multidimensional model, e.g. the star or snowflake schema, the physical implementation, namely creating data structure according to the multidimensional model, and, at the end, the data feeding.

The seventh step, **OLAP analysis**, is finding correlations by using OLAP tools, with which to explore and analyze multidimensional data for outlining relationships between discovered process behavior and external factors impacting on the process, e.g. people involved, seasonal patterns, workload and resource availability, process misapplication. For this purpose, a domain expert is asked to actively interacts with such tools drilling-down, rolling-up, slicing and dicing, and pivoting, for generating several meaningful outcomes in form of (hierarchical) tabular data or charts for more friendly investigation and comparison.

The eighth step, **Interpretation**, is inferring the analysis results: specialists try to give a basis to the assumptions made, discarding those improbable, confirming the most probable, or requiring further investigation. This could lead to additional iterations returning to any of the previous steps.

The ninth and final step, **Discovered knowledge enactment**, is managing the new awareness on processes: implementing discovered process models in information system, if process-aware, or using such models to replace, or partially modify, the current ones, or simply using them as new reference models, for reorganizing the real processes to reflect new models, preventing the recurrence of specific deviations and exceptions, as well as providing guidelines and recommendations to process improvement.

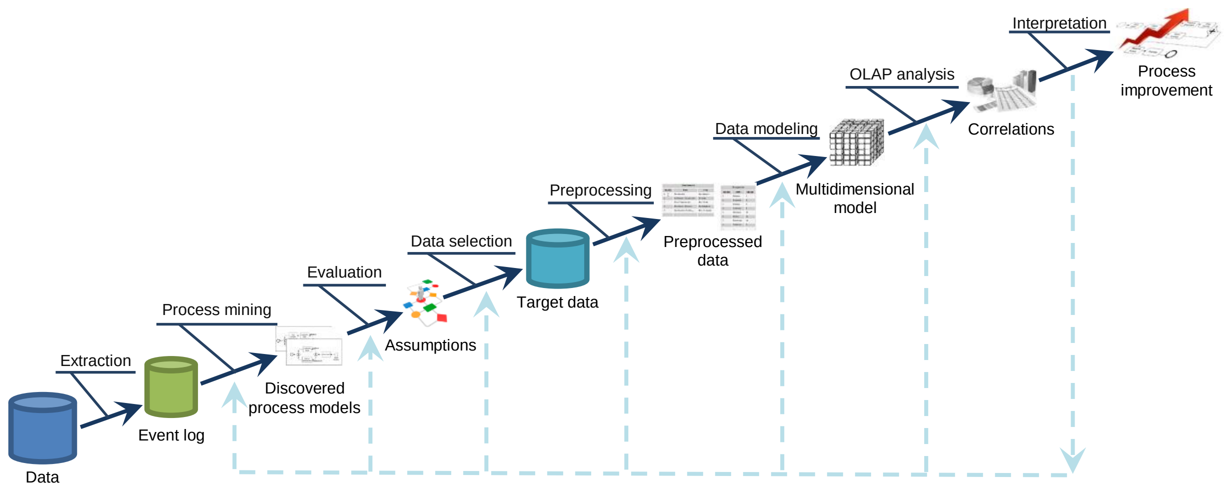

Figure 3. PDCD Methodology: Steps Overview.

V. Results and Discussion

The proposed methodology proves to be effective in the considered case study, leading to good results in getting a full knowledge of actual production processes and related context. The process mining activity alone allows to discover as-is processes, by providing models that ensure a close correspondence to the actual behavior of the processes because they are generated based on real event data. Therefore, the discovered models represent processes as they are actually performed during the examined period, but they did not provide any details about specific observed patterns.

One issue that often arises from stakeholders during the evaluation of a process model is “why this sequence of activities?”. The answer can rarely be inferred from the model itself. In our case study, it was not possible to answer such question even if the model presents a small number of activities. This became more and more complex in case involving a higher number of activities. In practice, it is only possible to make assumptions, which, however, must be validated in order to be “converted” into an answer.

The main goal of our methodology is precisely to try to give such answers. To this aim, the assumptions provided by the process experts are relevant for choosing the information to be investigated. This avoided to persist on irrelevant data or data not related with process under examination. The decision to address the DW world and to use OLAP tools revealed all the benefits in performing data analysis in a flexible and structured manner, observing information from different viewpoints and at different levels of detail. Furthermore, the data analysis phase cannot ignore the involvement of domain experts for attaining substantial results that will be further assessed by the same experts. The resulting suggestions could be used for:

- Accomplishing a new full cycle after generating more appropriate event logs;
- Repeating the analysis integrating the already used information, or using a different set of information;
- Establishing criteria to simulate changed processes, for checking the runs of processes before upgrading;
- Defining guidelines or take measures to improve processes and limit the exceptions.

The use of a systematic approach to provide criteria by which to argue the observed behavior in process models discovered by process mining, represents an added value to acquire a deeper understanding of the entire process. In addition, if the discovered model may be compared to a standard designed model, such criteria should support further assessment of detected deviations. Further assessment could determine which deviations to keep on the new process model and which ones to consider just as exceptions or, even, which ones to avoid because counterproductive.

VI. Related Work

To the best of our knowledge, there are no previous works that merge process mining and data analysis for discovering process models and then investigating external factors affecting such processes. The external factors are usually identified as the context of the process. In Business Process Management (BPM) the concept of context has several facets: in [8][15], it is described as the environment in which a business process may be used, while in [16], it is “The minimum set of variables containing all relevant information that impact the design and execution of a BP”, and in [17], the context is “any information reflecting changing circumstances during the modeling and the execution of a BP”. The work in [7] outlines the importance of considering the process context for improving BPs, learning from past experiences. In the above works, the concept of context is introduced to explain the benefits of the context-awareness in the BPM scope, and, in particular, in BPM design. However, also process mining techniques, as highlighted for the first time in [18], may greatly benefit in considering the process context, that is categorized into four classes: instance context, process context, social context and external context. In our work, we mainly considered the instance context, namely the factors that influence the singular process instances such as product customizations, assembling times and fixes.

The proposed methodology, inspired by KDD process [9], merges the BPM and DW. In particular, OLAP techniques are used to better understand the process models discovered by process mining, and the external causes, i.e., the context. A similar idea is in [19] where an approach for analyzing and preventing exceptions in BP is described. However, such approach is based only on generic Data Warehousing and data mining techniques. No process mining is applied and exceptional behaviors are defined by conditions over process execution data, i.e., subjectively, instead of comparing discovered models to standard ones.

In literature, there are further works that combine BPM and DW, with an extensive discussion being presented in [20]. The work in [21] describes a multidimensional approach for process modeling which enables the mapping of different aspects of a BP into a data cube and the support for a wide range of analysis. Such approach deals with only numerical data, a drawback overcome in [22] where the concept of OLAP data cube is merged with BP formalizing the notion of Process Cube. In Process Cube, events and process models are organized using different dimensions, and slice, dice, roll-up and drill-down operations are reformulated to be consistent with the new data structure. Improved versions of the Process Cube framework are the Process Mining Cube tool [23] and the PMCube explorer [24], that allow a more detailed multi-dimensional representation of a business process for a proper analysis.

Other works that combine BPM and DW, without merging them, show data mining as the main pillar. In [25] data mining is integrated with BPs for enabling an agile and exhaustive analysis of processes. Moreover, a methodology for successfully reusing data mining solutions during integration is introduced. Inspired by such approach, a formal framework for BP redesign is proposed in [26]: operational data collected during process runs are mined to explicitly represent the dynamics of the BP. This allows to re-design the process more efficiently. In [27] data mining is used to support decisions on resource allocation. Process context data, extracted from past process executions, are mined to acquire new knowledge for guiding optimal resource allocations in new process instances.

In summary, no work has so far combined process mining and DW as we did in this paper. They are strictly related for achieving a thorough knowledge of discovered processes by exploiting the context information, but they are not merged in order to quickly implement improvements in each single stage, such as a more efficient process mining algorithm or a better design of data cube.

VII. CONCLUSION AND FUTURE WORK

In a competitive and globalized business context, manufacturing companies need to adapt rapidly to new conditions in order to advance. Furthermore, production processes in the manufacturing field are quite complex, so it is needed to have a comprehensive understanding of such processes in order to adapt them to the new settings. In a case study, we investigated, in collaboration with domain experts, the process models discovered with process mining algorithms. We selected a large set of data from the company information system, we run process mining and we built a Data Mart. Using OLAP we then performed a thorough analysis and submitted the results to the production manager and his staff. Their interpretation contributed to a deeper understanding of the observed behavior and led to feedback on how to improve the production process.

We also generalized the approach by proposing a methodology for achieving a better awareness of the process models discovered with process mining.

REFERENCES

[1] J. Davis, T. Edgar, J. Porter, J. Bernaden, and M. Sarli, "Smart manufacturing, manufacturing intelligence and demand-dynamic performance," Comput. Chem. Eng., vol. 47, pp. 145–156, 2012.

[2] IEEE Task Force on Process Mining, "Process Mining manifesto," in Business Process Management Workshops, vol. 99 LNBIP, no. PART 1, pp. 169–194, 2012.

[3] W. M. P. van der Aalst, Process Mining: Discovery, Conformance and Enhancement of Business Processes. Springer Berlin Heidelberg, 2011.

[4] J. Mendling, Metrics for Process Models: Empirical Foundations of Verification, Error Prediction, and Guidelines for Correctness, vol. 6. Springer Publishing Company, 2008.

[5] S. J. J. Leemans, D. Fahland, and W. M. P. van der Aalst, "Discovering block-structured process models from event logs containing infrequent behaviour," in LNBIP, vol. 171, Springer International Publishing, pp. 66–78, 2014.

[6] A. Bettacchi, A. Polzonetti, and B. Re, "Understanding Production Chain Business Process Using Process Mining: A Case Study in the Manufacturing Scenario," in Advanced Information Systems Engineering Workshops: CAiSE 2016 International Workshops Proceedings, pp. 193–203, 2016.

[7] K. Ploesser, M. Peleg, P. Soffer, M. Rosemann, and J. Recker, "Learning from Context to Improve Business Processes," BPTrends, vol. 6, no. 1, pp. 1–7, 2009.

[8] United Nations Centre for Trade Facilitation and Electronic Business, "Core components technical specification - part 8 of the ebXML framework." p. 113, 2003.

[9] U. M. Fayyad, G. Piatetsky-Shapiro, and P. Smyth, "From Data Mining to Knowledge Discovery: An Overview," in Advances in Knowledge Discovery and Data Mining, USA: American Assoc. for Artificial Intelligence, pp. 1–34, 1996.

[10] M. Reichert and B. Weber, Enabling Flexibility in Process-Aware Information Systems: Challenges, Methods, Technologies. Springer Berlin Heidelberg, 2012.

[11] R. Kimball, The data warehouse toolkit. Wiley, 1996.

[12] M. Golfarelli, D. Maio, and S. Rizzi, "The Dimensional Fact Model: A Conceptual Model for Data Warehouses," Int. J. Coop. Inf. Syst., vol. 7, no. 2–3, pp. 215–247, 1998.

[13] SpagoBI Homepage, https://www.spagobi.org/ 2016.10.31.

[14] B. Flyvbjerg, "Five Misunderstandings About Case-Study Research," Qual. Inq., vol. 12, no. 2, pp. 219–245, 2006.

[15] M. Born, J. Kirchner, and J. P. Müller, "Context-driven Business Process Modelling," Jt. Proc. 4th Int. Work. Technol. Context. Bus. Process Manag. TCoB 2009. Conjunction with ICEIS 2009., pp. 17–26, 2009.

[16] M. Rosemann, J. Recker, C. Flender, and P. Ansell, "Understanding Context-Awareness in Business Process Design," Proc. 17th Australas. Conf. Inf. Syst., 2006.

[17] O. Saidani and S. Nurcan, "Context-awareness for adequate business process modelling," in Third International Conference on Research Challenges in Information Science, RCIS 2009, Morocco, pp. 177–186, 2009.

[18] W. M. P. van der Aalst and S. Dustdar, "Process Mining Put Into Context," IEEE Internet Comp., vol. 16, no. 1, pp. 82–86, 2012.

[19] D. Grigori, F. Casati, U. Dayal, and M.-C. Shan, "Improving Business Process Quality through Exception Understanding, Prediction, and Prevention," in Proceedings of the 27th Intl. Conf. on Very Large Data Bases, Italy, pp. 159–168, 2001.

[20] M. K. Shahzad, "Improving Business Processes using Process-oriented Data Warehouse," PhD dissertation, DSV, KTH, Stockolm, Sweden, 2012.

[21] S. Mansmann, T. Neumuth, and M. H. Scholl, "Multidimensional Data Modeling for Business Process Analysis," in Proceedings of the 26th international conference on Conceptual modeling (ER'07), pp. 23–38, 2007.

[22] W. M. P. van der Aalst, "Process Cubes: Slicing, Dicing, Rolling Up and Drilling Down Event Data for Process Mining," Asia Pacific Bus. Process Manag. First Asia Pacific Conf. AP-BPM 2013, Sel. Pap., vol. 159, pp. 1–22, 2013.

[23] A. Bolt and W. M. P. van der Aalst, "Multidimensional Process Mining Using Process Cubes," in BPMDS 2015, EMMSAD 2015, Held at CAiSE 2016 International Workshops Proceedings, pp. 102–116, 2015.

[24] T. Vogelgesang and H. J. Appelrath, "Multidimensional process mining with PMCube explorer," in Proceedings of the BPM Demo Session 2015, vol. 1418, pp. 90–94, 2015.

[25] D. Wegener and S. Rüping, "On reusing data mining in business processes - A pattern-based approach," in Business Process Management Workshops: BPM 2010, Revised Selected Papers, vol. 66 LNBIP, pp. 264–276, 2011.

[26] T.-M. Truong and L.-S. Lê, "Towards a Formal Framework for Business Process Re-Design Based on Data Mining," in BPMDS 2016, EMMSAD 2016, Held at CAiSE 2016 International Workshops Proceedings, pp. 250–265, 2016.

[27] R. Sindhgatta, A. Ghose, and H. K. Dam, "Context-Aware Analysis of Past Process Executions to Aid Resource Allocation Decisions," in Advanced Information Systems Engineering: CAiSE, Ljubljana, Slovenia. Proceedings, pp. 575–589, 2016.

 ISBN: 978-1-61208-536-4

Towards a Generalised e-Learning Business Process Model

Rawad Hammad, Mohammed Odeh and Zaheer Khan
Software Engineering Research Group, Department of Computer Science and Creative Technologies,
Faculty of Environment and Technology, University of the West of England, Bristol,
BS16 1QY, United Kingdom
Email: Rawad.hammad@uwe.ac.uk, Mohammed.odeh@uwe.ac.uk, Zaheer2.khan@uwe.ac.uk

***Abstract*— Modelling learning scenarios is central for e-learning domain. This has been manifested in the proliferation of the different Educational Modelling Languages, as well as in developed e-learning models. However, the existing modelled scenarios are deficient as they lack flexibility and the agility to respond to the dynamic nature of an e-learning process that is suitable to answer learners' needs. This paper proposes a novel approach to develop a generalised business process model from a set of related business processes sharing the same goals and associated objectives. The proposed approach has been applied in the e-learning domain, which demonstrated its ability to develop a generalised e-learning business process model that is derived from the existing pedagogical models and technology-enhanced learning artefacts. Moreover, the proposed approach has been evaluated to test its effectiveness in generalising a set of business processes, which paves the ground to apply it in different contexts. The generalised e-learning business process model has been modelled using the industrial standard Business Process Modelling Notations (BPMN 2.0) so that processes can be dynamically enacted in service-oriented environments and, at the same time, adapting to answering e-learners' learning requirements.**

Keywords- e-learning processes; business process models for e-learning; e-learning; technology-enhanced learning; process-based e-learning; business process generalisation.

I. INTRODUCTION

Various educational organisations are increasingly adopting e-learning/Technology-Enhanced Learning (TEL) due to their ability to meet different e-learners' needs and work with newly innovative e-learning models, such as connectivism and self-regulated learning [1]. This application of e-learning technologies differs from one organisation to another, which necessitates having a well-specified and generalised e-learning model. In this context, learning is the act by which behavioural change, knowledge, skills and attitudes are acquired [2], which can be described as a learning *process*. A process, from a computational perspective, involves activities which are performed by certain entities (i.e., human and/or machine) working in collaborative groups to achieve specific business goals [3]. However, evolved e-learning models rarely adopt the business process concept, which negatively impacts their agility and capability to respond to e-learners' demands [4]. Thus, this paper is an attempt to understand widely published models of e-learning business processes, classify these processes, and then generalise them to form a generic e-learning business process that is pedagogically sound and can adapt to different learning paths/processes based on e-learners' context.

The rest of this paper is organised as follows: Section II discusses related work; Section III describes the proposed approach to deriving a generalised business process model from a set of related business processes having the same goal; Section IV applies and demonstrates the generalisation approach/process in the e-learning domain; Section V discusses the proposed approach and reflects on the results of adopting a case study-based evaluation approach to determine the effectiveness of the proposed generalisation method; and Section VI concludes the paper with future research directions.

II. RELATED WORK

There exist various e-learning/TEL models, such as the Learning Management Systems and Learning Objects where the emphasis is on the role of technology in supporting learning and teaching. Such models are practice models; henceforth, they are considered as *e-learning artefacts*, mainly to distinguish them from pedagogical models underpinning e-learning. This section reviews both types (i.e., e-learning artefacts and pedagogical models) in order to form a better understanding of e-learning processes and potentials to improve these processes. Therefore, this section is divided into the following two sub-sections: (i) e-learning artefacts and (ii) e-learning pedagogy.

A. e-Learning/TEL Artefacts

The continuously changing learning contexts (e.g., learners' demands, institutional settings, subjects taught, etc.) have led to the proliferations of diverse e-learning artefacts. These artefacts stretch from simple ones, such as Learning Object (LO) through complex ones, such as Instructional Management Systems Learning Design (IMS LD). This section reviews three e-learning artefacts and reflects on their process-related concerns. *First*, LO is the essential element that exists in all other artefacts. LO usually refers to: (i) instructional contents developed to address certain learning objectives, (ii) assessment activity, and (iii) metadata to describe the LO and make it discoverable [5]. In spite of LO strengths, such as reusability and interoperability, it is content-oriented and lacks the well-structured representation of learning concerns, which limits its pedagogical value [6].

Second, the proliferation of different Educational Modelling Languages (EML), such as the Open University of Netherland EML (OU EML) [7] and the UNED University EML (PALO) [8] represent an advanced step towards decoupling the learning process from its contents instead of

 ISBN: 978-1-61208-536-4

having content-oriented artefacts (e.g., LO), where contents and processes are highly coupled. According to [9], OU EML has been acknowledged as the most powerful and expressive EML; and therefore, it has been standardised by the IMS Global Learning Consortium under the title "IMS LD". IMS LD embodies a containment framework of elements that can formally describe the design of any teaching-learning process/scenario [10]. It is the only interoperability specification that allows designers to describe Units of Learning (UoL), where a UoL is the smallest unit providing learning events for learners, satisfying one or more interrelated learning objectives [11]. However, IMS LD has shortcomings that include: (i) lack of flexibility (e.g., tiny changes to contents are not possible unless essential modifications to the activity structures, act, role-part, method, properties and conditions are done), (ii) interoperability-oriented concerns (e.g., cannot save or retrieve information to/from external sources) [12], (iii) dynamic grouping for users is not possible, (iv) user behaviour is not recorded, (v) adaptation is limited (i.e., no adaptation based on previous user behaviour), and (vi) complexity, since it works as an integrative layer with other specifications [13]. Further limitations are discussed in [14].

Third, the above-mentioned limitations have led to the development of more process-oriented e-learning artefacts, such as Workflow-based e-Learning Platform (WeLP) [15]. WeLP aims at facilitating and enhancing the performance of e-learning systems through separating processes (i.e., activities, roles, conditions, etc.) from other e-learning ecosystem components, such as e-learning contents and other technical components. To do so, e-learning procedures have been divided into the following four aspects: (i) *teaching* that targets lecturers, (ii) *learning* that targets students, (iii) *administration* that targets administration and personnel, and (iv) *infrastructure* that targets infrastructure, technical experts and technicians. These four aspects represent four sub-processes that will be used to plan and design the process of various e-learning activities. Each process represents a list of activities that ensure its successful implementation. However, WeLP remains at the very high level of abstraction, leans toward design, and lacks a real evaluation that can prove its impact in terms of developing better e-learning platforms. It intuitively analyses the relationships between the proposed sub-processes and activities but lacks detailed specification of activities. For instance, material delivery is a process by itself and cannot be squeezed into one simple activity.

To conclude, process-based approaches are either: (i) not adopted in e-learning artefacts systems (e.g., LO), (ii) semi adopted (e.g., IMS LD) but in a very complicated approach where the e-learning process is cemented into the system, (iii) adopted in a superficial way where underpinning pedagogy is ignored, or (iv) remains at the concept/abstract level (e.g., WeLP).

B. *e-Learning Pedagogy*

As stated above, all e-learning artefacts are underpinned by certain pedagogical models or theories. Therefore, significant analysis for the available pedagogical strands is necessary to inform the e-learning processes derivation. Developing a proper understanding of e-learning pedagogy enables us to: (i) formally specify available e-learning models, (ii) understand how these e-learning models can be used by stakeholders, (iii) generalise these process models, and (iv) better decide what contextual information is needed to customise the generalised model for each learner based on his/her needs. There exist two schools of thought regarding understanding pedagogy. The first school does not believe in theory because the learning phenomenon cannot be explained by simple theories [16]. The second school, adopted in this research, believes that learning theories are essential to understand pedagogy [17]. Being the proponent of the second school of thought, it is worth recalling that there is no agreement on one single classification for pedagogical strands. In addition, this research embraces Greeno et al's classification [18], where learning can be understood through the following three broad perspectives.

First, we have the associationist perspective, where learning is the process of connecting the elementary mental or behavioural units through a series of activities. Various learning theories/processes fall in this perspective, such as instructional design and direct instructions [19]. *Second*, there is the cognitive/constructive perspective, where learning is about achieving understanding. Learning here is interpreting and constructing meanings, while knowledge acquisition is the outcome of interaction between learner's new experiences and his/her previous structures/understanding. Learning by doing and problem-based learning fall in this perspective. *Third*, we have the situative perspective, where learning is situated in various social practices and contexts. The e-learners' relationship with their community shapes their knowledge, learning outcomes and ability to learn by participation [20]. Connectivism and community of practice learning theories fall in this perspective [21]. As explained above, each perspective encompasses various learning theories, but a more detailed discussion remains beyond the scope of this research. The next section proposes a manual method to develop a generalised business process model from a set of related business processes having the same goal.

III. The Proposed Approach to Develop a Generalised Business Process Model From a Set of Related Busesiness Processes Having The Same Goal

Process-based systems consist of various business processes. These business processes collectively aim at achieving the same business goals and objectives, but they may vary in the design of the process details (e.g., workflows, interactions, concurrent or sequential flow of activities, means of achieving the same objective and approaches to attend the tasks, etc.) Often, domain specific business processes possess common characteristics which can be generalised to promote reusability, consistency and interoperability among different business organisations. For instance, the direct instruction learning process refers to learning by following instructor-designed learning processes, while the self-regulated learning process refers to self-planning, self-monitoring and self-assessment for learning processes. So, the goal of both processes is the same, but they use different mechanisms to achieve that goal. Therefore, an effective generalisation approach is needed. To do so, *first*, the existing e-learning literature is surveyed, which includes: e-learning artefacts, pedagogical models, various e-learning designs and principles adopted in authoring tools. *Second*, business process modelling (e.g., Business Process Modelling Notations (BPMN)) and business process architecture (e.g., Riva method) literature is reviewed. *Third*, lessons learnt from the previous two steps have been used to develop the proposed generalisation approach. In this respect, Riva is a methodological approach [22] to derive business process architectures for a certain organisation from its essential business entities. Riva and BPMN work on two different levels, the former targets the process architecture (i.e., more abstract level), while the latter targets the activities implemented to achieve process goals. Investigating related literature from both domains (i.e., business process modelling/architecture and e-learning) helps devise the proposed approach from different perspectives. For instance, Riva steps in classifying Essential Business Entities to identify Units of Work and considers different analytical perspectives/abstraction levels have been adapted to deal with the domain-specific concerns from a process perspective. The proposed generic method to generalise business processes is based on the following steps:

1- Analyse all available business processes, their goals, activities, pedagogic models/theories influencing them and determine the boundary of these processes. This allows us to obtain insights about different e-learning processes, their scopes and whether they can be formally modelled using BPMN visual notations and corresponding machine readable formats (e.g., XML Metadata Interchange (XMI) and XML Schema Definition (XSD)).
2- If necessary, classify the early-identified business processes based on domain-specific concerns to bring further coherence to the proposed processes/activities (e.g., as depicted in Fig. 2: e-Learning Process (LP1) to LP 9 have been classified in three different categories). This classification can help in capturing the semantics of various e-learning processes.
3- Identify all process elements which include: (i) flow objects (events, activities and gateways), (ii) data (data objects, inputs, outputs and data stores), (iii) connecting objects (sequence flows, message flows, associations and data associations), (iv) swimlanes (pools and lanes) and (v) artefacts (group and text annotation). Some of these elements (e.g., text annotations) help to capture semantics of specific activities, which can be useful later on for business process enactment and execution in a Service-Oriented Architecture (SOA)-enabled environment.
4- Detect the common process elements and the special/unique ones from the early-identified process elements (i.e., the outcome of step 3). For instance, *user login* and *set profile* are common activities in various processes, while *plan your e-learning activity* is not common.
5- Generalise the special/unique process elements (e.g., the following two activities: (i) *"study a particular learning lesson"* and (ii) *"perform the following instructions"* can be generalised in the following activity: *"participate in the specified learning activity"*). Careful considerations for the terms used is needed as they reflect different underpinning learning approaches (e.g., *"perform"* usually entails participatory learning while *"study"* does not).
6- Define and specify the rules and the conditions that are essential to customise the generic e-learning process for a certain e-learner (i.e., generate a specialised business process from the generic one). For instance, define the following rule: e-learning process combines Self-Regulated Learning (SRL) elements for those e-learners who have metacognitive skills. Such rules allow selecting the suitable process elements from the generalised business process elements. Specifying this rule requires adopting certain specification/standard that is suitable for this research context (i.e., capturing the semantics of e-learning processes). For this research, the Semantic Web Rule Language (SWRL) has been selected due to its expressiveness and automated reasoning capabilities. Also, it is compatible with Web Ontology Language (OWL), which is used for contextualising e-learning processes. The above rule is translated to more machine readable format (e.g., "if then else rule"). For example, *'If a particular e-learner has metacognitive skills then suggest SRL elements for his/her e-learning process'*. In order to perform automated reasoning at process execution time, the above rule is translated to SWRL specifications, as depicted in Fig.1. A SWRL rule is composed of: (i) *antecedent* and (ii) *consequent*, that are separated by "->". Both antecedent and consequence are composed of atoms connected with conjunctions, where conjunction is represented as ",". Once the antecedent atoms are true the SWRL rule fires and execute the atoms on the left hand side. SWRL rules are executed using a software reasoner.

```
eLearner(?x), eLearnerBehaviouralModel(?lbm),
Skills(?s), LearningProcess(?lp),
hasAneLearnerBehaviouralModel(?x, ?lbm),
haseLearnerModelSkill(?lbm, ?s),
hasFollowLearnerLearningProcess(?x,?lp),skillType
(?s, ?str), matchesLax(?str, "Metacognitive") ->
recommendedProcessElement(?lp,"SRL")
```

Figure 1: SWRL Rule Syntax

7- Make the information required to execute the early-specified rules available (i.e., types of e-learner skills should be modelled in the e-learner behavioural model in order to make the above-mentioned rule executable). This is expressed in Fig. 1 by the atom *matchLax(?str, "Metacognitive")*.
8- Identify, if any, potential conflicts between process elements (e.g., SRL e-learning processes contradict with Direct Instruction, especially in selecting learning goals.

 ISBN: 978-1-61208-536-4

This has essential consequences on the process's roles and their actions).

9- Resolve the discovered contradictions by introducing intermediate process elements, further rules or making assumptions necessary to accurately specify the business process. For instance, *"Decide Learning Approach"* activity has been added to the generic e-learning process model, where this activity is backed by certain SWRL rules.

10- If the early-identified business processes have been classified, then make one level of generalisation for each category. For instance, in Fig. 2: LP1, LP2 and LP3 have been generalised and led to Upper-Level eLearning Process (ULP1) and similarly LP4 to LP7 have been generalised and led to ULP2 and so on.

11- Perform another level of generalisation for the outcome of the previous step (i.e., the early-generalised processes) using steps 4 to 10. For instance, ULP1, ULP2 and ULP 3 have been generalised and led to the generalised e-Learning Business Process.

12- Cross-Verify whether the generalised e-learning process model can adapt all different detailed process models and their activities by going through the generalised process model and confirming its ability to accommodate elements from the early-identified detailed e-learning processes.

In the next section, the above-proposed approach will be applied in e-learning domain to check its effectiveness in generalising an e-learning process model that could meet various e-learners' requirements.

IV. Applying The Proposed Approach to Develop a Generalised e-Learning Business Process Model

This section covers the following three concerns: (i) applying the early-proposed approach in e-learning domain to develop a generalised e-learning business process, (ii) the nine detailed e-learning business processes and (iii) the generalised e-learning business process model.

A. *The Proposed Approach to Develop a Generalised e-Learning Busines Processes*

This sub-section demonstrates how the early-proposed approach is applied in the e-learning domain. As previously-mentioned, e-learning processes have not been properly identified which necessitates carrying out a thorough analysis for pedagogical theories and models underpinning e-learning artefacts as indicated in the *first* step. This has led to identify nine e-learning processes, as described in the next sub-section. *Second*, the nine e-learning processes have been classified, as depicted in Fig. 2, based on domain-specific (i.e., pedagogical) concerns and scoped to cover learning-oriented aspects only. *Third*, all process elements have been identified. *Fourth*, common and unique elements have been identified. *Fifth*, various unique elements have been abstracted using generic terms, such as participate in assessment activities, where assessment can take different forms stretching from simple quizzes through project-based approaches.

Sixth, rules have been defined to explain which form will be chosen for a certain e-learner. *Seventh*, all constructs (e.g., feedback score, previous learning styles, etc.) required to execute the early-defined rules have been made available. *Eighth*, some contradictions (e.g., self-regulated e-learning processes versus instructor-directed ones) have been identified and resolved, as indicated in step *nine*, by introducing intermediate process elements. *Tenth*, three generalised e-learning processes have been developed. *Eleventh*, a final generalised e-learning process has been developed out of the outcome of the previous step. *Twelfth*, the final generalised e-learning process has been evaluated to ensure the inclusion of all detailed e-learning process elements, as will be explained later.

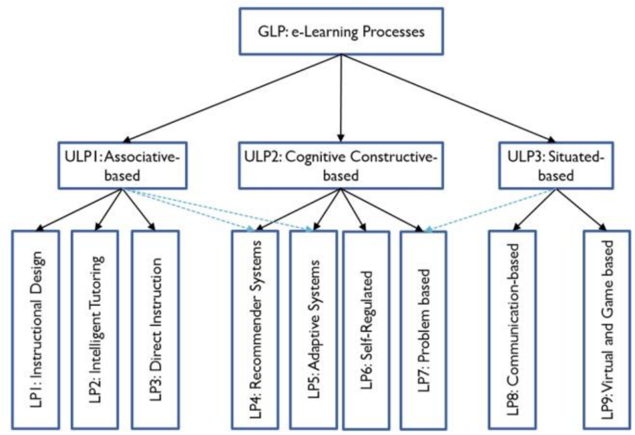

Figure 2: The Generalised and Detailed eLearning Processes

In this way, the generalised e-learning process is driven by pedagogy and informed by practice e-learning models. In the next two sub-sections, the nine e-learning processes will be briefly described under their classification, as depicted in Fig. 2. Then, the generalised e-learning process will be introduced.

B. *The Detailed e-Learning Business Processes*

This section covers nine detailed e-learning processes according to their pedagogical perspectives, as follows.

Associationist e-Learning Processes, which consist of the following three e-learning processes. *First*, there is the Instructional Design (ID) e-learning process, which is a typical behavioural/associationist e-learning process. Like any other e-learning process, ID e-learning process starts with common login activities. Successful candidates will be able to explore the learning space provided by the e-learning system to the learners to interact with contents/activities and perform all the tasks to accomplish their goals. Then, the e-learners will be able to select the topic required to study, perform the learning activity (e.g., read the learning objectives and proceed to the lesson if they wish). To check e-learners' understanding, they are supposed to participate in the assessment activity specified by the instructor, which will usually lead to useful feedback. This feedback is automated and is quite generic - not specific for each e-learner. Well-designed ID processes embody remedial contents for those who were not able to accomplish their objectives. e-Learners are allowed, in such e-learning processes, to seek support from academic staff or initiate collaborative activities with their peers.

Second, there is the Intelligent Tutoring Systems (ITS) e-Learning Process. ITS represents a wide spectrum of systems evolved in different ways that adopt various mechanisms including expectation and misconception tailoring, constraints-based modelling, model tracing, separate in class instruction, integrated class instruction, feedback provision, and misconceptions modelling. The ITS e-learning process based on misconception modelling will be modelled to represent this type of processes because modelling the expectation and misconception based on principal instruction is very common in ITSs, as shown in different studies (e.g., [23]). The main added value of ITS process is its ability to deliver a specific learning to each e-learner based on his/her model as well as the mechanism provided to provide feedback to e-learners.

Third, Direct Instruction (DI) e-learning process offers more emphasis on the practice and consequently acting up on this practice via feedback. Therefore, the e-learner behaviour is observed by instructor in order to provide the relevant feedback that is suitable for the e-learner progress towards the attainment of the learning objectives. Observation can take different forms and similarly feedback as well. Feedback is composed of: *evaluative part*, which is related to the learning outcome and indicates the performance level achieved and *the informational component*, and consists of additional information relating to the concept, task, mistakes or how to proceed [24].

Cognitive Constructive e-Learning Processes, which includes many processes. Below are some of the most used processes in current artefacts. *First*, we have Problem-Based e-learning process (PBL). PBL is not problem solving, but it ensures that learning happens in the context of problem solving or real world scenario. It is composed of the following steps [25]: (i) identify concepts of the problem that need clarification, (ii) define the problem, (iii) analyse the problem, brainstorm about solutions or causes, (iv) structure solutions or causes, (v) state learning objectives, (vi) self-study directed towards learning objectives, and (vii) report things learned and application to the problem. Usually, assessment is measured against competencies acquired to show mastery in the field.

Second, Self-Regulated e-Learning (SRL) process occurs when the e-learner takes the initiative with or without the help of others to diagnose their learning needs, formulate learning goals, identify resources for learning, select and implement learning strategies and evaluate their learning outcomes [26]. The SRL process is composed of the following activities [27]: (i) *plan*, e-learner provides input

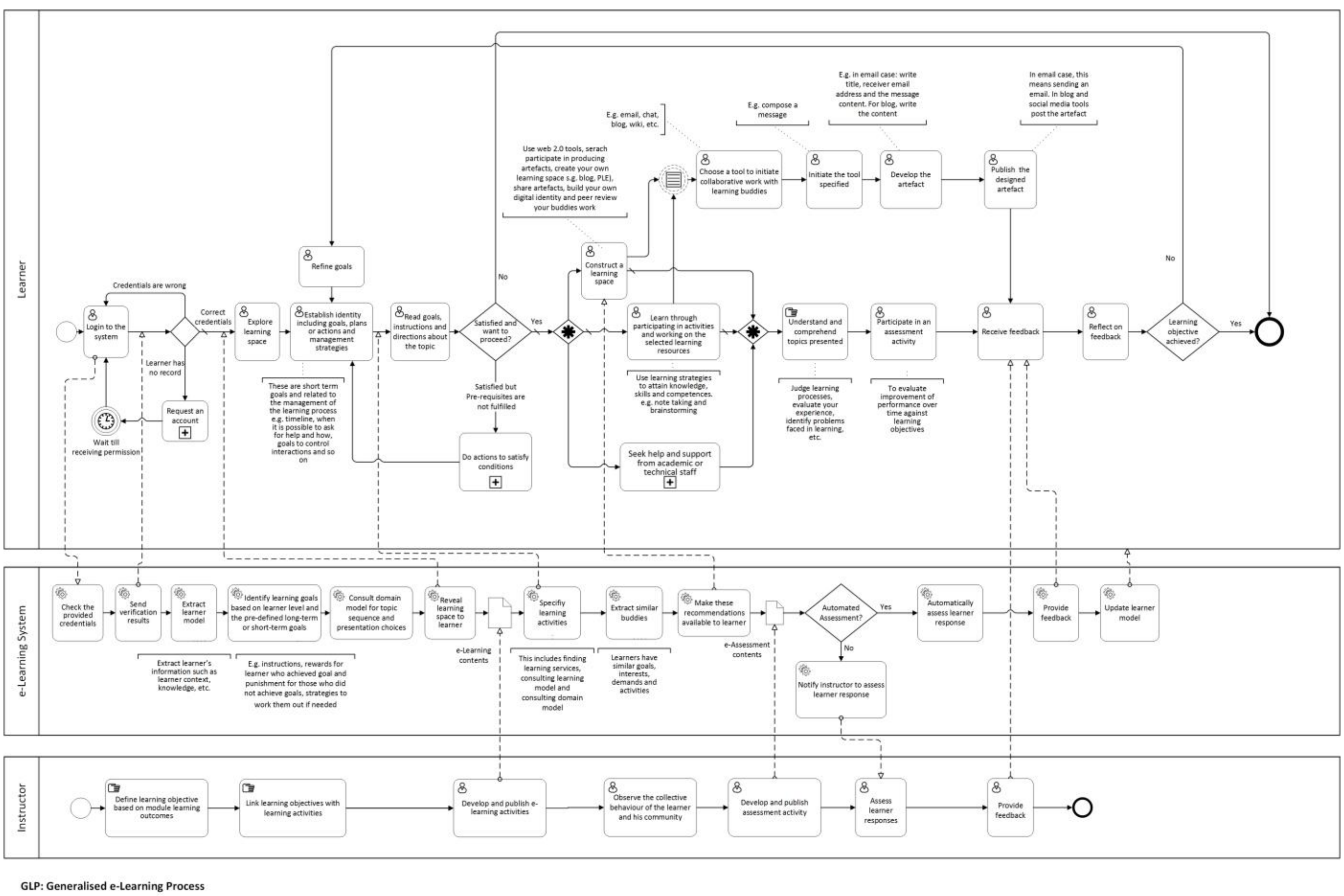

Figure 3: The Generalised e-Learning Process

regarding goals, preferences (e.g., profile-setting), (ii) *prepare*, e-learner finds and selects learning resources (e.g., explore or find contents), (iii) *learn*, e-learner works to attains knowledge, skills and competences using learning strategies and techniques (e.g., time management), and (iv) *reflect*, e-learner reflects and reacts on strategies, achievements and usefulness (e.g., self-evaluating).

Third, there is the Recommender Systems (RecSys) e-learning process. RecSys is applied in various domains, however, its application in e-learning significantly varies because of e-learning particularities (e.g., long terms educational goals) [28]. RecSys internal process focuses on two main aspects, either recommending learning resources or finding peers who share interests, goals and characteristics with the e-learner. Each type has different sequence of activities (e.g., finding peers RecSys check the e-learners' history to identify e-learners with similar learning patterns). In contrast, resources' recommendations RecSys require further check for the learning model, the domain model and the previous feedback.

Fourth, Adaptive Systems e-learning process varies from one system to another, but generally consists of extracting the e-learner model, checking which learning goal, objectives or tasks need to be accomplished, checking the domain model to capture the proper content suitable for that learner, as well as proper presentation techniques, presenting contents and finally updating learner model based on feedback.

Situated e-Learning Processes, which covers two main processes. *First*, Communication/Participation-based situated e-learning process that is dominated by the learner participation and communication with peers and instructor to learn new concepts. It shows how interactions can be done in situated learning environments. In such learning processes, the instructor is mainly facilitator rather than instructor. Connectivism learning theory is an active example on this category because it shows the roles of the non-human appliances in learning processes [21]. *Second*, we have Virtual-Enhanced e-learning (VEL) or Game-Enhanced e-Learning (GEL) processes, which represent the use of virtual world and game-enhanced e-learning systems. Such models establish an identity for each e-learner, allow the e-learner to explore the whole environment, plan for progress, work according to plan, gain some achievements as a result of understanding the concepts or the knowledge presented and proceed with the next steps [29]. Generalising the above-mentioned e-learning business processes is introduced in the next section.

C. *The Generalised e-Learning Business Process*

Fig. 3 shows the final outcome of applying the early-proposed approach to develop a generalised e-learning process that can lead to different e-learning processes based on the hybrid input captured from the e-learner's context. This context must have different behavioural information about the e-learner (e.g., his/her knowledge and learning preferences), topics, programme, peers, institutions, etc. The generalised e-learning process includes the following three roles: the e-learner, the instructor and the e-learning system. Generally, this e-learning process model consists of four key activities, as detailed below.

First, the e-learner needs to login into the system. This includes certain seamless activities (e.g., check the e-learner's credentials) to be carried out by the system. Then successful login leads to initiating the early-specified "learning space" where the e-learner sees whatever is available on the system (e.g., modules and courses). The *Learning Space* provides contents/activities (e.g., learning or assessment activities) designed by instructors. However, the learning space and other activities in the business process model are adaptive, dynamic and responsive as they differ from one e-learner to another. This is mainly because this e-learning business process model is supported by a comprehensive ontological model that captures the semantics of the e-learning process to meet the demands of the e-learner. This ontological model has been developed based on a detailed survey of e-learning models and artefacts. This ontological model can not be covered here in detail due to space limitation and hence will be covered elsewhere. Mainly, it consists of the following eight main constructs: (i) *eActor*: models roles, which interact with the software system for certain purposes, (ii) *e-Learning Facilitating Tool*: models the wide range of software tools (e.g., wiki, e-mail, etc.) used in e-learning context to facilitate and support e-learners, (iii) *Pedagogy*: models different pedagogical strands/classification of various e-learning processes, (iv) *Learning Process*: involves activities which are performed by stakeholders to achieve specific goals, (v) *eActivity*: models actions done by a specific actor (e.g., e-learner) using a facilitating tool or combination of them to achieve a goal, (vi) *eContext*: models information that characterises the situation of an entity (e.g., location of learning, environmental attributes, etc.), (vii) *eContent:* models subject domain contents available for e-learners and (viii) *Presentation*: models the way chosen by a specific actor (e.g., instructional designer) to deliver contents.

For the above ontological model, a large number of classes, properties and relations have been designed and developed. Using the above details, the generalised e-learning process can be adapted for a specific e-learner behavioural model, which includes his/her skill, knowledge, preferences, etc. *Second*, the e-learner initiates his/her e-learning process and performs the specified activities. This includes various variations based on the captured contextual information, as explained above. *Third*, an assessment step is needed either by quick quiz, project or other formative assessment tools in order to assess the e-learner's understanding for the presented topic and update his/her model accordingly. *Four*, a decision needs to be made whether the goal of the early-initiated e-learning process has been met or not. If so, the process will be terminated, otherwise the goal or other process elements (e.g., learning contents) need to be further refined to achieve the overall goal of the e-learning process.

As explained in the above discussion, significant variations of the generalised e-learning process can be achieved through out the conditions and gateways available in the BPMN model. One variation could be pure behavioural e-learning process, where the e-learner role remains at the minimum level (i.e., knowledge recipient). Another variation could be self-regulated or problem-based learning process, which allows further participation. A combination of various elements from both types (i.e., a hybrid e-learning process) is possible, as well. This reflects the dynamic nature of the e-learning process. One additional note here is the different interpretations of e-learning activities. For instance, self-regulation and self-monitoring processes might be used interchangeably by some of the e-learners, while they are not, because the latter represents only one phase of the former. To resolve this issue, we have broken them into more obvious sub-tasks (e.g., identifying management strategies and refining goals) to make the e-learning process more traceable and achievable. Finally, it is worth mentioning the scope of the above-developed generalised e-learning process since it only reflects *fine-grained learning-oriented* processes that occur as part of a module. *Coarse-grained* processes that can cover module or programme scale or non-learning-oriented processes are not covered in this research and will remain for future work.

V. DISCUSSION

The proposed generalisation approach is a bottom-up approach, where various e-learning processes have been reviewed from the literature and underpinning theories. The proposed approach to develop a generalised business process from a set of related business processes comprises practice (i.e., how the work is done) and theory (i.e., models/theories underpinning business logic). In this case, the generalised e-learning business process model can be described as *driven by pedagogy and is informed by various e-learning practice models*. This is based on the lessons learnt from the educational domain where e-learners rarely follow one learning theory/approach to achieve their learning objectives [30]. They usually combine elements from different e-learning processes which can be achieved by the proposed hybrid and generic e-learning process model.

Incorporating pedagogy in various stages of developing the generalisation approach is essential since pedagogy explains the added value of using technology in education. For instance, wiki can be used for various purposes, but proper use of pedagogy (i.e., careful consideration for: (i) planning for learning process including the e-learner goals, preferences, knowledge, etc., (ii) the goal of the e-learning process, (iii) the overall settings of the organisation, etc.) can make the use of wiki educationally effective. The adopted classifications of the nine e-learning process models according to their pedagogical strands illuminates further reflections on understanding how different e-learning processes are driven and how they can be assessed against the attainment of their final goals. It also shows the role that Business Process Modelling Notation can play in documenting such rich and dynamic processes and to what extent these technologies can capture the semantics of the e-learning domain. Additional feedback on the modelled e-learning processes is expected to be gained from domain experts and other stakeholders (e.g., instructors, e-learners, institutions, etc.) because modelling processes in BPMN allows them to be understood by non-technical audience, and therefore pave the ground for process improvement.

Various evaluation methodologies have been used to evaluate similar artefacts, such as: dataset-driven evaluation, user studies and real life testing or case studies. Dataset-driven or offline experiment evaluation approaches are widely used in evaluating e-learning artefacts [31]. Datasets used in such experiments can be: (i) extracted from a real system interaction history or (ii) artificially constructed to test the validity of the proposed approach [32]. Real case studies are challenging to adopt due to: (i) the comprehensiveness of information required about pedagogy, learning style, learner knowledge, etc. which means that current e-learning systems do not have such a comprehensive set of data, (ii) time restrictions, (iii) the need for a mature system instead of a prototype, and so on. Therefore, the early-proposed generalisation approach has been evaluated bottom-up by designing a hypothetical case study to test its effectiveness. In this case study, representative and sufficient enough cases have been devised which are based on certain assumptions to check whether the generalised e-learning business process can adapt different e-learning processes/paths. In other words, it tests whether it is possible for a certain e-learner to receive a tailored e-learning business process based on his/her learning profile?

To realise the above-mentioned data-driven approach, the following experimental setup is used: (i) PC with MS Windows 7, service pack 1, 64 bit OS, 4.00 GB RAM, (ii) Eclipse Java EE IDE for web developer version: MARS.1, release 4.5.1, (iii) BPMN 2.0, (iv) Protégé Ontology Editor to develop the e-learning ontological model, specify and instantiate it using Web Ontology Language (OWL 2.0), (v) Pellet Reasoner and (vi) SWRL (Semantic Web Rule Language). For testing, a set of comprehensive test cases/scenarios, acceptance criteria have been derived from the generic e-learning framework requirements [1] and details are covered elsewhere due to space limitations. As a result, the proposed approach demonstrates its ability to deliver behavioural, cognitive or situated e-learning processes based on the e-learner's contextual information. It also confirms its ability to construct a hybrid e-learning approach via combining elements from different categories (e.g., self-regulated e-learning and game-based e-learning processes) based on the e-learner information. Therefore, the proposed e-learning process model is pedagogically independent because various pedagogical models can be equally represented. In addition, it is computationally independent because a standard-based approach has been used for modelling purposes.

This work paves the ground for developing a more mature prototype, where real case study and real users are involved to test the validity of this approach in meeting various e-learners' demands through a flexible process-based approach. Enacted business processes will be orchestrated over cloud or SOA-enabled environment so that stakeholders or e-learners' demands can be met through a set of software services. Also, the proposed approach and the generalised e-learning business process model is technology independent and have no restrictions if compared to other solutions, such as IMS Learning Design. It is also more detailed in terms of covering several e-learning scenarios that could be applied in different disciplines. Additionally, it handles the e-learning processes in a more comprehensive approach than other approaches used in various Adaptive e-Learning Systems or Recommender Systems.

VI. CONCLUSION AND FUTURE WORK

This paper proposed a novel approach to developing generalised e-learning business processes model from a set of related e-learning business processes sharing the same goals and objectives. It has been applied in the e-learning domain, which demonstrates its ability to derive business processes based on surveying the existing models of learning taking into consideration pedagogical models underpinning current e-learning models and technology-enhanced learning artefacts. The proposed hybrid and generalised e-learning business process model is flexible and capable to respond to the dynamic nature of the e-learning processes. Additionally, it has been evaluated to prove its effectiveness. Two further research directions are being accomplished; first is the development of a comprehensive ontological model to effectively contextualise the proposed process models, and hence resolve semantic e-learning heterogeneities. Second, is the enactment of these process models and orchestration of their activities over an SOA-enabled environment.

REFERENCES

[1] R. Hammad, M. Odeh, and Z. Khan, “Towards a generic requirements model for hybrid and cloud-based e-learning systems” in The IEEE 5th International Conference on Cloud Computing Technology and Science (CloudCom), Bristol, UK, Dec. 2013, pp. 106-111.

[2] R. D. Boyd and J. W. Apps. “Redefining the Discipline of Adult Education,” San Francisco, Jossey-Bass, June1980.

[3] M. A. Ould, “Business Processes: Modelling and Analysis for Re-Engineering and Improvement” Bath, Wiley, June 1995.

[4] R. Hammad, M. Odeh and Z. Khan, "eLearner experience model," in The 17th International Arab Conference on Information Technology (ACIT), Morocco, 2016, pp. 20-28.

[5] D. A. Wiley, *Connecting learning objects to instructional design theory: A definition, a metaphor, and a taxonomy* In Wiley, D.A. (Ed.), “The Instructional Use of Learning Objects,” Agency for Instructional Technology, Jan 2002. Available: http://186.113.12.12/discoext/collections/0309/0001/02740001.pdf.

[6] N. Friesen, “Three objections to learning objects and e-learning standards,” In R. McGreal, (Ed.) “Online Education using Learning Objects,” Routledge, London, Dec. 2004, pp. 59-70.

[7] R. Koper and J. Manderveld. “Educational modelling language: modelling reusable, interoperable, rich and personalised units of learning,” British Journal of Educational Technology, 35(5), pp. 537-551. Dec 2004, doi:10.1111/j.0007-1013.2004.00412.x.

[8] M. Rodriguez-Artacho, “PALO language overview,” Universidad Nacional de Education a Distancia. Madrid. Jan 2002, Available: http://sensei.lsi.uned.es/palo/PALO-TR.pdf.

[9] A. Rawlings, P. Van Rosmalen, R. Koper, M. Rodriguez-Artacho, and P. Lefrere, “Survey of educational modelling languages (EMLs),” Version 1,” CEN/ISSS WS/LT Learning Technologies Workshop, Netherland, September 2002.

[10] C. M. Stracke, "Interoperability and quality development in e-learning: overview and reference model for e-learning standards," in the proceedings of the Asia-Europe E-Learning Colloquy (E-ASEM), Seoul, Jan. 2006, pp. 2-26.

[11] M. Derntl, S. Neumann, D. Griffiths, and P. Oberhuemer. “The conceptual structure of IMS learning design does not impede its use for authoring,” IEEE Transaction on Learning Technologies 5(1), August 2011, pp. 74-86, doi: dx.doi.org/10.1109/TLT.2011.25.

[12] J. Torres, J. M. Dodero, I. Aedo, and P. Diaz, “Designing the execution of learning activities in complex learning processes using LPCEL,” Proceedings of the 6th International Conference on Advanced Learning Technologies, Netherland, July 2006, pp. 415-419.

[13] S. Neumann, M. et al, “Report of the results of an IMS learning design expert workshop,” International Journal of Emerging Technologies (iJET) 5(1), Nov. 2009, pp. 58-72.

[14] D. Burgos, “What is wrong with the IMS learning design specification? constraints and recommendations,” Proceedings of Lernen Wissen Adaption (LWA) Conference, Germany, Oct. 2010, pp. 281-288.

[15] J. Yong, “Workflow-based e-learning platform,” Proceedings of the 9th International Conference on Computer Supported Cooperative Work in Design, May 2005, pp. 1002-1007.

[16] R. Gagné Mills., “The conditions of learning,” New York: Holt, Rinehart and Winston, April 1965.

[17] D. H. Schunk, “Learning theories: an educational perspective,” 6th edition, Massachusetts USA: Pearson, Oct. 2012.

[18] J. Greeno, A. Collins, and L. Resnick, “Cognition and learning,” in D. Berliner and R. Calfee (Eds.) “Handbook of Educational Psychology”, NY: Simon & Schuster Macmillan, June 1996, pp. 189-198.

[19] J. B. Watson, “Psychology as the behaviourist views it,” Psychological Review, vol. 20, pp. 158-177, 1913.

[20] E. Wenger and B. Wenger, “Learning in landscapes of practice: a framework,” in E. Wenger, M. Fenton-O'Creevy, S. Hutchinson, C. Kubiak, and B. Wenger (Eds.) Learning in Landscape of Practice: boundaries, identity, and knowledgeability in practice-based learning, Routledge, Oxon, pp. 13-29, May 2015.

[21] S. Downes, “Places to go: Connectivism & connective knowledge,” Innovate: Journal of Online Education, vol. 5(1), Oct. 2008, pp. 11-18.

[22] M. A. Ould, “Business Process Management: A Rigorous Approach,” UK: BCS, The Chartered Institute, Meghan-Kiffer Press, Jan. 2005.

[23] W. Ma, O. O. Adesope, J. C. Nesbit, and Q. Liu, “Intelligent tutoring systems and learning outcomes: A meta-analysis,” Journal of Educational Psychology, vol. 106, June 2014, pp. 901-918, doi: dx.doi.org/10.1037/a0037123.

[24] S. Narciss, “Feedback strategies for interactive learning tasks,” in ed., J. M. Spector, D. M. Merrill, J. v Merrienboer, and M. Driscoll P. (Eds.), Handbook of Research on Educational Communications and Technology , 3rd edition, New York: Lawrence Erlbaum Associates, Feb. 2008, pp. 125-143.

[25] V. Jurković (Ed), “Guide to problem-based learning,” Slovene Association of LSP Teachers, UK, August 2005.

[26] M. S. Knowles, “Self-Directed Learning,” New York, Association Press, Jan. 1975.

[27] F. Karin, A. Nussbaumer, and D. Albert. “A psycho-pedagogical framework for self-regulated learning in a responsive open learning environment,” Proceedings of the International Conference eLearning Baltics Science (eLBa Science 2010). July 2010, pp. 14-23.

[28] J. Lu, D. Wu, M. Mao, W. Wang, and G. Zhang, “Recommender system application developments: A survey,” Decision Support System, vol. 74, June 2015, pp. 12-32, doi: dx.doi.org/10.1016/j.dss.2015.03.008.

[29] M. Gil-Ortega and L. Falconer, “Learning spaces in virtual worlds: bringing our distance students home,” Journal of Applied Research in Higher Education, vol. 7(1), Jan. 2015 pp. 83-98, doi: dx.doi.org/10.1108/JARHE-02-2014-0026.

 ISBN: 978-1-61208-536-4

[30] T. Mayes and S. De Freitas, "Review of e-learning theories, frameworks and models," Joint Information Systems Committee, London, UK. July 2004, [retrieved: Dec, 2016] https://curve.coventry.ac.uk/open/file/8ff033fc-e97d-4cb8-aed3-29be7915e6b0/1/Review%20of%20e-learning%20theories.pdf.

[31] K. Verbert, H. Drachsler, N. Manouselis, M. Wolpers, R. Vuorikari, and E. Duval, "Dataset-driven research for improving recommender systems for learning," Proceedings of the 1st International Conference on Learning Analytics and Knowledge, Feb. 2011, pp. 44-53.

[32] M. Erdt, A. Fernandez, and C. Rensing, "Evaluating recommender systems for technology enhanced learning: A quantitative survey," IEEE Transactions on Learning Technologies, vol. 8(4), June 2015, pp. 326-344.

 ISBN: 978-1-61208-536-4