Data-Driven Modelling: Towards Interpreting and Understanding Process Evolution of In-Service Engineering Projects

Lei Shi[1(✉)], Linda Newnes[1], Steve Culley[1], James Gopsill[2], and Chris Sinder[2]

[1] Department of Manufacturing, University of Bath, Bath, UK
`l.shi@bath.ac.uk, leishi.mail@gmail.com`
[2] Department of Manufacturing, University of Bristol, Bristol, UK

Abstract. Product service plays an essential role in day-to-day operations of nowadays manufacturing industries. However, the changing demands of the market/customers, the increasing complexity of product functionalities and the extended product lifecycles present challenges to related In-Service projects. In order to handle the increasing number of projects and to control the costs and resource consumptions, it is critical to improve the efficiency and automation of process management. Within this context, this paper introduces some data-driven approaches to interpret and represent changes of project process over time in an automatic manner. These approaches aim to help project actors improve their understanding of process structure and the efficiency of process management, and also enable them to investigate process changes from more dynamic perspectives. To evaluate the approaches, a dataset from an aerospace organisation is considered in this paper.

Keywords: In-Service · Engineering project process · Process management · Process evolution

1 Introduction

As an important part of product-lifecycle management (PLM), product service is considered to have direct impacts on the opportunities of marketing, strategy and finance of an enterprise [1]. Under this circumstance, a general trend of modern manufacturing is to reduce the boundary between product service and manufacturing process, or in some instances, to make product service to be part of manufacturing process [2, 3]. Product-service system (PSS) is a formal solution to achieve this. The focus has been aimed at combining services with product design and production according to the specific requirements from customers together with the certain selection of information from the market [4]. The use of PSS has fundamentally changed the managerial and operational approaches of most engineering projects. The direct benefits it can bring include the improvement of sustainability, profitability, market share, as well as the reduction of through-life costs [4, 5].

In practice, the In-Service department of an enterprise is the specialised division that handles day-to-day service related work [6]. The main duty of the department is to solve the issues regarding routine maintenance or emergency repairs of products, whilst collect feedback from customers and then use it to improve the product design in future. The changing demands of the market/customers, increasing complexity of product structure and the extended product lifecycle are the factors that present challenges to the In-Service departments in most manufacturing industries, especially the high-value-manufacturing (HVM) [7]. From an operation perspective, these challenges mainly include, (i) developing and managing the processes for large number of projects under the time and resource constraints, and (ii) improving the efficiency and quality of service work and reducing its cost and resource consumptions.

On the other hand, human decision-making still plays an essential role in many process development and process management approaches [8, 9]. The emphasis of human factors in these approaches means the decision makers (i.e., the project actors) are required to have both macro and micro level understanding of process evolutions in a real-time manner, before they can make any appropriate decisions. However, gaining such comprehensive understanding is not an easy task in practice, which could be prevented by certain issues that caused by the data accessibility, information overload or knowledge gap [10, 11].

Within this context, this paper introduces some data-driven approaches to automatically interpret the changes of engineering project process over time. These approaches aim to help project actors improve their understanding of process structure and efficiency of process management, and also enable them to investigate process changes from more dynamic perspectives.

The paper is organised as follows. Section 2 reviews the related work. Section 3 describes the approach developed. Section 4 includes the evaluation using a collection of industrial data. Section 5 concludes the paper.

2   Related Work

A formal definition of "*servitization of business*" was introduced by Vandermerwe and Rada [12], which refers a series of models that were specifically designed for enterprises on the purpose of adding their product values, increasing their profitability and market shares. In recent years, clear evidence shows that *service* plays an increasingly important role in many manufacturing industries, especially in the industry who produces complex products [3]. The concept of *servitization* directs the strategy transformation of manufacturers in high-value-manufacturing (HVM). As an immediate consequence, most of them have moved from selling products to delivering product-service systems (PSS) [4].

Following the globalisation trend, the design, production and service of products in manufacturing industries nowadays require significant amount of collaborations [13, 14]. As a result, the handling of related engineering projects can be quite a challenge at times. It is mainly caused by factors such as, (i) dealing with distributed and heterogeneous data repositories, (ii) analysing large amount of project data on both macro and micro levels, (iii) planning and managing distributed resources; (iv) exchanging and

sharing project data between project actors, departments and external collaborators, and (v) maintaining or improving efficiency and quantity of service process.

On the other hand, the conflict between the increased amount of service requirements and the limited availability of resources can also be a concern. For instance, the manufacturer Boeing has 20,910 airplanes in service cycle at year-end of 2014. With an average 3.6 % of increment in each year, this number will be doubled in 2033 [15]. To handle the increasing amount of In-Service projects, manufacturers need to find a sustainable way to improve their efficiency of process development and management, and also to control their operating costs and resource consumptions.

Business Process Management (BPM) is an approach being proposed to model and re-develop the existing business process, which has been used to improve the process efficiency, effectiveness and adaptability [16]. It covers the fields of process re-engineering, standardisation, optimisation and simulation [17, 18]. As stated by Davenport [17], the re-engineering of business process has direct impacts on the improvement of innovation and sustainability for an enterprise. Meanwhile, the process standardisation is considered to be useful to reduce the process variability, so that it can be used to control warranty costs and resource consumption, as well as to improve the effectiveness on related decision making tasks [19].

According to Melão and Pidd [20], the understating of business process is one of the critical requirements to implement effective process management. Moreover, the analysis of project process is considered to be critical for discovering the process norms and handling the process exceptions [21]. To improve the understanding, event-based data has been applied to model or simulate the business process from bottom-up perspectives [22]. The use of information technology is considered to be essential for most BPM models, as the technology can effectively improve the automation of their process modelling and process analysis functions [17]. Recent research also highlighted that the use of "big data" is a promising way to improve the capability and rationality of BPM [23]. As stated in [24], advanced approaches from data mining and machine learning are necessary to be integrated with BPM when large amount of project data is required to be dealt with. The use of such approaches could help reduce the time consumption and the need of human effort, and also to enable project actors to improve their understanding capability of process dynamics.

There are various existing approaches and tools for managing business process, although improving the level of automation is still an ongoing task. Within this context, this paper introduces a data-driven approach to automatically model project process and represent process evolution.

3   Data-Driven Process Management

This work deals with data that is generated by the operation process of engineering projects. The essential information contained in the data needs to be extracted, and then used to construct the actual process of the projects. Each information item is required to have a textual form. For example, communication related or technical data contains email, fax, report or documentation, which would include textual content by their nature;

other types of data, such as image or drawing, may not contain sufficient textual content, hence the metadata and annotations may need to be used.

3.1   Process Interpretation

The operation process of an engineering project can be represented by a collection of activities. These activities are typically organised in a dynamic form in order to adapt both internal and external condition changes during project execution phases. In practice, the internal condition changes may refer to the characteristic changes of a project, and the external condition changes may refer to the environmental changes with regard to the project.

For a process, the data generated by the current/previous activity, together with the data provided by certain external sources, are applied to make the decision on what the next activity is supposed to be (see Fig. 1). During project execution, the iteration of decision-making may need to be performed a number of times, in order to generate the finalised project process. As the input data, project characteristics are considered as key factors to determine the quantity of the iterations and the type/dependency of the activities. On the other hand, the generated data by these activities also determines the possible changes of project characteristics in future. Therefore, the data generated by such iteration steps can be used to measure the evolution of project process.

There are two benefits of using this representation (as shown in Fig. 1) to interpret project process: (i) it emphasises the time dimension of the process and related activities, and (ii) it filters out the less important information contained in the process structure.
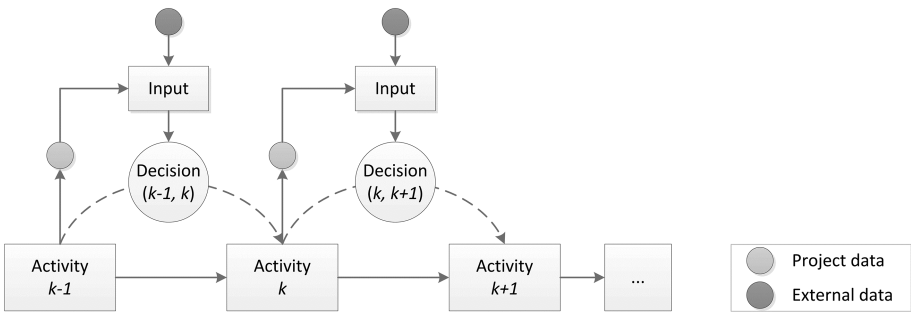

Fig. 1.   A project process with data and decisions

3.2   Evolution Identification

As discussed in Sect. 2, understanding the process structure and its evolution is a critical requirement for implementing process standardisation, process optimisation, and handling process exceptions. As a dynamic variable, the evolution of a project process can reflect the actual changes of the project characteristics, operation performances and implementation constraints. In this work, the occurrence of certain activities of a process

is used as an indication to measure the temporal changes of the process structure. On this basis, the distribution of each single activity is be generated and analysed.

As shown in Fig. 2, the representation of modelled process is structured as a linear format. To investigate the process evolution, this linear representation is segmented into multiple partitions based on pre-defined intervals.
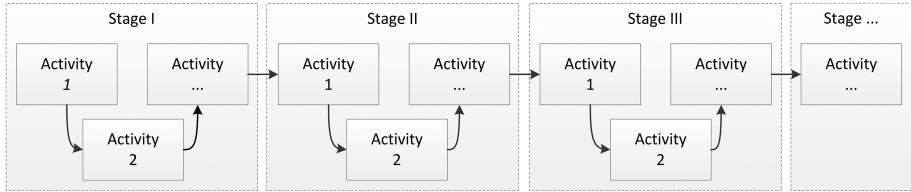
Fig. 2. The linear interpretation of a project process

There are four different interval types being defined in this work, (i) absolute step interval, (ii) normalised step interval, (iii) absolute time interval, and (iv) normalised time interval. The *step* variable indicates the atomic component of the modelled process, e.g., a single activity. The *time* variable indicates the timestamp of a single activity.

Giving a modelled process that includes 20 activities, which is denoted by $[a_1, a_2, \ldots, a_{20}]$. If the setting of absolute step interval is 5, then the process should be segmented into $20/5 = 4$ partitions, such as $[a_1, \ldots, a_5]$, $[a_1, \ldots, a_{10}]$, $[a_1, \ldots, a_{15}]$ and $[a_1, \ldots, a_{20}]$. If the setting of normalised step interval is 40 %, then the process should be segmented into $ceiling[20/(20 * 40 \%)] = 3$ partitions, i.e., $[a_1, \ldots, a_8]$, $[a_1, \ldots, a_{16}]$ and $[a_1, \ldots, a_{20}]$. Assuming the modelled process has a timeline that equals 10 days, when the setting of absolute time interval is 5 (days), the process should be segmented into $10/5 = 2$ partitions. If the setting of normalised time interval is 40 %, then the process should be segmented into $ceiling[10/(10 * 40 \%)] = 3$ partitions.

4 Evaluation

To evaluate the proposed approaches, a dataset captured from the In-Service department of an aerospace manufacturer is considered. This evaluation aims to investigate, (i) whether the project process can be automatically constructed from the project data, and (ii) whether the process evolution can be automatically identified and represented. The detailed information about the dataset, evaluation process and evaluation results are introduced in the following sections.

4.1 Data Collection

The applied dataset in this evaluation contains 396 In-Service projects that were completed between 2013 and 2014. For each project, all the essential data was recorded during project execution. The project data mainly includes communication related (35.11 %), operation related (49.10 %) and test/evaluation related (15.79 %). By

considering the knowledge captured from the senior staff in the department, the information items contained by the project data are classified into 21 types, each of which is considered to be generated from, or associated with, a particular activity. For example, the information item OM indicates the activity "sending an outgoing message to the customer"; IM indicates the activity "receiving an incoming message from the customer"; S&F indicates the activity "performing a stress and fatigue test", etc.[1]

4.2 Creating Process Interpretation and Identifying Process Evolution

In order to determine the activity types, automatic data analysis techniques are applied, such as natural language processing (NLP) and named entity recognition (NER). These techniques analyse the project data on both metadata level and content level. After the data analysis, all the activities (with the timestamps) contained in the project data are identified and extracted, and then the project process is modelled and interpreted in a sequence format. Figure 3 shows some processes that are generated by applying the approach and the dataset. In this figure, each row indicates a generated project process, and each '*Tx*' indicates an activity. For each process, its contained activities are organised in a chronological order.

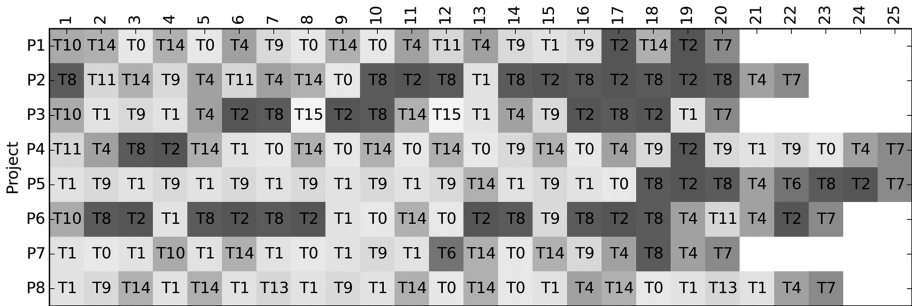| Project	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
P1	T10	T14	T0	T14	T0	T4	T9	T0	T14	T0	T4	T11	T4	T9	T1	T9	T2	T14	T2	T7					
P2	T8	T11	T14	T9	T4	T11	T4	T14	T0	T8	T2	T8	T1	T8	T2	T8	T2	T8	T2	T8	T4	T7			
P3	T10	T1	T9	T1	T4	T2	T8	T15	T2	T8	T14	T15	T1	T4	T9	T2	T8	T2	T1	T7					
P4	T11	T4	T8	T2	T14	T1	T0	T14	T0	T14	T0	T14	T0	T9	T14	T0	T4	T9	T2	T9	T1	T9	T0	T4	T7
P5	T1	T9	T1	T9	T1	T9	T1	T9	T1	T9	T1	T9	T14	T1	T9	T1	T0	T8	T2	T8	T4	T6	T8	T2	T7
P6	T10	T8	T2	T1	T8	T2	T8	T2	T1	T0	T14	T0	T2	T8	T9	T8	T2	T8	T4	T11	T4	T2	T7		
P7	T1	T0	T1	T10	T1	T14	T1	T0	T1	T9	T1	T6	T14	T0	T14	T9	T4	T8	T4	T7					
P8	T1	T9	T14	T1	T14	T1	T13	T1	T9	T1	T14	T0	T14	T0	T1	T4	T14	T0	T1	T13	T1	T4	T7		

Fig. 3. The sequence interpretation of modelled processes

To identify process evolution, each of the modelled processes is segmented based on pre-defined intervals. In this evaluation, the interval setting applied in process segmentation is *normalised step interval* (NSI). The total number of intervals is set as four, so the intervals are 0–25 %, 0–50 %, 0–75 % and 0–100 %, respectively. For example, the 0–25 % interval means the segmented process partition should contain 25 % of the total activities; similarly, the 0–50 % interval means the process partition should contain 50 % of the total activities.

To investigate process evolution on a detailed level, the activity distribution of each interval is taken into account. The detailed information of each activity distribution is shown in Fig. 4.

---

[1] Due to confidentiality reasons, the full descriptions of these activities are not included in this paper, and some activity names have been deliberately masked.
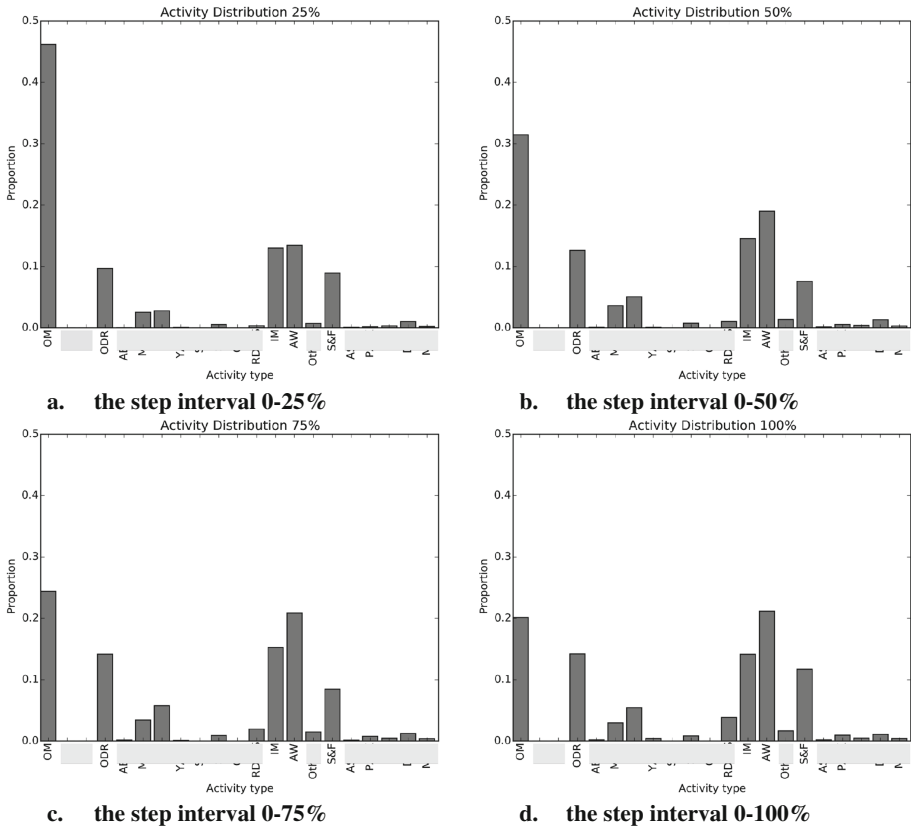
Fig. 4. The activity distributions with normalised step intervals

Figure 4a shows the activity distribution being generated from the initial project stage (0–25 %). As shown in the figure, the top activities involved by these processes are OM (46.17 %), IM (12.98 %), AW (13.45 %), ODR (9.66 %) and S&F (8.91 %). OM and IM are the communication related data, which indicate the commutations between the customer (e.g., an airline operator) and the In-Service department. Meanwhile, ODR and AW are the operation related data. ODR is generated by the customer, for the purpose of describing service requirement or inquiry. AW is generated by the In-Service department, for the purpose of providing technical solution or responding to the inquiry. In fact, most of the work is planning related at this stage. Hence, the service requirements, general inquiries and communications are supposed to take high proportions. It can be seen that the activity distribution generated from the segmented processes corresponds to the mentioned facts in practice.

Figure 4b shows the activity distribution being generated from the initial stage to early-mid stage (0–50 %). The top activities involved by these processes are OM (31.46 %), AW (19.01 %), IM (14.52 %), ODR (12.64 %) and S&F (7.58 %). After the initial stage, most of the planning work ought to be finished, and this fact may reduce the amount of outgoing communication (OM). Meanwhile, most of the general inquires

raised by customers are supposed to be solved, so that the detailed service requirements (ODR) from the customers can be formally submitted to the In-Service department. The department therefore needs to issue the technical solutions accordingly. As shown in Table 1, the patterns of the activity distribution reflect the facts: OM has the decrement of 14.71 %; ODR and AW have the increments of 2.97 % and 5.56 % respectively.

Table 1. Changes of the distribution over time

	OM	IM	ODR	AW	S&F
NSI (25 %–50 %)	−0.1471	+0.0155	+0.0298	+0.0556	−0.0133
NSI (50 %–75 %)	−0.0704	+0.0071	+0.0154	+0.0184	+0.0090
NSI (75 %–100 %)	−0.0431	−0.0108	+0.0004	+0.0033	+0.0324

Figure 4c shows the activity distribution being generated from the initial stage to mid-late stage (0–75 %). The top activities involved by these processes are OM (24.42 %), AW (20.85 %), IM (15.24 %), ODR (14.17 %) and S&F (8.48 %). At this stage, the major service work needs to be completed. Hence, the quantity of submitted service requirements and issued technical solutions tend to be increased, by comparing to the previous stages. According to Table 1, the patterns of the activity distribution correspond to the facts: ODR and AW have the increments of 1.54 % and 1.84 % respectively.

Figure 4d shows the activity distribution being generated from the initial stage to the final stage (0–100 %). The top activities involved by these processes are AW (21.18 %), OM (20.11 %), ODR (14.21 %), IM (14.15 %) and S&F (11.73 %). At this stage, most of the major service work has been completed, so that the amount of operation related activities should be decreased or remained at the similar level. Meanwhile, the amount of test/evaluation related activities (S&F) is expected to have an increment. As shown in Table 1, the patterns of the activity distribution again correspond to the facts: both ODR and AW remain stable; S&F has the increment of 3.24 %.

According to the evaluation, it can be seen that, (i) a project process can be automatically generated from the project data by applying the proposed approach; (ii) the activity distribution of segmented process is useful to investigate and understand the process changes over time.

From a project management perspective, the generated activity distributions with their patterns can provide explicit indications to project actors, which could enable them to gain comprehensive understandings of process structures from dynamic perspectives. Such results can also improve their awareness of process norms and process exceptions. Using different interval settings, the project actors are able to investigate modelled processes on different granularity levels. When dealing with large number of projects, these approaches could help the project actors reduce the time and efforts put into the process understanding, and also improve the rationality of related decision-making tasks.

5 Conclusions

To enhance the understanding of process structure and process evolution of engineering projects, data-driven approaches on process interpreting, segmentation and analysis have been introduced in this paper. The application of these approaches aims to reduce human efforts in process understanding, and also to improve the efficiency and automation of process management. By using an industrial dataset, the evaluation of this work reveals that the introduced approaches have the capability of automatically modelling project process and representing the process changes over time. These approaches are considered to have the potential to help project actors understand the process dynamics, standardise/optimise the existing processes, and also improve the awareness of process norms/exceptions.

References

1. Gebauer, H., Krempl, R., Fleisch, E.: Service development in traditional product manufacturing companies. Eur. J. Innov. Manag. **11**(2), 219–240 (2008)
2. Mont, O.K.: Clarifying the concept of product–service system. J. Cleaner Prod. **10**(3), 237–245 (2002)
3. Neely, A., Benedettini, O., Visnjic, I.: The servitization of manufacturing: further evidence. In: 18th European Operations Management Association Conference (2011)
4. Baines, T.S., Lightfoot, H.W., Evans, S., Neely, A., Greenough, R., Peppard, J., Roy, R., Shehab, E., Braganza, A., Tiwari, A., Alcock, J.R., Angus, J.P., Bastl, M., Cousens, A., Irving, P., Johnson, M., Kingston, J., Lockett, H., Martinez, V., Michele, P., Tranfield, D., Walton, I.M., Wilson, H.: State-of-the-art in product-service systems. Proc. Inst. Mech. Eng., Part B: J. Eng. Manuf. **221**(10), 1543–1552 (2007)
5. Kreye, M., Goh, Y.M., Newnes, L.B.: Uncertainty in through life costing within the concept of product service systems: a game theoretic approach. In: DS 58-7: Proceedings of ICED 09, The 17th International Conference on Engineering Design, vol. 7, Design for X/Design to X, Palo Alto, CA, USA, 24−27 August 2009
6. Carey, E., Culley, S., Weber, F.: Establishing key elements for handling in-service information and knowledge. In: The 19th International Conference on Engineering Design, ICED13, Seoul, South Korea (2013)
7. Xie, Y., Culley, S., Weber, F.: Applying context to organize unstructured information in aerospace industry. In: DS 68-6: Proceedings of the 18th International Conference on Engineering Design (ICED 11), Impacting Society Through Engineering Design, vol. 6, Design Information and Knowledge, Lyngby/Copenhagen, Denmark, 15−19 August 2011 (2011)
8. Van den Bergh, J., Thijs, S., Viaene, S.: Focusing on BPM's human factor: an interview with Els Van Keymeulen of Schoenen Torfs. Transforming Through Processes. SpringerBriefs in Business Process Management, pp. 41–43. Springer International Publishing, Switzerland (2014)

9. Han, Y., Kauranen, A., Kristola, E., Merinen, J.: Human interaction management–adding human factors into business processes management. Special report for information systems integration, Helsinki University (2007)
10. Bettis-Outland, H.: Decision-making's impact on organizational learning and information overload. J. Bus. Res. **65**(6), 814–820 (2012)
11. Pfeffer, J., Sutton, R.I.: The Knowing-doing Gap: How Smart Companies Turn Knowledge into Action. Harvard Business Press, Brighton (2013)
12. Vandermerwe, S., Rada, J.: Servitization of business: adding value by adding services. Eur. Manage. J. **6**(4), 314–324 (1988)
13. Ming, X.G., Yan, J.Q., Wang, X., Li, S., Lu, W.F., Peng, Q., Ma, Y.: Collaborative process planning and manufacturing in product lifecycle management. Comput. Ind. **59**(2), 154–166 (2008)
14. Hicks, B.: The language of collaborative engineering projects. In: DS 75-6: Proceedings of the 19th International Conference on Engineering Design (ICED13), Design for Harmonies, vol. 6, Design Information and Knowledge. Seoul, Korea (2013)
15. Huls, D.: Current Market Outlook 2014−2033. Boeing Commercial Airplanes, USA (2014)
16. van der Aalst, W.M., ter Hofstede, A.H., Weske, M.: Business process management: a survey. In: Aalst, W.M., Hofstede, A.H., Weske, M. (eds.) BPM 2003. LNCS, vol. 2678, pp. 1–12. Springer, Heidelberg (2003)
17. Davenport, T.H.: Process Innovation: Reengineering Work Through Information Technology. Harvard Business Press, Brighton (2013)
18. Laguna, M., Marklund, J.: Business Process Modeling, Simulation and Design. CRC Press, Boca Raton (2013)
19. Banuelas Coronado, R., Antony, J.: Critical success factors for the successful implementation of six sigma projects in organisations. TQM Mag. **14**(2), 92–99 (2002)
20. Melão, N., Pidd, M.: A conceptual framework for understanding business processes and business process modelling. Inf. Syst. J. **10**(2), 105–129 (2000)
21. Grambow, G., Oberhauser, R., Reichert, M.: Event-driven exception handling for software engineering processes. In: Daniel, F., Barkaoui, K., Dustdar, S. (eds.) BPM Workshops 2011, Part I. LNBIP, vol. 99, pp. 414–426. Springer, Heidelberg (2012)
22. Weijters, A., Van der Aalst, W.: Process mining: discovering workflow models from event-based data. In: Belgium-Netherlands Conference on Artificial Intelligence. Citeseer (2001)
23. Chen, H., Chiang, R.H., Storey, V.C.: Business intelligence and analytics: from big data to big impact. MIS Q. **36**(4), 1165–1188 (2012)
24. Shi, L., Gopsill, J.A., Newnes, L., Culley, S.: A sequence-based approach to analysing and representing engineering project normality. In: 2014 IEEE 26th International Conference on Tools with Artificial Intelligence (ICTAI), pp. 967–973. Limassol, Cyprus (2014)