

A Process Mining Methodology for People Analytics: With a Case Study on Recruitment Analysis

Emiel Caron and Nijs Niessen

Tilburg University, Department of Management, Warandelaan 2, PO Box 90153, Tilburg, The Netherlands

Keywords: HR Analytics, People Analytics, People Management, Process Mining, Process Mining Methodology.

Abstract: In today's competitive business environment, acquiring data about Human Resource (HR) processes and optimizing operational excellence are some of the main objectives. Process mining as a specific form of HR Analytics aims at translating data, stored in the HR information systems of the organizations into insights about the organization's HR processes. These insights are then translated into improvements for HR processes and to enhancements in the compliance to rules and regulations. However, so far there is no concrete and comprehensive framework for the execution of process mining analysis that can be used practically by the HR professionals. In this paper, we develop a methodology for HR Analytics, based on PM²; a step-by-step method with the best practices and approaches to process mining projects in the HR domain. Additionally, we establish definitions such as the HR event log, the business process, etc. that capture the specifics of the HR domain. Finally, we demonstrate the effectiveness of the proposed methodology by applying it to a case study on the recruitment process. The results show that the methodology successfully identifies areas for improvement and provides insights that can enhance the overall HR recruitment process.

1 INTRODUCTION

Human Resource Management (HRM) as a business function aims to effectively deploy and manage employees within an organization in order to gain a competitive advantage. A new trend within the field of Human Resources (HR) is the use of HR Analytics (HRA) or people analytics (Marler and Boudreau, 2017), which is, as opposed to the traditional HRM, a data-driven approach to better understand and optimize HR processes and practices, like recruitment and selection, performance management, employee turnover, training and development, and so on.

HR analytics are founded on the Human Resource Information Systems (HRIS). Chauhan, Sharma & Tyagi (2011) define the HRIS as: “[...] the integration of software, hardware, support functions and system policies and procedures into an automated process designed to support the strategic and operational activities of the human resources department and managers throughout the organization.” By providing timely, accurate and understandable information the HRIS's support decision-making for HR professionals. HRIS serve as integral components of the business processes they support. Business processes are defined as a chain of

linked activities that transform inputs into value added output to achieve the company's objectives, by applying one or more resources (Christensen et al., 2016).

Potential techniques to produce HR analytics, that can utilize the process data within HRIS, are process mining techniques (Aalst, 2012), that focus on the discovery, monitoring and improvement of processes by extracting knowledge from event logs in information systems. In this research, we want to engage with three process mining research challenges, identified by R'bigui and Cho (2017), by using process mining in an HR context. These are: 1) combining process mining with other type of analysis and to enrich event log data, 2) improving the usability for non-experts, in this case HR professionals, and 3) improving the understandability for non-experts, to produce process models that can be used for HR decision-making.

Moreover, the objective of this research is fourfold: 1) to extend and specialize PM² - the process mining project methodology (Van Eck et al., 2015) - for the specific development of HR and people analytics, 2) to define the concept of an event log in the case of a HR information system, 3) to apply the specialized methodology on a real case study on

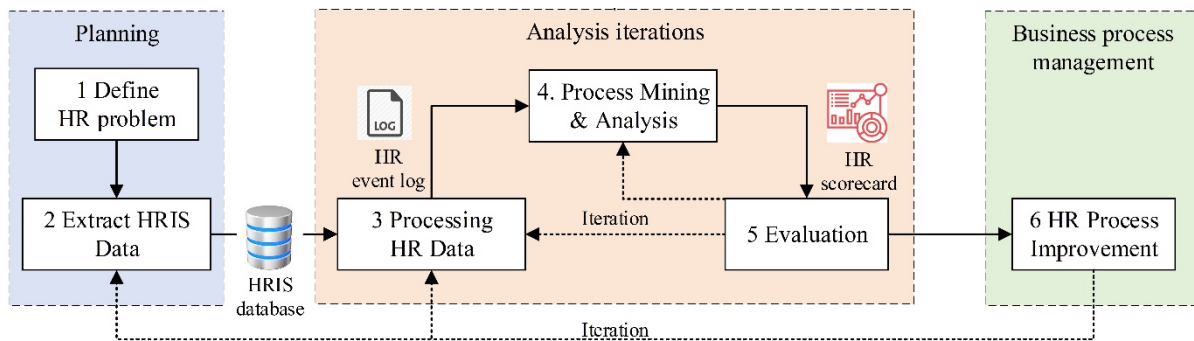


Figure 1: PM^{2HR}, Process mining methodology for HR, adapted from Van Eck et al. 2015.

employee recruitment and selection, and 4) to enhance the current limited scientific literature (Arias et al. 2018, Jooseok et al. 2018) on the utilization of process mining in the field of HRM.

The remainder of this paper is structured as follows. In Section 2 we develop the process mining framework for people analytics. Subsequently, in Section 3 we apply the framework on a case study for recruitment analysis. Conclusions are drawn in Section 4.

2 METHODOLOGY FOR HR ANALYTICS: PM^{2HR}

We propose a specialisation of the six stages (Section 2.1-2.6) of the PM² process mining project methodology (Van Eck et al., 2015), to make it useful for HR processes and analytics, called PM^{2HR}. The initial planning of the project consists of: the definition of the HR decision problem and the related data extraction from the HR data source (stages 1-2). After that extracted event data then goes through several analysis iterations. In each iteration, HR data processing, process mining & analysis, and evaluation are executed (stages 3-5). Lastly, gained insights are used to modify the actual HR business process (stage 6) that takes place in the company, i.e. process improvements as part of business process management. An overview of PM^{2HR} is shown in Figure 1. In the following sub sections we discuss HR specific perspectives for the six stages.

2.1 Stage 1. Define the HR Business Problem

The goal of the planning step is to start the people analytics project and determine the initial HR business questions. Van Eck et al. (2015) define three activities in this stage: identifying business questions,

selecting business processes, and composing a project team. People analytics, can help answer a variety of HR-related business questions. Example HR business process and related business questions are:

- Recruitment and selection: What is the time-to-hire for open positions? How many candidates does it take to make a successful hire?
- Employee onboarding: How long does it take for new employees to become fully productive? Which onboarding activities are most effective in achieving this?
- Performance management: How often are performance reviews conducted? What is the average performance rating across the organization?
- And so on for other HR business processes that are supported HRIS like training and development, compensation and benefits, employee engagement, turnover and retention.

Moreover, the PM^{2HR} methodology addresses two important research challenges (R'bigui and Cho, 2017): enhancing the usability and understandability of process mining models for non-experts, specifically HR practitioners. By involving HR practitioners directly in the project team alongside process mining experts, both the managerial and analytical perspectives are integrated from the outset. This approach results in process models that better align with HR objectives.

2.2 Stage 2. Extract HR Data

In the data extraction stage the relevant event data, and optionally ex-ante process models, are extracted from the HRIS, that support the HR processes identified in the previous stage. This stage consists of three activities (Van Eck et al., 2015): determining scope, extracting raw event data and transferring process knowledge. HRIS provide a wide variety of functions in a company, Chauhan et al. (2011) separate these functions into three categories:

operational, tactical, and strategic. HRIS are often, partially, process-aware information systems (Van der Aalst, 2009), as they are driven on the basis of HR process models (see Stage 1). In this step the output is the *HRIS database*, that captures all raw data required to answer the business question, so specific event log data and other relevant HR data.

2.3 Stage 3. Processing HR Data

After the initialization phase, the analysis iterations begin with the processing of the data from the HRIS database. The objective of this stage is to create HR event logs from the extracted HRIS data and prepare them for the mining & analysis stage. Processing activities include (Van Eck et al., 2015): creating views, aggregating events, enriching logs and filtering logs.

Here we focus on the generic data requirements for using process mining in the domain of HRA. The HR event log is a necessary element for process mining in the HR domain. We define the *HR event log* as: a (processed) data set that satisfies the three minimum (1-3) properties for process mining as specified in the XES standard¹: 1. a Case ID, 2. an Activity name, and 3. a Time-stamp, and has additional properties (4-6) that are related to the HR decision problem. The additional properties are:

4. Cases relate to people, typically employees. In HR, process instances are typically related to employees, where the instances describe the employee's life cycle with regards to its recruitment, internal mobility, educational path, etc.

5. People data needs to follow strict legislation. This property follows from 4. Since the data is extracted from the HRIS and it contains data about people, it is necessary that it complies to certain rules and regulations, like the General Data Protection Regulation² (GDPR) in Europe. In particular, some of its principles include that the data is required to be processed lawfully and in a transparent manner as well as accurate and collected for explicit and legitimate purposes.

6. The inclusion of attributes related to the HR The HR event log must include attributes that are specific to the HR process is analyzed. Examples of HR specific attributes include details related to the: employee, job, organization, HR process, etc. For example, the division (i.e. Finance, Marketing) that is responsible for performing an activity or the job position of the employee (i.e. Manager, Senior Manager, Junior Analyst, etc.).

¹ <https://xes-standard.org/>

2.4 Stage 4. Process Mining & Analysis

In this stage the HR event logs are analyzed using process mining techniques to answer the research questions and gain insights into the HR business via dedicated analytics. In this stage four types of activities are identified (Van Eck et al., 2015): process discovery, conformance checking, enhancement, and process analytics. The first three activities are the main classes of process mining techniques (Aalst, 2012). Process analytics constitute the class of all other analysis techniques, e.g. data mining, statistics, dashboards and scorecards etc., which can be applied in the context of the business process, often in combination with the first three activities. In addition, process mining covers different analytical perspectives (Aalst, 2016): the control-flow perspective that focuses on the ordering of activities, the organizational perspective that focuses on the resource, and the case perspective that includes all properties of the cases.

From an HR professional's perspective, the primary goal of process analytics is to develop specific metrics and key performance indicators (KPIs) that can be used to manage HR business processes effectively. These metrics are then combined into an HR performance scorecard or dashboard, making them easily accessible and visible to HR professionals. By doing so, the focus shifts to developing metrics that directly relate to the business problem at hand. To compute these metrics, various classes of process mining techniques are utilized.

In Table 1, we present an example HR performance scorecard with relevant metrics for the recruitment process. Muensterman et al. (2009) and Laumer et al. (2014) define metrics and targets related to the recruitment process in organizations over the dimensions: process time (e.g. time-to-hire), process costs (cost-per-hire) and recruitment process quality. The HR performance scorecard uses information from the derived process model, visualized by process mining and captured in process statistics. Where some process models can be very complex and inelible, the scorecard allows for a structured way of describing what is happening in different models at a glance. Similar scorecards can be produced for other HR business processes.

² <https://gdpr-info.eu/>

Table 1: Example HR performance scorecard with metrics related to recruitment.

Metrics	Definition
Time to Screen CV (Recruiter)	Time it takes Recruiter to complete screening
Time to Screen CV (Hiring Manager, HM)	Time it takes Recruiter to complete screening
Time to Hire	Time it takes to reach status Hired after initial submission of application
Time to Select	Time between screening by HM and when the offered candidate receives an offer.
Time Interview to Offer	Time it takes for a candidate to receive an offer after first interview
Ratios	Definition
Requisition Hit rate	% of applications that were screened by the HM
Interview rate	% of applicants that have been interviewed after a screening by the HM

2.5 Stage 5. Evaluation

During this stage, the findings are evaluated to refine research questions and generate improvement ideas. This stage includes three main activities: diagnose, verify, and validate. Diagnosing the findings involves separating expected and unexpected findings and refining new research questions. For instance, if the average time to promotion between two functions is much shorter than expected, further questions may be raised to analyse this finding in more detail and determine whether the effect is global or only observed in a subset of employees. The next step is to verify and validate the results. Verification entails comparing the findings to the original data to ensure their correctness. For example, verification might reveal that an unexpected finding was caused by errors in data preparation. On the other hand, validation involves comparing the findings to the claims, hypotheses, or knowledge of stakeholders. In this activity, HR professionals are crucial to validate the findings, as process analysts may misinterpret the results. Validation may also reveal the cause of unexpected findings, such as when it is discovered that two functions are very similar and should be combined into one, as seen in the example above. At this stage, there are two possible follow-ups: refined research questions can be investigated in a subsequent analysis iteration, or improvement ideas can be applied in the HR decision-making support phase. Overall, the evaluation stage is essential to ensure that the process mining findings are valid and reliable, and

that they provide actionable insights for HR process improvement.

2.6 Stage 6. HR Process Improvement

The final stage relates to use the gained insights to improve the HR processes. This stage involves two activities, adapted from van Eck et al. (2015): HR policy improvement and supporting operations. For example, if the analysis identifies issues with the internal mobility of employees, improvement ideas can be used to guide HR decision-making and modify the internal mobility process through HR policy change. It is important to note that improving HR policy is a different area of expertise from process mining, and it may require a separate project. One potential improvement is to change recruitment policy for certain positions in response to the finding that external hires often leave while internal hires do not. This can help to reduce employee turnover and ensure that the organization retains valuable talent. The supporting operations activity involves using the improvement ideas to support daily HR operations. This can include implementing new tools or processes that align with the improved HR policies. For example, if the HR policy is changed to encourage internal mobility, new training programs can be developed to help employees adopt new skills and prepare for new roles.

3 CASE STUDY: INSIGHTS IN RECRUITMENT PROCESSES

In this section we apply and evaluate PM^{2HR} in a case study to provide insights on the recruitment process of a company in the global technology industry.

Stage 1. Define the HR Business Problem.

The company's primary focus is to enhance operational excellence in its recruitment and selection process, aiming to execute it more strategically, consistently, and reliably. In simpler terms, the business goal is to optimize the recruitment process to achieve better outcomes. The HR professional defines the business problem. In this case it is about “gaining insights in the performance of its recruitment process such that suggestions for process improvements are developed”, and questions like “How does the actual recruitment and selection process look like?”, “Are there any bottlenecks in the recruitment process?”, “What metrics can be used to measure the performance of the recruitment process”,

and “What are suggestions for process improvements?”.

Stage 2. Extract Recruitment Data.

The data is extracted from the Applicant Tracking System (ATS) database by SQL scripts. The ATS is integrated in the HRIS and it represents a specialized information system for the recruitment process. When the ATS data is extracted from the HRIS database, the data is transformed and stored into three different tables: Requisitions, Applications, and Application status. The tables contain information about the requisition ID, the activity performed (i.e. screen Curriculum Vitae (CV) by recruiter, interview, hire), which division is responsible for the requisition, personal details of the applicant (name, date of birth, education etc.), submission details of the application and several time-stamps indicating the start and the end date of the requisition. In this case, the data extracted from the ATS database, after preparation, satisfies the properties of the HR event log (see Section 2.3). Therefore, process mining can be applied meaningfully.

Stage 3. Prepare Recruitment Event Log.

The extracted data has a csv file format. In order to use process mining, it is necessary that the data is transformed to an HR event log. The data transformation is implemented using Python. The data structure was altered such that the three different tables were merged into one single table, to satisfy the properties of the HR event log. Therefore, the result of the data processing and transformation is the HR event log. In particular, it contains the three requirements of an event log: a Case ID (req.user.id), an activity name (applicant.status), a timestamp (start.time, end.time) and other additional attributes that are listed as other columns: background, education, division, sub.division, requisition.region. The HR event log contains data of the years 2018-2019, with 250,000 cases (applications), nearly 600,000 events (application activity), and 2,500 different process variants.

Stage 4. Process Mining And Analysis.

The next step is to construct a process mining model using the HR event log. Disco software³ is used to further explore the HR event log, examine its quality and analyse the recruitment process. In this case, the HR event log analysis is both explorative and goal driven. Thus, the HR professional implements a control flow analysis as well a performance analysis. First, the normative recruitment process (NRP) is discovered using fuzzy miner together with the

complete recruitment process and secondly the process efficiency is measured using certain metrics and ratios. The tasks in the NRP are: Screen CV by Recruiter → Screen CV by Hiring Manager (HM) → Interview candidate → Make offer → Hired / Close application.

Discovering The Actual Recruitment Process.

In Figure 2, a process model for the full recruitment process is discovered, by using the filtering parameters to obtain a clear and eligible model. The main activities are clearly visible and an almost desirable recruitment is shown.

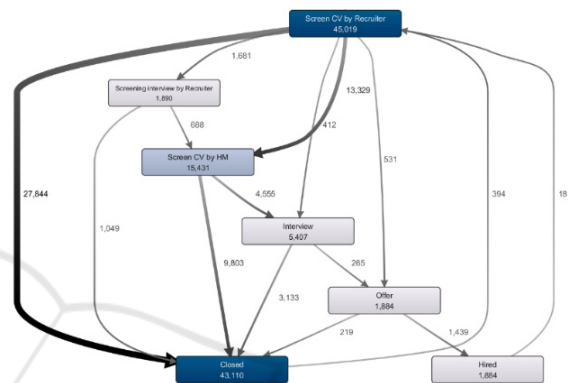


Figure 2: Discovered (filtered) recruitment process.

A new addition to the activities of the NRP is the ‘Screening interview by Recruiter’ step. In the figure, some unexpected paths are still visible: Cases going from ‘Screen CV by Recruiter’ to either ‘Interview’ or ‘Offer’ and cases going from ‘Closed’ or ‘Hired’ back to ‘Screen CV by Recruiter’. When drilling down on these cases that go from ‘Closed’ to ‘Screen CV by Recruiter’, the variants show that most of these steps occur because the case got either re-opened by the recruiter, or maybe the applicant applied again, triggering the ATS to change the status to ‘Screen CV by Recruiter’. In Table 2, the five most frequent variants for hires are given. Comparing the results to the variants of the NRP shows that most of the variants and their frequency stay the same, except that a new activity is introduced for the desirable variant: ‘Interview 2’ is added.

Process Efficiency: Recruitment Performance.

Following the recruitment process performance scorecard (see Table 1), we produce the metrics using the software. Results given in the scorecard, depicted in Table 3, show that the recruitment process does not reach the targets set by business experts. The differences between the target and the actual data

³ <https://fluxicon.com/disco/>

Table 2: Recruitment process hire variants.

Hire variants	Relative
Screen CV by Recruiter→Offer→Hired	21.65%
Offer→Hired	14.45%
Screen CV by Recruiter→Hired	13.58%
Screen CV by Recruiter→Screen CV by HM→Interview→Interview2→Offer→Hired	9.63%
Screen CV by Recruiter→Screen CV by HM→Offer→Hired	2.70%
Other hire variants	23.15%

ranges from 0.5 days to almost 26 days. Especially the metrics for ‘Time to Select’ and ‘Time Interview to Offer’ are much higher than their targets, while the ‘Time to Hire’ is only marginally higher than the target. Following the pipeline requisitions, promising candidates go through the real requisitions relatively fast, with a minimum of two days. After further investigation, applications to these real requisitions skip activities like ‘Screen CV by HM’ or ‘Interview’, which are recorded in the pipeline requisition, resulting in a lower mean score for ‘Time to Hire’ and exclusion from the ‘Time to Select’ and ‘Time Interview to Offer’ metric, which are measured from ‘Screen CV by HM’ and ‘Interview’.

Table 3: Recruitment performance scorecard.

Metric	Target	Mean	Delta	Median	Min / Max
Time to Screen CV by Recruiter	5	5.5	0.5	2.6	0 / 40.4w
Time to Screen CV by HM	3	12.1	9.1	7.0	0 / 30.7w
Time to Hire	30	30.2	0.2	20.7	2 / 189
Time to Select	25	50.8	25.8	45.9	2 / 289
Time Interview to Offer	20	41.3	21.3	35.2	2 / 175
Requisition Hit rate	30%				
Interview rate	29.7%				

Process Efficiency: Educational Perspective.

For an analysis of the educational perspective, a comparison between the educational levels of the job functions is made. The three levels are Bachelor, Master, PhD. When comparing the metrics in Table 4, the scorecard shows that the Bachelor-level applications are screened faster on average than their PhD-counterpart and score below average. This could be the result of the requirements for a Bachelor-level function are defined more clearly and are more common than the ones for a PhD-level function. What stands out is that the ‘Time to Select’ and ‘Time Interview to Offer’ are very high for PhD-level functions.

Table 4: Bachelor, Master, and PhD recruitment scorecard.

Metric	Target	BSc	MSc	PhD
Time to Screen CV by Recruiter	5	5	5.3	5.6
Time to Screen CV by HM	3	12.3	12.1	13.8
Time to Hire	30	31.6	28.4	27.2
Time to Select	25	48.3	50.1	85.2
Time Interview to Offer	20	44.1	45.7	74.4
Requisition Hit rate		36.2%	31.5%	19.5%
Interview Rate		27.7%	31.5%	34.2%

Stage 5. Evaluation.

During this stage, the HR professional evaluates the results of the recruitment process mining analysis to determine whether the metrics provide meaningful insights that answer the business questions at hand. The process model and its statistics are then presented to and discussed with relevant business experts and stakeholders. Based on this discussion, it may be determined that the initial questions have not been fully answered, or new follow-up questions have arisen, which may require additional iterations of PM^{2HR}.

Stage 6. HR Process Improvement.

Once all the business questions have been satisfactorily answered, or the desired insights have been obtained, it is possible to take targeted action to improve process performance. In the context of recruitment, this could involve focusing on faster screening times by recruiters, streamlining the process of scheduling interviews with applicants, and so on. An essential aspect of process improvement is conducting additional iterations of analysis. This involves updating the dataset and repeating the analysis to obtain the latest results, which can help evaluate the effectiveness of the actions taken. If the targeted actions have led to progress, then they have been successful in improving process performance. However, if the results are not as expected, additional actions may be necessary to address the identified issues. Frequent iterations of analysis on recent data not only allow for continual improvement of process performance but also help ensure that the organization remains in control. This approach helps to identify potential problems early on and make appropriate adjustments, ultimately leading to more effective and efficient business operations.

In general, the results from the case study were highly valuable for the company's HR department, as they provided crucial day-to-day insights into the operational excellence of the recruitment process. These insights were challenging to obtain and generate before, highlighting the usefulness of process mining.

4 CONCLUSIONS

This paper provides an overview of the PM^{2HR} methodology for HR analytics. This process mining project methodology for HR analytics is a comprehensive approach to multiple HR decision-making problems and it is developed based on the PM² methodology (Van Eck et al., 2015). In the following, we list the sources of added value of the process mining technique in the HR context:

- Process mining allows the HR practitioner to conduct analyses not possible with existing tools, such as the classical data mining and statistical tools.
- Process mining analyses the entire collection of HR event data.
- Process mining allows HR to have a more effective way of measuring process efficiencies and inefficiencies as well as simply providing an overview of the actual HR process.

To evaluate, the PM^{2HR} methodology for HR analytics we applied it on a case study related to recruitment analysis. Overall, the application of the methodology was successful as it produces relevant insights in the recruitment process, with the support of the HR performance scorecard. A key element in the methodology is the establishment of the definition of the HR event log. Another highlight of our research is the guidance we provide to the HR professional in assessing to which HR functions process mining could be used.

To conclude, this paper demonstrates that the use of process mining in the HR domain creates new insights that enable better HR-decision making, particularly in the recruitment and selection process. We demonstrate that the PM^{2HR} methodology for HR analytics is a comprehensive user guide for process mining analyses in the HR domain. We also illustrate that these analyses lead to significant process improvements. However, we believe that there is still need for applying and evaluating the PM^{2HR} methodology for HR analytics in HR process mining projects of other organizations and conducting more case studies.

ACKNOWLEDGEMENTS

We would like to thank W. van der Zanden and M.E. Savvidi for their contribution to this work.

REFERENCES

- Van der Aalst, W.M.P. (2009). Process-Aware Information Systems: Lessons to Be Learned from Process Mining. In: Jensen, K., van der Aalst, W.M.P. (eds) Transactions on Petri Nets and Other Models of Concurrency II. Lecture Notes in Computer Science, vol 5460. Springer, Berlin, Heidelberg.
- van der Aalst, W.M.P. (2012). Process mining. *Communications of the ACM*, 55:76–83.
- van der Aalst, W.M.P. (2016). Process mining: Data science in action. Springer Berlin Heidelberg, 2 edition.
- Arias, M. & Saavedra, R. & Marques Sarmay M., Munoz-Gama, J. & Sepulveda, Marcos. (2018). Human resource allocation in business process management and process mining: A systematic mapping study. *Management Decision*. 56. 10.1108/MD-05-2017-0476.
- Chauhan, A., Sharma, S. K., & Tyagi, T. (2011). Role of HRIS in improving modern HR operations. *Review of Management*, 1(2), 58-70.
- Christensen, C. M., Hall, T., Dillon, K. & Duncan, D. S. (2016). Know Your Customers' "Jobs to Be Done". *Harvard business review*, 94(9), 54–62.
- van Eck, M. L., Lu, X., Leemans, S. J., & van der Aalst, W. M. (2015). PM²: A Process Mining Project Methodology. In *International Conference on Advanced Information Systems Engineering* (pp. 297-313). Springer, Cham.
- Laumer, S., Maier, C., & Eckhardt, A. (2015). The impact of business process management and applicant tracking systems on recruiting process performance: an empirical study. *Journal of Business Economics*, 85(4), 421-453.
- Jooseok, L., Lee, S., Kim, J. & Choi, I. (2018). Dynamic human resource selection for business process exceptions. *Knowledge and Process Management*. 26. 10.1002/kpm.1591.
- Marler, J.H. & Boudreau, J.W. (2017) An evidence-based review of HR Analytics, *The International Journal of Human Resource Management*, 28:1, 3-26, DOI: 10.1080/09585192.2016.1244699.
- Muenstermann, B., Eckhardt, A., & Weitzel, T. (2010). The performance impact of business process standardization: An empirical evaluation of the recruitment process. *Business Process Management Journal*, 16(1), 29–56.
- R'bigui, H. & Cho, C.. (2017). The state-of-the-art of business process mining challenges. *International Journal of Business Process Integration and Management*. 8. 285. 10.1504/IJBPIIM.2017.10009731.