

Editorial: "Business Process Intelligence: Connecting Data and Processes"

WIL VAN DER AALST, Technische Universiteit Eindhoven

J. LEON ZHAO, City University of Hong Kong

HARRY JIANNAN WANG, University of Delaware

This introduction to the special issue on Business Process Intelligence (BPI) discusses the relation between data and processes. The recent attention for Big Data illustrates that organizations are aware of the potential of the torrents of data generated by today's information systems. However, at the same time, organizations are struggling to extract value from this overload of data. Clearly, there is a need for data scientists able to transform event data into actionable information. To do this, it is crucial to take a process perspective. The ultimate goal of BPI is not to improve information systems or the recording of data; instead the focus should be in improving the process. For example, we may want to aim at reducing costs, minimizing response times, and ensuring compliance. This requires a "confrontation" between process models and event data. Recent advances in process mining allow us to automatically learn process models showing the bottlenecks from "raw" event data. Moreover, given a normative model, we can use conformance checking to quantify and understand deviations. Automatically learned models may also be used for prediction and recommendation. BPI is rapidly developing as a field linking data science to business process management. This article aims to provide an overview thereby paving the way for the other contributions in this special issue.

Categories and Subject Descriptors: **H.2.8 [Database Applications]:** Data mining and **H.4.1 [Office Automation]:** Workflow management

General Terms: Design, Management, Performance

Additional Key Words and Phrases: Process mining, Business Process Intelligence, Process Modeling, Performance analysis, Compliance checking

ACM Reference Format:

Wil van der Aalst, J. Leon Zhao, Harry Jiannan Wang 2014. Business Process Intelligence: Connecting Data and Processes. *ACM Transactions on Management Information Systems*. XX, XX, Article XX (Sept 2014), XX pages.

DOI: <http://dx.doi.org/10.1145/0000000.0000000>

1. TOWARDS PROCESS SCIENCE

Our capabilities to store and process data have been increasing exponentially since the 1960s [Hilbert and Lopez 2011, Manyika 2011]. Today, data are collected about anything, at any time, and at any place. Moreover, the emerging "internet of things" and the rapid growth of mobile devices provide new forms of data much closer to our everyday reality (social, physical, and economical). Figure 1 illustrates the different types of data being collected in today's world cumulating into the *Internet of Events* (IoE). As a result, organizations are seeking ways to exploit the information readily available [Van der Aalst 2014a]. The "Big Data" hype illustrates this trend. Moreover, a new discipline has emerged: *Data Science* (DS). Data science aims to collect, analyze, and interpret data from a variety of sources (social interaction, business processes, cyber-physical systems) [Van der Aalst 2014a, Chiang et al. 2012]. To turn data into actionable information, a comprehensive understanding of the context of the data and the ability to mine and visualize large amounts of data are essential. Therefore, the role of the data scientist is:

- to ensure that the right data is recorded and stored (provenance) and to be able to extract relevant data in a complex IT landscape,
- to use a wide variety of analytical techniques to extract value (models, insights, predictions, recommendations, and visualizations) from data, and
- to present the results to end users, assist in the interpretation of results, and be able to put it in context.

Just like computer science emerged from mathematics when computers became abundantly available in the 1980s, we can now see that today's data tsunami is creating the need for data scientists. Although data is enabling new and exciting forms of analysis, one should not forget that in the end we would like to improve operational business processes based on data and not just collect data. Therefore, we relate data science to *Business Process Management* (BPM) [Van der Aalst 2013a].

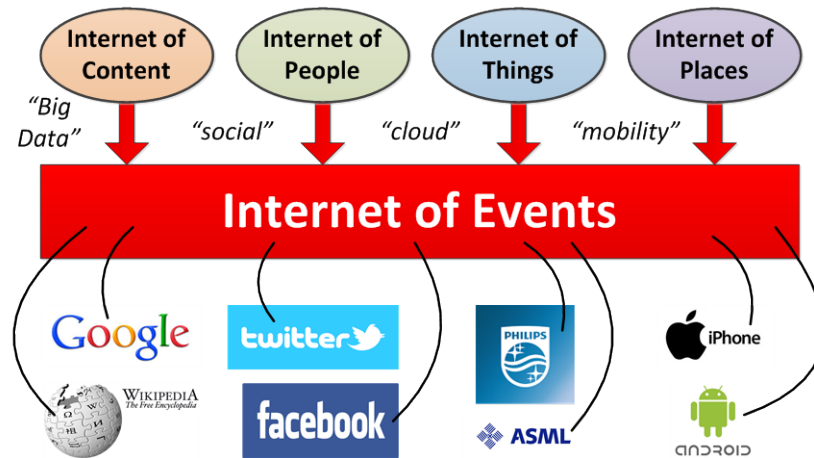


Fig 1. The “Internet of Events” is composed of the “Internet of Content”, the “Internet of People”, the “Internet of Things”, the “Internet of Places” [Van der Aalst 2014a].

BPM is the discipline that combines knowledge from information technology and knowledge from management sciences and applies this to operational business processes [Van der Aalst 2013a, Dumas et al. 2013, Reichert and Weber 2012, Weske 2007, Stohr and Zhao 2001]. BPM received considerable attention in recent years due to its potential for significantly increasing productivity and saving costs. Moreover, today there is an abundance of BPM systems. These systems are realized using generic software products that are driven by explicit process designs to enact and manage operational business processes. However, BPM is definitely not limited to BPM systems. BPM has a broader scope: from process automation and process analysis to operations management and the organization of work [Van der Aalst 2013a, Zhao, Kumar, and Stohr 2000]. On the one hand, BPM aims to improve operational business processes, possibly without the use of new technologies. For example, by modeling a business process and analyzing it using simulation, management may get ideas on how to reduce costs while improving service levels. On the other hand, BPM is often associated with software to manage, control, and support operational processes. Traditional BPM approaches do not exploit event data in a systematic way: the focus of BPM practitioners is often on modeling processes and measuring Key Performance Indicators (KPIs). However, it is obvious that the BPM discipline will become more data-driven.

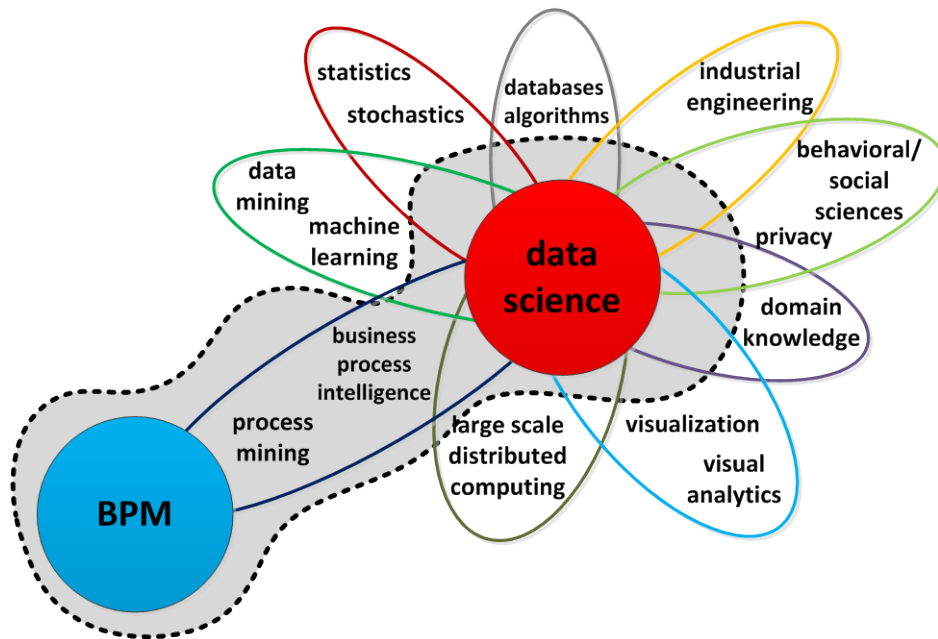


Fig. 2. Overview of the different skills relevant for today’s data scientist [Van der Aalst 2014b] and the relation to Business Process Management (BPM). The highlighted area demarcates the focus of this special issue.

Figure 2 shows the focus of this special issue. Next to the typical data science skills, the connection to BPM is emphasized in the figure. Next to data scientists, there is an urgent need for *process scientists*, who are analysts that combine BPM skills with data science skills to improve processes by exploiting event data.

2. BUSINESS PROCESS INTELLIGENCE

Business Process Intelligence (BPI) is on the interface between data science and BPM. BPI is closely related to topics such as *Business Intelligence* (BI) and *Process Mining* (PM).

Boris Evelson of Forrester Research defines BI as “a set of methodologies, processes, architectures, and technologies that transform raw data into meaningful and useful information used to enable more effective strategic, tactical, and operational insights and decision making”. Although this definition does not exclude a process focus, BI methodologies and tools are typically not process-aware.

Process mining aims to discover, monitor and improve real processes by extracting knowledge from event logs readily available in today’s information systems as well as from corporate policy manuals [Van der Aalst 2011, Li et al. 2010, Wang et al. 2009]. The starting point for process mining is an *event log*. Each event in such a log refers to an activity (i.e., a well-defined step in some process) and is related to a particular case (i.e., a process instance). The events belonging to a case are ordered and can be seen as one “run” of the process. Event logs may store additional information about events. In fact, whenever possible, process mining techniques use extra information such as the resource (i.e., person or device) executing or initiating the activity, the timestamp of the event, or data elements recorded with the event (e.g., the size of an order).

Event logs can be used to conduct three basic types of process mining [Van der Aalst 2011]. The first type of process mining is *discovery*. A discovery technique takes an event log and produces a model without using any a-priori information. Process discovery is the most prominent process mining technique. For many organizations it is surprising to see that existing techniques are indeed able to discover real processes merely based on example behaviors stored in event logs. The second type of process mining is *conformance*. Here, an existing process model is compared with an event log of the same process. Conformance checking can be used to check if reality, as recorded in the log, conforms to the model and vice versa. The third type of process mining is *enhancement*. Here, the idea is to extend or improve an existing process model thereby using information about the actual process recorded in some event log. Whereas conformance checking measures the alignment between model and reality, this third type of process mining aims at changing or extending the a-priori model. For instance, by using timestamps in the event log one can extend the model to show bottlenecks, service levels, and throughput times.

Orthogonal to three types of process mining (process discovery, conformance checking, and enhancement), one can look at the *offline* setting (i.e., historic data on completed cases) or the *online* setting (i.e., also consider event data on still running cases that can still be affected by corrective actions). For example, discovered and subsequently enhanced process models can be used for *operational support* and thus enable *predictive analytics*, e.g., predicting the remaining flow time of a case, recommending a suitable activity or resource, etc.

Another application of the three basic types of process mining is *comparative* process mining, which uses process cubes [Van der Aalst 2013b] to compare different subprocesses or groups of cases. The focus is no longer on a single process and a homogenous set of cases. Comparative process mining can be used to identify differences between two branches of the same organization or two groups of customers.

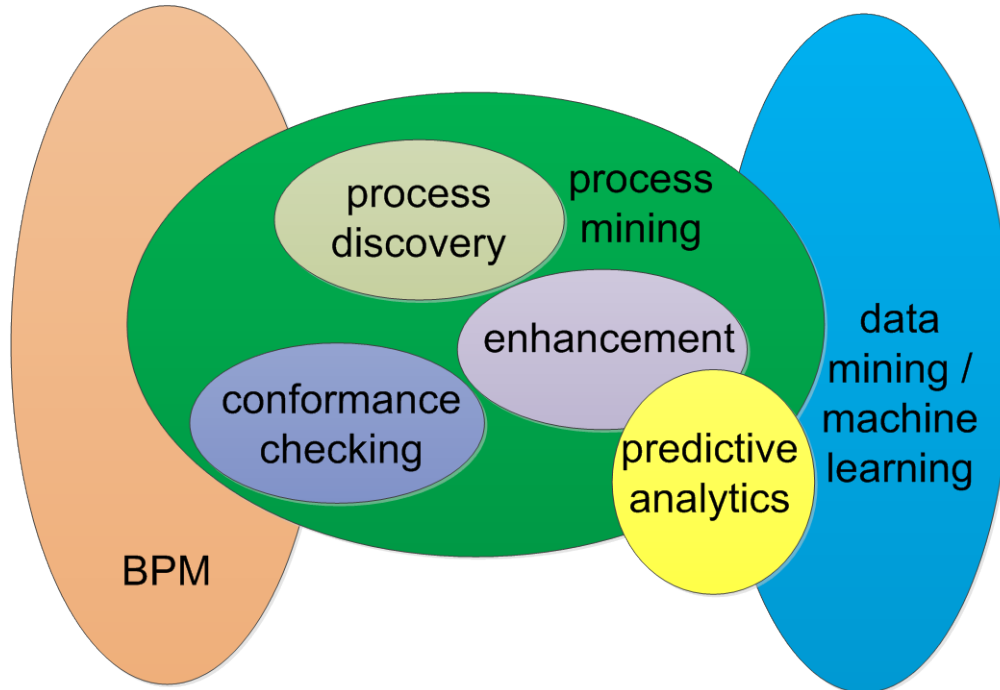


Fig. 3. Positioning process mining as the linking pin between BPM and data-driven approaches like data mining.

Figure 3 positions process mining as a bridge between process-centric approaches and more data-centric approaches like data mining and machine learning. It is difficult to distinguish between process mining and BPI. BPI could be defined as BI with particular attention for processes. However, this is exactly what process mining is about. Therefore, we use the terms process mining and BPI interchangeably.

Data science, consisting of data mining and machine learning, is concerned with four main questions [Van der Aalst 2014a]:

- Reporting: *What happened?*
- Diagnosis: *Why did it happen?*
- Prediction: *What will happen?*
- Recommendation: *What is the best that can happen?*

The first two questions aim at understanding the past. The last two questions use knowledge learned from past experiences to say something about the future. Obviously, all four types of questions can be posed in the context of process improvement. Typical BPI questions are:

- What is the process that people in an organization really follow?
- Where are the bottlenecks in my business process and how to remove them?
- Where do organizations deviate from the expected or idealized process?
- What are the “highways” in my process?
- What factors are influencing the service level provide to customers?
- Can we predict problems (delay, deviation, risk, etc.) for running cases?
- Can we recommend countermeasures?
- How to redesign the process given a set of goals?
- Are non-compliant cases more costly?

Process mining/BPI software aims to answer the above questions. Mainstream BI software like IBM Cognos Business Intelligence (IBM), Oracle Business Intelligence (Oracle), SAP BusinessObjects (SAP), WebFOCUS (Information Builders), MS SQL Server (Microsoft), MicroStrategy (MicroStrategy), NovaView (Panorama Software), QlikView (QlikTech), SAS Enterprise Business Intelligence (SAS), TIBCO Spotfire Analytics (TIBCO), Jaspersoft (Jaspersoft), and Pentaho BI Suite (Pentaho) are *not* process oriented and therefore less suitable for answering the above questions. Process mining tools like ProM (open source), Disco (Fluxicon), Perceptive Process Mining (Perceptive Software), ARIS Process Performance Manager (Software AG), QPR ProcessAnalyzer (QPR), Celonis Discovery (Celonis), Interstage Process Discovery (Fujitsu), Discovery Analyst (StereoLOGIC), and XMAalyzer (XMPro) aim at answering at least a subset of the above questions. Existing tools provide widely applicable and very robust types of analysis. However, Process mining/BPI is a relatively young discipline with many open questions.

3. IN THIS SPECIAL ISSUE ON BPI

There is a growing interest in BPI. This is not surprising as it is clear that management information systems need to become more process-centric while utilizing the event data at hand. This special issue presents state-of-the-art results in BPI. Out of twenty three submissions, eight papers were accepted after an initial screening and multiple rounds of reviews and revisions. A brief overview of these papers is given next.

Two papers are related to the process discovery aspect of BPI. In “On the Discovery of Declarative Control Flows for Artful Processes,” Di Ciccio and Mecella focus on the study of artful processes, where decisions are based on the experience, intuition, and knowledge of the process actors. Compared with operational business processes, artful processes are much less structured, very flexible, and even completely unknown a priori. The authors propose an algorithm to automatically discover artful processes, which is critical for better understanding knowledge-intensive artful processes and providing improvements in many scenarios. Wang et al. present a unique application of process mining to knowledge management for online Q&A communities in “An Analytical Framework for Understanding Knowledge Sharing Processes in Online Q&A Communities”. They innovatively assign dialog act tags to posts in discussion threads and treat each thread as a process instance of dialog act tags. Then, different process patterns are discovered and analyzed to identify patterns that are more likely to lead to helpful knowledge sharing.

Conformance checking is another important part of BPI. In “Compliance Checking of Organizational Interactions,” Jiang et al. look into the problem of regulatory process compliance checking. They propose a normative structure named Norm Nets (NNs) to formally model narrative process regulations and develop mechanisms to automatically check process compliance by mapping NNs to Colored Petri Nets. Their approach enables the systematic analysis of process compliance with dynamic interrelated regulations and business partner interactions.

The next two papers investigate the intelligent task-agent assignment problem in BPI. In “Process Analytics Approach for R&D Project Selection,” Silva, Ma, and Yang, propose a research analytics framework to facilitate the reviewer assignment task in the scientific proposal selection process. Text mining and clustering techniques are

used to cluster proposals and social network analysis and bibliometric analysis are applied to generate reviewers' network with no conflict of interests and their research qualification scores. Based on that, a reviewer assignment function is developed to optimize the research relevance between proposals and reviewers. In "Mining Agents' Goals in Agent-Oriented Business Processes," Yan et al. propose a framework to mining process agents' goals from process event logs. Belief-Desire-Intention (BDI) logic is used to formally represent agents' goals and a goal-mining algorithm is developed. This framework can be used to detect and analyze inconsistencies between the design and execution of business processes in terms of agents' goal and overall process goal, which facilitate process improvements and redesign.

The data perspective of BPI is discussed in the following two papers. In "Classification Models for RFID-based Real-Time Detection of Process Events in the Supply Chain: An Empirical Study," Keller, Thiesse, and Fleisch aim to address the noisy process data problem in RFID-enabled supply chains. A number of data mining techniques have been applied to filter and aggregate RFID raw data and a customized decision-rule-based classifier is developed. The method proposed in this paper can greatly improve the process data quality and thus pave the way for better real-time process monitoring and optimization. Business process integration has been an important issue for B2B collaboration. In "A Data Flow Perspective for Business Process Integration," Guo, Sun, and Vogel propose a data flow formalism to facilitate inter-organizational process integration, which provides a set of constructs to model inter-organizational data exchange and a method to calculate public data set to facilitate collaboration.

Lastly, Partington et al. present a comprehensive case study of applying process mining to the administrative and clinical data from four healthcare providers in "Process Mining for Clinical Processes: A Comparative Analysis of Four Australian Hospitals". This case study provides many useful insights into conducting practical process mining projects.

The guest editors would like to thank all reviewers who provided timely and constructive reviews for this special issue; without their dedication and expertise, this special issue would not be possible. We also want to thank Professor Hsinchun Chen for his kind guidance and strong support in the process of developing this special issue. Last but not least, we extend our thanks to Cathy Larson for her help in every step of the submission and review process.

REFERENCES

- W.M.P. van der Aalst, 2011. *Process Mining: Discovery, Conformance and Enhancement of Business Processes*. Springer-Verlag, Berlin.
- W.M.P. van der Aalst, 2013a. *Business Process Management: A Comprehensive Survey*. ISRN Software Engineering, pages 1-37. doi:10.1155/2013/507984.
- W.M.P. van der Aalst, 2013b. *Process Cubes: Slicing, Dicing, Rolling Up and Drilling Down Event Data for Process Mining*. In M. Song, M. Wynn, and J. Liu, editors, *Asia Pacific Conference on Business Process Management (AP-BPM 2013)*, volume 159 of *Lecture Notes in Business Information Processing*, pages 1-22. Springer-Verlag, Berlin.
- W.M.P. van der Aalst, 2014a. *Data Scientist: The Engineer of the Future*. In K. Mertins, F. Benaben, R. Poler, and J. Bourrieres, editors, *Proceedings of the I-ESA Conference*, volume 7 of *Enterprise Interoperability*, pages 13-28. Springer-Verlag, Berlin.
- W.M.P. van der Aalst, 2014b. *Process Mining: Data Science in Action*. Coursera course, <https://www.coursera.org/course/procmin>.
- R. H. L. Chiang, P. Goes, and E. A. Stohr, 2012. *Business Intelligence and Analytics Education*, and

- Program Development: A Unique Opportunity for the Information Systems Discipline, *ACM Transactions on Management Information Systems*, vol. 3, no. 3, pp. 12:1–12:13.
- M. Dumas, M. La Rosa, J. Mendling, and H. Reijers, 2013. *Fundamentals of Business Process Management*. Springer-Verlag, Berlin.
- M. Hilbert and P. Lopez, 2011. The World's Technological Capacity to Store, Communicate, and Compute Information. *Science*, 332(6025):60-65.
- J. Li, H.J. Wang, Z. Zhang, J. L. Zhao, 2010. A policy-based process mining framework: mining business policy texts for discovering process models, *Information Systems and e-Business Management*, Volume 8, Number 2 / March, Pages 169-188.
- J. Manyika, M. Chui, B. Brown, J. Bughin, R. Dobbs, C. Roxburgh, and A. Byers, 2012. *Big Data: The Next Frontier for Innovation, Competition, and Productivity*. McKinsey Global Institute, 2011.
- M. Reichert and B. Weber, 2012. *Enabling Flexibility in Process-Aware Information Systems: Challenges, Methods, Technologies*. Springer-Verlag, Berlin.
- E.A. Stohr and J. L. Zhao, 2001. Workflow Automation: Overview and Research Issues, *Information Systems Frontiers*, Volume 3, Issue 3, pp. 281-96.
- H.J. Wang, J.L. Zhao, L.J. Zhang, 2009. Policy-Driven Process Mapping (PDPM): Discovering process models from business policies, *Decision Support Systems*, 48 (1), pp. 267-281.
- M. Weske, 2007. *Business Process Management: Concepts, Languages, Architectures*. Springer-Verlag, Berlin.
- J. L. Zhao, A. Kumar, and E. A. Stohr, 2000. Workflow-centric Information Distribution through Email, *Journal of Management Information Systems*, Volume 17, No. 3, pp. 45-72.