Integrating Data for Business Process Management

Hong-Linh Truong and Schahram Dustdar
Distributed Systems Group, Vienna University of Technology
Email:{truong,dustdar}@infosys.tuwien.ac.at

Abstract

To be able to utilize Web-scale resources for business processes and to adapt these processes to the dynamic change of environments, business process management suites/systems (BPMS) must be able to gather, integrate and manage various types of data in the lifecycle of business processes. We discuss the issue of integrating data for business process management. We provide an overview on the current state of how data is integrated into business process management and recommend new directions.

1 Introduction

For an organization, its business processes are the key to its success. Therefore, it is of paramount importance for the organization to have a powerful business process management (BPM) approach that enables it to rapidly create new, capable business processes and to improve and adapt existing business processes to the changing environment in which the processes are executed. The term "BPM" is used loosely here and we mean that BPM includes all activities that help the organization to achieve such capable and adaptable business processes. Over the last a few years, we have observed that, instead of relying only on a single organization's resources (software services and humans) to perform process activities, business processes in an organization have increasingly relied on Web-scale resources by utilizing and assembling multiple-organizational or individual resources. This paradigm shift has been supported by emerging technologies, such as the SOA and Software-as-a-Service (SaaS) model and the integration of humans into business processes (e.g., BPEL4People/WS-HumanTask[1]). Together with more complex business requirements, this paradigm shift makes business processes more complex and difficult to manage and understand. This change has a profound impact on technologies used in BPM, affecting all phases of the lifecycle of business processes, including process design, modeling, execution, monitoring, and optimization.

To be able to utilize the Web-scale resources for business processes and to adapt these processes to the dynamic change of environments, BPM suites/systems (BPMS) must be able to gather, integrate, and manage various types of data in order to efficiently manage processes. It means that many types of data should be integrated and associated through the lifecycle of business processes. To date, those types of data that characterize the main five themes, named process strategy, process architecture, process ownership, process measurement, and process improvement [11], of a business process are voluminous, complex and difficult to collect and manage. In particular, by utilizing SaaS and user-generated services and by allowing mass customization, BPMS

Bulletin of the IEEE Computer Society Technical Committee on Data Engineering

[1] `http://xml.coverpages.org/bpel4people.html`

should be able to closely interact with SOC management frameworks and utilize data provided by these frameworks. However, it is very challenging to integrate BPMS and SOC management frameworks together. This paper examines which types of data are needed, why they are important, and how they are currently integrated and utilized for business processes (Section 2). We also suggest to develop a unified model and techniques for linking and managing integrated data for BPM during the evolution of business processes (Section 3).

2 Data Associated with Business Processes

In order to examine types of data that should be integrated for BPM, let us consider data required in the lifecycle of a business process that has been studied in literature. Table 1 provides further description of these types.

Type	**Phases**	**Description**	**Example**
Strategy	Design, Modeling	including data about business strategy and IT strategy, expected SLA and KPIs for the process	"by 2010, more than 50% activities should be performed by third-party Web services", "utilize only shipping services offering the shipping time less than 3 days"
Capability	All	including capabilities associated with a service, a human or a human's offerings, as well as concerns describing when and how the service or human can be used.	"providing company credits", "the average response time to credit requests is 2 days", "a service provided by a media freelancer in Vienna", "an expert in BPEL design"
Contract	All	including software service contracts, contracts for business artifacts/data, business compliance rules	"pay-per-use with service credit", "the law enforcement is the European court", "the data is free, but owned by the provider"
Patterns and processes	All	including common patterns, discovered process models and patterns, existing processes relevant to the process.	"delegation pattern", "one-to-many service interaction pattern", "3 similar processes in the repository"
Business rules	Execution	including rules specifying business constraints and compliance policies used to execute the process	"a failed service is replaced by only similar services inside Europe", "when the order process delays 2 days, send a notification"
Performance	Monitoring, Optimization, Design	including IT performance and business performance metrics. Furthermore, historical performance and monitoring data for offline optimization and refinement, QoS metrics of services and humans	"the availability in the last 10 days is 70%", "today's number of failures is 3" , "the average response time in the last 100 calls is 4 days" , "the number of successful completed activities in the last 10 days is 70"

Table 1: Types of possible data relevant to a business process

Through the lifecycle of a business process, various types of data are needed for different purposes. All these types of information are important for deciding the techniques used in, for example, structure and behavior design, runtime execution and monitoring, and off-line and runtime optimization. Our first observation is that a majority of BPMS support only certain types of data associated with capability and performance. They allow the user to design services and humans in business processes, and monitor and optimize the processes, e.g., [8]. Our second observation is that there is a track record on advanced process patterns and model analysis, such as [16, 1, 15]. Furthermore, many performance data is collected, such as [17]. However, many tools and BPMS cover only a small part of these types of data. There is a lack of frameworks and techniques to support BPMS to harness multi-organizational and individual resources for business processes. We believe that the following points should be addressed in BPMS:

- integrating and managing data about resources capability and availability in the Web-scale
- ensuring contract compliance for business processes consisting of services and humans in the Web-scale

- managing and integrating reusable patterns and processes, and performance data

In our view, integrating these types of data into and providing them as an internal feature of BPMS are extremely important as they help to solve challenges raised by the gap between the business level and IT level as well as to adapt business processes to changing environments. In the following, we discuss these main points.

2.1 Integrating Capability and Availability Data in the Web-scale

Two main types of resources for a business process are software services and humans. Data describing the capability and availability of services and humans is critical for all phases of a business process. It can substantially improve the design and adaptation of business processes. For example, the design of a business process could start from scratch if we are not aware of existing resources that can be assembled. Capability data is virtually required in all lifecycle phases, but most BPMS use the capability data only for the design phase, some for the optimization phase at runtime. While data about few services and humans might be enough for the design (to prove that the functionality is working), rich data about services and humans would increase the possibility to analyze what-if scenarios in the process modeling and to adapt processes to situations at runtime. Unfortunately, managing Web-scale resources for business processes is still at an early stage. Most BPMS just assume that the designer knows where the resources are. But this assumption is hard, if not impossible, to be hold when business processes are relied on Web-scale resources.

BPMS used by an organization face many challenges when integrating resources outside the organization, in particular, commodity software services provided as SaaS and humans acting as external services. Many BPMS have already supported the design and execution of business processes whose activities are performed by services, in particular, Web services, but do not offer mechanisms to search and find relevant services. Furthermore, this search will not be limited to functional aspect of services (e.g., account management or payment) but also other concerns, such as licensing, location, and trust. With respect to the role of humans in business processes, humans can be actors who design the process as well as who perform activities in the process. In the first aspect, managing a person's capabilities, skill and team as well as his/her participation in the business process design could potentially help to improve the design of business processes by quickly locating the right person for the right task. This aspect requires BPMS to be integrated with social and team networks which is lacking in most BPMS. In the second aspect, some techniques, such as BPEL4People, have enabled the integration of humans in the Web as a part of business processes. They are, however, very premature. For example, they allow to specify human and software activities but neglect the discovery of human resources. In most cases, the user has to enter the information about human services. Harnessing mass user-generated services as one way of outsourcing, e.g., empowered by freelancers, is currently not in the focus of BPMS. To overcome these problems, it is necessary to integrate BPMS with service discovery and registry capabilities and social networks of humans. From the management point of view, solutions based on cloud computing, such as Platform-as-a-Service (PaaS) for BPMS, could also address these issues. PaaS providers could be responsible for managing resource capabilities, while an organization just focuses on utilizing these resources (e.g., the Boomi platform[2] for software services).

2.2 Integrating Service- and Data-Related Contract Concerns

The design and execution of business processes have to ensure that resources used and artifacts manipulated and produced will comply to certain contracts. Currently, the evaluation of contract concerns associated services and artifacts in business processes is focused only on a small number of concerns, notably service-related QoS/SLA metrics, e.g., [5], and mostly at the design time. However, resources and artifacts are bound to many other concerns, such as quality of data, intellectual property rights, law enforcement, data distribution, data disposition, to name just a few. These concerns are typically associated with DaaS (data-as-a-service), such as StrikeIron[3]

[2] `www.boomi.com`

[3] `http://www.strikeiron.com/strikeironservices.aspx`

and Amazon Web services[4], which are utilized by organizations to retrieve and store business artifacts. These concerns are important for both service and data aspects compliance in business processes.

With current composition and data mapping techniques, BPMS allow the designer to functionally compose data sources and services in an easy manner. However, the designer lacks a mechanism to validate whether these data sources and services being complied with the above-mentioned expected concerns. Partially, it is due to the fact that services and data are not well described, but also this research topic is not in the focus of existing BPMS. To overcome these issues, we should enrich current QoS/SLA metrics and techniques with quality of data, service and data licensing, and data governance metrics. Furthermore, compliance evaluation techniques for these concerns should be integrated into all phases of BPM.

2.3 Integrating Reusable Patterns and Processes, and Performance Data

Several research approaches have been carried out for understanding processes and patterns. But this kind of data is not well integrated into existing BPMS, if we consider how BPMS can utilize these data to recommend the process design, modeling and optimization in a (semi)automatic manner. Furthermore, currently BPMS lack a connection to existing business processes that might be reused. This is not only due to a small handful of research efforts on mining process repositories [7] but also due to the lack of shared process repositories[5]. To support the (automatic) search and reuse of patterns and process models, it is expected to have a service-based repository for sharing business processes, either in the Web-scale or the individual organization level. Existing work has demonstrated the usefulness of documented best practices, detected patterns, and mining results, such as patterns used for design and modeling [6] and mining results used for recommending processes [10]. When we are able to manage reusable patterns and processes, then these works can be combined with other techniques, such as similarity analysis of processes [2], to provide powerful mechanisms to the design of new processes.

With respect to performance data, currently most BPMS support only a few metrics of IT performance, such as failure, availability, and response time, collected from the monitoring of the execution of processes. Some support the optimization of the process at runtime based on these performance metrics. However, historical performance metrics are not well integrated into BPMS for supporting the design, modeling, and runtime adaptation. To date, many performance analysis works have been done but we lack a standard way to link and manage performance data throughout the lifecycle of business processes. We should consider mining, process analysis, and performance analysis results to be associated with different levels of abstraction of business processes, such as individual activities and workflow regions, to provide a unified view on the performance in order to support the process refinement and optimization at different levels. In addition, as the business performance is measured through KPIs, it is interesting to establish the correlation between IT performance and business performance metrics; this is not well researched and understood. The performance data is also strongly linked to patterns and process models, and process repositories, and thus they should be managed and provided together.

3 Unified Data Management for BPM

Given a requirement, a business process is designed, modeled, executed, monitored and optimized. Although various types of data related to the process alone might be provided by different tools, the current situation is that we lack a mechanism to link all kinds of data inherently in the lifecycle of business processes. From our analysis of integrating data for BPM, we propose two main points for a unified data management system that should be integrated into BPMS:

- a unified, scalable and flexible model for integrating diverse types of data required by BPM.

[4] `http://aws.amazon.com/`

[5] For example, the Process Wiki (`http://wiki.process.io`) is a place where we can find a few business processes, while in the myExperiment (`http://www.myexperiment.org`) hundreds of scientific workflows are shared.

- techniques for managing the integrated data for business processes during their evolution.

Stimulated by our work in Web services evolution management [13], we devise a conceptual model of integrated data for BPM in Figure 1. In this model, we have different types of data, described under different specifications and linked through a meta-model. The instance data belonging to each type will be linked as an external source, such as modeling process description, process execution description, performance data, documented best practices, detected patterns, service capability registry, and human capability registry, thus allowing different specifications and diverse types of data to be included. This meta-model can be built based on XML in which a type of data is represented by a concept describing the type of data, the schema location, and the source of instance data. The collected data is then managed over the time by incorporating temporal aspects into the acquisition, management, and retrieval of the data. Furthermore, social aspects, such as teams and social networks, can be associated with particular types of data which are understood, analyzed, created and manipulated by a team of people. Currently, we are focusing on integrating modeling models [12], performance data and detected patterns [14, 4], software and human service registry [13], and human-provided services [9] with a focus on self-adaptive design, execution and optimization of SOA-based business processes.

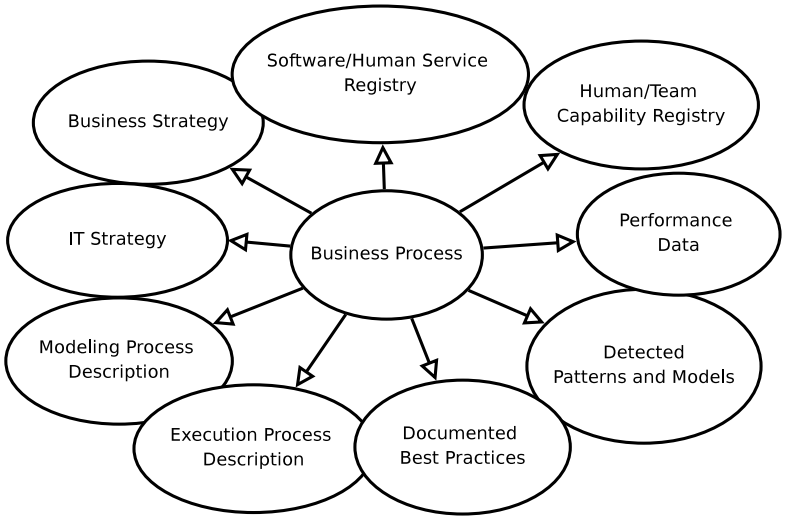

Figure 1: Unified model for integrating different types of data associated with business processes

4 Conclusion

In this paper we outline the current state of integrating data for business process management (BPM). As we have identified, since business processes increasingly rely on Web-scale resources, such as software services deployed under SaaS and human-provided services, there will be a need to integrate many types of data, to analyze and correlate these types of data, and to make them available in all phases of the lifecycle of business processes, in order to support the efficient design, adaptation and self-management of business processes.

References

[1] Alistair P. Barros, Marlon Dumas, and Arthur H. M. ter Hofstede. Service interaction patterns. In Wil M. P. van der Aalst, Boualem Benatallah, Fabio Casati, and Francisco Curbera, editors, *Business Process Management*, volume 3649, pages 302–318, 2005.

[2] Remco M. Dijkman. Diagnosing differences between business process models. In Dumas et al. [3], pages 261–277.

[3] Marlon Dumas, Manfred Reichert, and Ming-Chien Shan, editors. *Business Process Management, 6th International Conference, BPM 2008, Milan, Italy, September 2-4, 2008. Proceedings*, volume 5240 of *Lecture Notes in Computer Science*. Springer, 2008.

[4] Schahram Dustdar, Thomas Hoffmann, and Wil M. P. van der Aalst. Mining of ad-hoc business processes with teamlog. *Data Knowl. Eng.*, 55(2):129–158, 2005.

[5] Genady Grabarnik, Heiko Ludwig, and Larisa Shwartz:. Management of service process qos in a service provider - service supplier environment. In *E-Commerce Technology and the 4th IEEE International Conference on Enterprise Computing, E-Commerce, and E-Services, 2007. CEC/EEE 2007. The 9th IEEE International Conference on*, pages 543–550, July 2007.

[6] Thomas Gschwind, Jana Koehler, and Janette Wong. Applying patterns during business process modeling. In Dumas et al. [3], pages 4–19.

[7] Zhilei Ma1, Branimir Wetzstein1, Darko Anicic, Stijn Heymans, and Frank Leymann. Semantic business process repository. *Proceedings of the Workshop on Semantic Business Process and Product Lifecycle Management (SBPM 2007)*, 251:92–100.

[8] Carlos Pedrinaci, Dave Lambert, Branimir Wetzstein, Tammo van Lessen, Luchesar Cekov, and Marin Dimitrov. Sentinel: a semantic business process monitoring tool. In *OBI '08: Proceedings of the first international workshop on Ontology-supported business intelligence*, pages 1–12, New York, NY, USA, 2008. ACM.

[9] Daniel Schall, Hong Linh Truong, and Schahram Dustdar. Unifying human and software services in web-scale collaborations. *IEEE Internet Computing*, 12(3):62–68, 2008.

[10] Helen Schonenberg, Barbara Weber, Boudewijn F. van Dongen, and Wil M. P. van der Aalst. Supporting flexible processes through recommendations based on history. In Dumas et al. [3], pages 51–66.

[11] P. A. Smart, H. Maddern, , and R. S. Maull. Understanding business process management: Implications for theory and practice. *British Journal of Management*, 2008.

[12] Huy Tran, Uwe Zdun, and Schahram Dustdar. View-based reverse engineering approach for enhancing model interoperability and reusability in process-driven soas. In Hong Mei, editor, *ICSR*, volume 5030 of *Lecture Notes in Computer Science*, pages 233–244. Springer, 2008.

[13] Martin Treiber, Hong Linh Truong, and Schahram Dustdar. Semf - service evolution management framework. In *SEAA*, pages 329–336. IEEE, 2008.

[14] Hong Linh Truong and Schahram Dustdar. Online interaction analysis framework for ad-hoc collaborative processes in soa-based environments. *T. Petri Nets and Other Models of Concurrency*, 2:260–277, 2009.

[15] Wil van der Aalst, Ton Weijters, and Laura Maruster. Workflow mining: Discovering process models from event logs. *IEEE Transactions on Knowledge and Data Engineering*, 16(9):1128–1142, 2004.

[16] Wil M. P. van der Aalst, Arthur H. M. ter Hofstede, Bartek Kiepuszewski, and Alistair P. Barros. Workflow patterns. *Distributed and Parallel Databases*, 14(1):5–51, 2003.

[17] Ping Zhang and Nicoleta Serban. Discovery, visualization and performance analysis of enterprise workflow. *Comput. Stat. Data Anal.*, 51(5):2670–2687, 2007.