Towards Configurable Data Collection for Sustainable Supply Chain Communication

Gregor Grambow, Nicolas Mundbrod, Vivian Steller and Manfred Reichert

Institute of Databases and Information Systems
Ulm University, Germany
`{gregor.grambow,nicolas.mundbrod,vivian.steller,manfred.reichert}@`
`uni-ulm.de`
`http://www.uni-ulm.de/dbis`

Abstract. These days, companies in the automotive and electronics sector are forced by legal regulations and customer needs to collect a myriad of different indicators regarding sustainability of their products. However, in today's supply chains, these products are often the result of the collaboration of a large number of companies. Thus, these companies have to apply complex, cross-organizational, and potentially long-running data collection processes to gather their sustainability data. Comprising a great number of manual and automated tasks for different partners, these processes imply great variability. To support such complex data collection, we have designed a lightweight, automated approach for contextual process configuration.

Key words: Process Configuration, Business Process Variability, Data Collection, Sustainability, Supply Chain

1 Introduction

In todays' industry many products are the result of the collaboration of various companies working together in complex supply chains. Cross-organizational communication in such areas can be quite challenging due to the fact that different companies have different information systems, data formats, and approaches to such communication. These days, state authorities, customers and the public opinion demand sustainability compliance from companies, especially in the electronics and automotive sector. Therefore, companies have to report certain sustainability indicators as, e.g., their greenhouse gas (GHG) emissions or the amount of lead contained in their products. Such reports usually also involve data from suppliers of the reporting company. Therefore, companies launch a sustainability data collection process along their supply chain. This often involves also the suppliers of the suppliers and so on.

As sustainability data collection is a relatively new and complicated issue, service providers (e.g., for data validation or lab tests) are also involved in such data collection. A property that makes these data collection processes even more complex and problematic is the heterogeneity in the supply chain: companies use

different information systems, data formats, and overall approaches to sustainability data collection. Many of them even do not have any information system or approach in place for this and answer with low quality data or not at all. Therefore, no federated system or database could be applied to cope with such problems and each request involves an often long-running, manual, and error-prone data collection process. The following simplified scenario illustrates issues with the data collection process in a small scale.

> *Scenario: Sustainability Data Collection*
> An automotive company wants to collect sustainability data relating to the quantity of lead contained in a specific part. This concerns two of the companies suppliers. One of them has an IHS in place, the other has no system and no dedicated responsible for sustainability. For the smaller company, a service provider is needed to validate the manually collected data to ensure that it complies with legal regulations. The IHS of the other company has its own data format that has to be explicitly converted to be useable. This simple scenario already shows how much complexity can be involved even in simple requests and gives an outlook on how this can look like in bigger scenarios involving hundreds or thousands of companies with different systems and properties.

In the SustainHub[1] project, we develop a centralized information exchange platform that supports sustainability data collection along the whole supply chain. We have already thoroughly investigated the properties of such data collection in the automotive and electronics sectors and published a paper about challenges and state-of-the-art regarding this topic [1]. With this paper, we propose an approach that enables an inter-organizational data collection process. The main point thereby is the capability of this process to automatically configure itself in alignment with the context of its concrete execution.

To guarantee the utility of our approach as well as its general applicability, we have started with collecting problems and requirements directly from the industry. This involved telephone interviews with representatives of 15 European companies from the automotive and electronics sectors, a survey with 124 valid responses from companies of these sectors, and continuous communication with a smaller focus group to gather more precise information. Among the most valuable information gathered there was a set of core challenges for such a system: as most coordination for sustainability data exchange between companies is done manually, it can be problematic to find the right companies, departments, and persons to get data from and also to determine, in which cases service providers must be involved (DCC1). Moreover, this is aggravated by the different systems and approaches different companies apply. Even if the right entity or person has been selected, it might still be difficult to access the data and to get it in a usable format (DCC2). Furthermore, the data requests rely on a myriad of contextual factors that are only manage implicitly (DCC3). Thus, a request is

---

not reusable because an arbitrary number of variants can exist for it (DCC4). A system aiming at supporting such data collection must explicitly manage and store the requests, their variants, all related context data, and also data about the different companies and support manual and automated data collection.

The remainder of this paper is organized as follows: Section 2 shows our general approach for data collection with processes. Section 3 extends this with additional features regarding context and variability. This is followed by a brief discussion of related work in Section 4 and the conclusion.

2 Data Collection Governed by Processes

The basic idea behind our approach for supporting data collection in complex environments is governing the whole procedure by explicitly specified processes. Furthermore, these processes are also automatically enacted by a PAIS (Process-Aware Information System) that is integrated into the SustainHub platform. That way, the process of data collection for a specific issue as a sustainability indicator can be explicitly specified by a process type while process instances derived from that type govern concrete data collections regarding that issue. Activities in such a process represent the manual and automatic tasks to be executed as part of the data collection by different companies. This approach already covers a number of the elicited requirements. It enables a centralized and consistent request handling (cf. DCC1) and also supports manual as well as automated data collection (cf. DCC2). One big advantage lies in the modularity of the realization as process. If a new external system shall be integrated, a new activity component can be developed while the overall data collection process does not need to be adapted. Finally, it also enables the explicit specification of the data collection process (cf. DCC4). By visual modeling the creation and maintenance of such processes is facilitated. However, the realization via processes can only be the basis for comprehensive and consistent data collection support. To be able to satisfy the requirements regarding contextual influences, various types of important data, and data request variants, we propose an extended process-based approach for data collection illustrated in Figure 1.
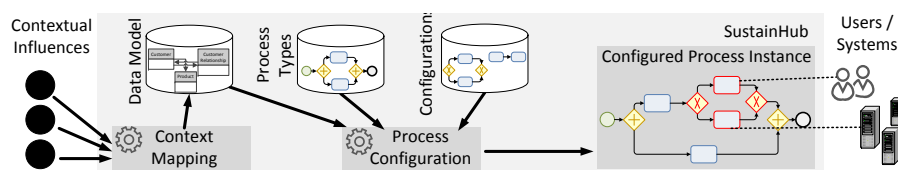

Fig. 1: SustainHub Configurable Data Collection Approach

To generate an awareness of contextual influences (e.g. the concrete approach to data collection in a company, cf. DCC3) and make them usable for the data

collection process, we have defined an explicit context mapping approach (discussed in Section 3.1). This data is necessary for the central step of our approach, the automatic and context-aware process configuration (discussed in Section 3.2), where pre-defined process types and configuration options are used to automatically generate a process instance containing all necessary activities to match the properties of the current requests situation (cf. DCC4). As basis for this step, we have elaborated a data model where contextual influences are stored (cf. DCC3) alongside different kinds of content-related data. This data model integrates process-related data with customer-related data as well as contextual information. We will now briefly introduce the different kinds of incorporated data by different sections of our data model. At first, such a system must manage data about its customers. Therefore, a customer data section comprises data about the companies, like organizational units or products. Another basic component of industrial production that is important for many topics as sustainability are substances and (sustainability) indicators. As these are not specific for one company, they are integrated as part of a master data section. In addition, the data concretely exchanged between the companies is represented within a separate section (exchange data). To support this data exchange, the system must manage certain data relating to the exchange itself (cf. DCC1): For whom is the data accessible? What are the properties of the requests and responses? Such data is captured in a runtime data section in the data model. Finally, to be able to consistently manage the data request process, concepts for the process and its variants as well as for the contextual meta data influencing the process have been integrated with the other data. More detailed descriptions of these concepts and their utilization will follow in the succeeding sections.

3 Variability Aspects of Data Collection

This section deals with the necessary areas for automated process configuration: The mapping of contextual influences into the system to be used for configuration and the modeling of the latter.

3.1 Context Mapping

As stated in the introduction, a request regarding the same topic (in this case, a sustainability indicator) can have multiple variants that are influenced by a myriad of possible contextual factors (e.g. the number of involved parties or the data formats they use). Hence, if one seeks to implement any kind of automated variant management, a consistent manageable way of dealing with these factors becomes crucial. However, the decisions on how to apply process configuration and variant management often cannot be mapped directly to certain facts existing in the environment of a system. Moreover, situations can occur, in which different contextual factors will lead to the same decision(s) according to variant management. For example, a company could integrate a special four-eyes-principle approval process for the release of data due to different reasons

like if the data is for a specific customer group or if the data relates to a specific law or regulation. Nevertheless, it would be cumbersome to enable automatic variant management by creating a huge number of rules for each and every possible contextual factor. Therefore, in the following, we propose a more generic way of mapping for making contextual factors useable for decisions regarding the data collection process.

In our approach, contextual factors are abstracted by introducing two separate concepts in a lightweight and easily configurable way: The *Context Factor* captures all different possible contextual facts existing in the systems' environment. Opposed to this, the *Process Parameter* is used to model a stable set of parameters directly relevant to the process of data collection. Both concepts are connected by simple logical rules as illustrated on the left side of Figure 2. In this example, a simple mapping is shown. If a contact person is configured for a company (CF1), the parameter 'Manual Data Collection' will be derived. If the company is connected via a tool connector (CF2), automatic data collection will be applied (P3). If the company misses a certain certification (CF3), an additional validation is needed (P2).
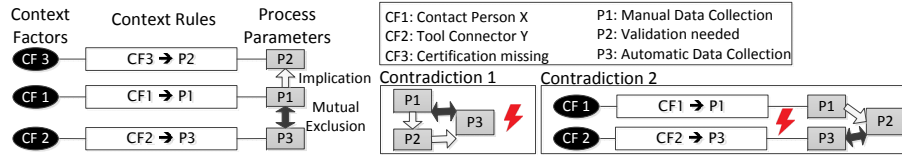

Fig. 2: Context Mapping

When exchanging data between companies, various situations might occur, in which different decisions regarding the process might have implications on each other. For example, it would make no sense to collect data both automatically and manually for the same indicator at the same time. To express that we have also included the two simple constraints 'implication' and 'mutual exclusion' for the parameters. For an example, we refer to Figure 2, where, for example, manual and automatic data collection are mutually exclusive.

Although we have put emphasis on keeping the applied rules and constraints simple and maintainable, there can still exist situations, in which these lead to contradictions. One case (Contradiction 1 in Figure 2) involves a contradiction only created by the constraints, where one activity requires and permits the occurrence of another activity at the same time. A second case (Contradiction 2 in Figure 2) occurs when combining certain rules with certain constraints, in which a contradicting set of parameters is produced. To avoid such situations, we have integrated a set of simple correctness checks for constraints and rules.

3.2 Process Configuration

In this section, we will introduce our approach for process configuration. Therefore, we not only considered the aforementioned challenges, we also wanted to keep the approach as easy and lightweight as possible to enable users of Sustain-Hub to configure and manage the approach. Furthermore, our findings included data about the actual activities of data collection and their relation to contextual data. Data collection often contains a set of basic activities that are part of each data collection process. Other activities appear mutually exclusive, e.g. manual or automatic data collection, and no standard activity can be determined here. In most cases, one or more context factors impose the application of a set of additional coherent activities rather than one single activity.

In the light of these facts, we have opted for the following approach for automatic process configuration: For one case (e.g. a sustainability indicator) a process family is created. The latter contains a *Base Process* with all basic activities for that case. Additional activities that are added to this Base Process are encapsulated in *Process Fragments*. These are automatically added to the process on account of the parameters of the current situation that is represented in the system by the already introduced *Process Parameters* and *Context Factors*. Thus, we only rely on one single change pattern to the processes, an insert operation. This operation has already been described in literature, for its formal semantics, see [2]. Thus our approach avoids problems with other operations as described by other approaches like Provop [3]. Figure 3 shows a simple example of a Base Process that has been configured with Process Fragments (configured areas are marked red). For simplicity, this example uses a subset of the activities of the scenario from the introduction.

To keep the approach lightweight and simple, we decided to model both the Base Process and the fragments in a PAIS (Process-Aware Information System) that will be integrated into our approach. Thus, we can rely on the abilities of the PAIS for modeling and enacting the processes and also for checking their correctness.
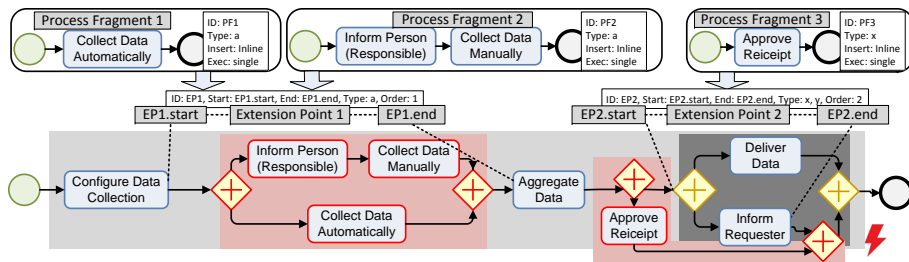

Fig. 3: Process Fragments

To enable the system to automatically extend the base process at the right points with the chosen fragments, we have added the concept of the *Extension*

Point (EP). Both the latter and the fragments have parameters, the system can match to find the right EP for a fragment (see Figure 3 for two example EPs and three fragments with matching parameters). Regarding the connection of the EPs to the Base Processes, we have also evaluated multiple options as, e.g., connecting them directly to activities. Most of such options introduce limitations to the approach or impose a fair amount of additional complexity (cf. [3] for a more detailed discussion). For these reasons we have selected an approach involving two so-called *connection points* of an EP with a Base Process. These points are connected with nodes in the process as shown in Figure 3. Taking the nodes as connection points allows us to reference the nodes' Id for the connection point because this Id is stable and would only change in case of more complicated configuration actions (cf. [3]). If the Base Process contains nodes between the connection points of one EP, an insertion would be applied in parallel to these (cf. EP2 in Figure 3), otherwise sequentially (cf. EP1). Furthermore, if more than one fragment should be inserted at one EP, they will be inserted in parallel to each other (cf. EP1 and fragments 1 and 2 in Figure 3).

By relying on the capabilities of the PAIS we have kept the number of additional correctness checks small. However, the connection points are not checked by the PAIS and could impose erroneous configurations. To keep correctness checks on them simple we rely on two things: The relation of two connection points of one EP and block-structured processes [4]. The first fact spares us from having to check all mutual connections of all connection points as two always belong together. The second implies certain guarantees regarding the structure of the processes. So we only have to check a small set of cases, as e.g., the erroneous definition of EP2 in Figure 3 that would cause a violation to the block structure as shown in the figure.

4 Related Work

Regarding the topic of process configuration, various approaches exist. Most of them focus on the modeling of process configuration. One example is C-EPC [5] that enables behavior-based configurations by integrating configurable elements into a process model. Another approach with the same focus is ADOM [6]. It allows for the specification of constraints and guidelines on a process model to support variability modeling. For all of these approaches two main shortcomings apply: First, they strongly focus on the modeling and neglect execution. Second, configuration must be manually applied by a human, which can be complicated and time-consuming. The approach most closely related to ours is probably Provop [3]. It allows storing a base process and pre-configured configurations to it. Compared to our approach Provop is more fine-grained, complicated, and heavyweight whereas our approach utilizes a set of simplifications that enable a far more lightweight approach. For further reading on the configuration topic, see [7] for an overview of configuration approaches and our predecessor paper for SustainHub [1].

5 Conclusion

In this paper, we have shown a lightweight approach to automatic and contextual process configuration required in complex domains. We have investigated concrete issues in an example domain relating to sustainability data collection in supply chains. With our approach, we have centralized the data and process management uniting many different factors in one data model and supporting the whole data collection procedure by processes executed in a PAIS. Moreover, we have enabled this approach to apply automated process configurations conforming to different situations by applying a simple model allowing for mapping contextual factors to parameters for the configuration. In future work, we plan to evaluate our work with our industrial partners and to extend our approach to cover further aspects regarding runtime variability, automated monitoring, and automated data quality management.

Acknowledgement

References

1. Grambow, G., Mundbrod, N., Steller, V., Reichert, M.: Challenges of applying adaptive processes to enable variability in sustainability data collection. In: 3rd Int'l Symposium on Data-Driven Process Discovery and Analysis. (2013) 74–88
2. Rinderle-Ma, S., Reichert, M., Weber, B.: On the formal semantics of change patterns in process-aware information systems. In: Proc. 27th Int'l Conference on Conceptual Modeling (ER'08). Number 5231 in LNCS, Springer (October 2008) 279–293
3. Hallerbach, A., Bauer, T., Reichert, M.: Configuration and management of process variants. In: Int'l Handbook on Business Process Management I. Springer (2010) 237–255
4. Mendling, J., Reijers, H.A., van der Aalst, W.M.: Seven process modeling guidelines (7pmg). Information and Software Technology **52**(2) (2010) 127–136
5. Rosemann, M., van der Aalst, W.M.P.: A configurable reference modelling language. Information Systems **32**(1) (2005) 1–23
6. Reinhartz-Berger, I., Soffer, P., Sturm, A.: Extending the adaptability of reference models. IEEE Trans on Syst, Man, and Cyber, Part A **40**(5) (2010) 1045–1056
7. Torres, V., Zugal, S., Weber, B., Reichert, M., Ayora, C., Pelechano, V.: A qualitative comparison of approaches supporting business process variability. In: 3rd Int'l Workshop on Reuse in Business Process Management (rBPM 2012). BPM'12 Workshops. LNBIP, Springer (September 2012)