

Towards Process-based Composition of Activities for Collecting Data in Supply Chains

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. Manufacturing companies more and more face the challenge of ensuring sustainable production. In particular, they continuously need to report sustainability data about their products and manufacturing processes that is categorized by various sustainability indicators. However, in a supply chain, such data collection also involves the companies suppliers. Thus, companies must issue cross-organizational data collection processes with potentially high numbers of responders. Due to the heterogeneity in a supply chain and the necessary involvement of services from external sustainability service providers, such processes are often long-running and error-prone. In response to that, we propose an approach for automatically and contextually assembling the required activities and services and managing them by an explicitly specified and enacted process.

Keywords: Process Configuration, Business Process Variability, Data Collection, Sustainability, Supply Chain

1 Introduction

Nowadays, companies collaborate in supply chains in order to assemble complex products like cars or electronic devices. Such companies face a specific challenge: state authorities and the market require sustainable production. Therefore, companies are increasingly forced to report sustainability data about their production that is classified by sustainability indicators like, for example, greenhouse gas (GHG) emissions or the amount of lead contained in products. However, to report respective data, in turn, a company must request it from its suppliers. In general, a sustainability data request could be passed through multiple levels of the supply chain as illustrated in Figure 1.

Sustainability data requests involve great heterogeneity: different companies follow different approaches in sustainability data management. Some of them have different in-house systems (IHSs) for this (cf. S1.1 in Figure 1), whereas others rely on manual management (cf. S1.3) or even have no proper approach to it at all. Furthermore, in various cases, services of external service providers may be required. For example, a company reporting GHG emissions in production (cf.

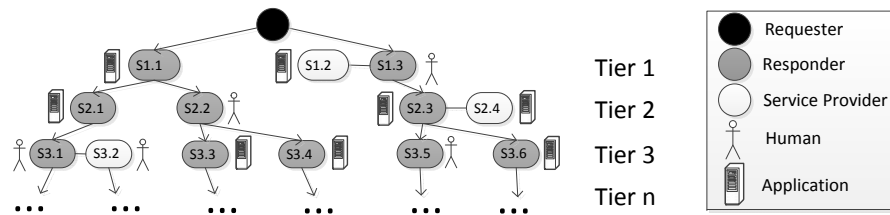


Fig. 1: Supply Chain Data Collection

S1.3) might need an external service provider to validate the data (cf. S1.2) after having collected their own data and received relating data from its suppliers (cf. S2.3).

Due to these properties, data collection process can be long-running, tedious, and error-prone, which may even involve legal fines for the reporting companies. Due to the heterogeneity in tools and approaches, it is not possible to apply a federated information system or data base to all companies. Furthermore, as such data exchange not only involves the mere exchange of data but involves various kinds of activities relating to such data, simple data integration approaches, as e.g., utilizing ontologies are also not sufficient. The following scenario gives a small-scale industrial example for such a data collection process.

Scenario: Sustainability Data Collection

An automotive company wants to collect sustainability data relating to a specific part. In particular, a regulation requests the reporting of the quantity of lead in that part. This also concerns sub-parts of that part that are delivered by two suppliers of the company. One of them is a bigger company with a IHS in place. The other one is a smaller company with no system and no dedicated responsible for sustainability. The IHS of the bigger company has its own data format that has to be explicitly converted to be useable. For the smaller company, a service provider is needed that will validate the manually collected data to ensure that it complies with legal regulations. 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 project¹, we are developing a centralized information exchange platform (also called SustainHub) that supports sustainability data collection along the entire supply chain. For this purpose, we have investigated and discussed the challenges for sustainable supply chain communication as well as the state-of-the-art [4].

As this paper focuses on the composition of activities and services, first of all we summarize the challenges a system must tackle to enable this. (DCC1) Most of

¹ SustainHub (Project No. 283130) is a collaborative project within the 7th Framework Programme of the European Commission (Topic ENV.2011.3.1.9-1, Eco-innovation).

the activities in sustainability data exchange are still executed manually. Taking into account that such data exchange takes place in complex supply chains, it can be problematic to even find the right person in the right department in the right company or the right service of the right service provider. To enable automated support, such information must be explicitly stored and managed. (DCC2) Different companies have different ways to manage relevant sustainability data. Some use IHSs, whereas others rely on manual data management. A system supporting data collection in a supply chain, therefore, must be able to access it in both ways, i.e. it must support manual as well as automated data collection. (DCC3) The requests in sustainability data exchange rely on a myriad of different factors (e.g., legal requirements, IHS used in a company). To support repeatable data collection, a system must be aware of such contextual factors and manage them centrally. (DCC4) Due to the great number of different factors, each request is not far from being unique and has to be executed manually. A system supporting such data exchange should enable the centralized definition of the data collection process and also its different variants. Both definitions should be as simple as possible to not burden users with a cumbersome and error-prone modeling process.

The remainder of this paper is organized as follows: Section 2 presents our approach for sustainability data collection and exchange. Section 3 gives a brief overview about related work followed by a conclusion.

2 Automated Process for Data Collection

Basically, our approach for supporting the complex process of sustainability data collection involves the idea to govern that process by an explicitly specified process, which is automatically enacted by a Process-Aware Information System (PAIS). That way, each request can be managed in a centralized way (cf. DCC1) and be specified explicitly (cf. DCC4) via a process template. Furthermore, this approach bears another advantage: For the different activities in such a process, custom components can be applied. These components can be used for manual activities as well as connections to different IHSs. In addition, other components can realize services of various service providers. This facilitates support for manual and automatic data collection (cf. DCC2). Further, makes the processes modular and reusable.

However, such an approach would involve rigidly pre-specified process templates and still no awareness of meta data like contextual factors. A large number of process templates would have to be specified in advance incorporating every possible combination of eventualities. A human would then have to select the right one being aware of each and every parameter of the current request. This would be tedious and error-prone. In response to this, we have created an approach capable of automatically incorporating contextual factors into a system and, based on them, creating various variants of pre-specified processes exactly matching the current request situation. Figure 2 illustrates the different components of this approach.

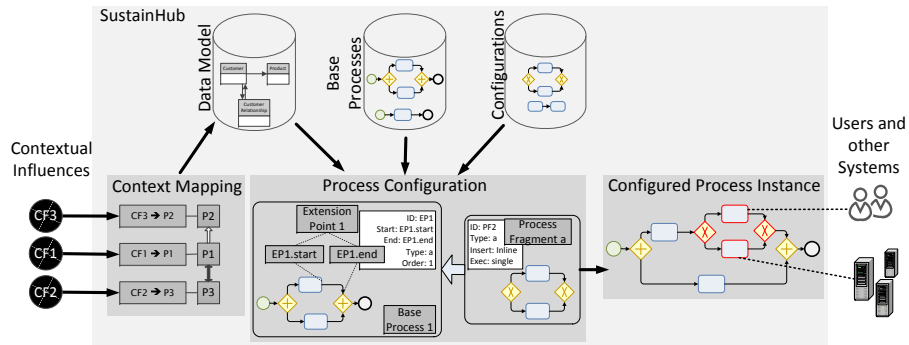


Fig. 2: Automated Process and Service Configuration

To be able to automatically process contextual factors and to utilize them for the automatic creation of process variants, our approach applies a set of different components: to incorporate contextual factors, the 'Context Mapping' component maps these to parameters directly usable for process configuration. The latter is applied by the 'Process Configuration' component that creates process variants (cf. DCC4) based on two entities: base processes incorporating the basic activities necessary for a specific type of request and process fragments depicting sets of activities suitable for a specific situation. By combining of these two, a configured process instance is created, as shown in Figure 2.

However, to be able to automatically apply such configurations, the 'Process Configuration' component must be aware of various facts and their relations. This includes data about the companies connected to SustainHub to be able to deliver activities to the right persons or IHSs as well as basic domain data accessible for all companies, like sustainability indicators. Furthermore, the actual data exchange (e.g. the data of a request) and the content exchanged must be stored and available to SustainHub as well. All this data must be mutually connected as well as connected to other data serving as basis for automatically managing context data and process configuration and enactment. Therefore, we apply a data model that is more comprehensive as those usually applied in PAIS integrating types of data usually found in other systems (e.g., ERP systems). We will not explain all entities here, however, we will introduce six sections of it containing entities for different purposes: 'customer data' (e.g. organizational models or IHS references), 'master data' (e.g. indicators or substances), 'runtime data' (e.g. request data), 'content data' (e.g. values actually exchanged), 'context data' (contextual factors), and 'process data' (e.g. data necessary for managing and configuring processes automatically). Having all the data in place, we now go into detail about the two main components for context mapping and process configuration.

2.1 Context Mapping

As already stated, variants of sustainability data requests can depend on a myriad of contextual factors. Consequently, a system enabling automated management of

such variants must apply a consistent way of managing and storing such factors (cf. DCC3). Moreover, there is often no one-to-one mapping between a factor and a variant or a certain set of activities. For example, a company might apply a special four-eyes-principle approval process in two cases. The first involves data relating to a specific law and involving high fines. The second concerns data relating to a specific customer group that the company has no high trust in. Enabling variant management by creating one rule to apply one certain variant (or adaptation) to a base process would result in a high number of rules. This would bloat the needed data and make modeling and maintaining cumbersome. To avoid this, we apply a simple and lightweight mapping of contextual factors to process parameters that can be directly used for the process variant configurations. As illustrated in Figure 2, our approach features simple logical mapping rules (e.g. $CF1 \wedge CF2 \rightarrow P1$) and also the option to apply simple consistency rules to avoid erroneous configurations (e.g. P1 and P3 mutually exclude each other). A simple example for a context factor as shown in the scenario in the introduction would be the approach and tool a company uses for sustainability data management. This can be mapped to concrete process parameters, e.g., that the company uses a specific IHS for a specific data collection task. That way, a distinct set of parameters for the selection of configurations can be created.

2.2 Process Configuration

When a stable set of parameters is in place for a specific request, it must be determined what exactly shall be inserted into its base process and where to insert it. This requires that options for both of these decisions are available in the variant model of SustainHub. As stated in DCC2, both the definition of the base processes and the configurations should be as simple as possible to not burden the users with a complicated modeling process. Therefore, we aimed for a simple and lightweight way of modeling.

Our studies have shown that for most requests, a basic set of activities is mandatory, e.g. configuration of the data collection or the final data delivery. Therefore, we have decided to allow for the modeling of base processes with mandatory activities for different cases (as e.g. sustainability indicators) and to only extend these processes with additional activities instead of also applying deletions. These base processes are then annotated with *extension points* to indicate where an extension is feasible. As illustrated in Figure 2 such extension points have two connections with the process to clearly indicate, where the insertion should happen and a set of meta information, SustainHub can then use to determine, which fragment would be suitable at that position. Furthermore, a set of options is used to determine, if fragments should be inserted as sub-process or directly in the base process (inline) and in which order they are to be inserted if multiple extension points are at the same position.

Our studies showed that a specific set of activities is often cohesively needed in a specific situation (e.g. if data has to be collected manually, that activity involves also other activities like informing the responsible person). Therefore, we have decided to apply whole fragments instead of fine-grained adaptations adding

single activities. Moreover, the approach becomes easier to model and maintain that way as we allow to model such fragments same as the base processes in a PAIS. That way, that PAIS manages their structural correctness and other basic factors.

2.3 Example Scenario

In this section, we show the application of our approach to the concrete industrial example scenario applied in the introduction. Figure 3 illustrates this including a base process, different extension points, context factors, process fragments, and a resulting configured process. The base process comprises three activities for configuring the request, aggregating the results, and delivering them to the requester (cf. Figure 3). It also has two extension points, EP1 for the data collection activities, EP2 for post processing activities. Via a specific parameter (Order) it is ensured, that EP1 fragments are inserted before EP2 fragments. Furthermore, various fragments are in place. In Figure 3, four of them are shown, for automatic and manual data collection as well as for external validation of manually collected data and for conversion of automatically collected data.

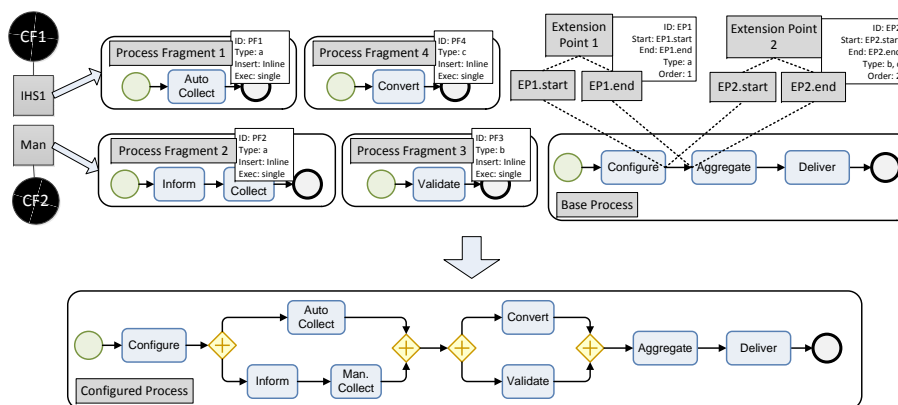


Fig. 3: Process Configuration Scenario

For this request, the system has the facts in place, that two responders are involved and that one of them has an IHS while the other relies on manual data collection (for which the responsible person is also modeled within SustainHubs' data model). Thus, these context factors can be mapped to two process parameters indicating one responder with different properties each. Each of the parameters is configured to imply two fragments, one for data collection, one for post processing. These fragments are then automatically integrated into the base process, all of them inline, as configured. Fragment 1 and 2 are integrated in parallel via EP1 and fragment 3 and 4 also via EP2. Finally, the resulting configured process is shown in the lower section of Figure 3.

3 Related Work

Our approach presented in this paper enables the automated assembly of various activities and services necessary for sustainability data requests applying contextually configured processes for that. Therefore, and due to the lack of space, we limit our review of related work to process configuration approaches. Examples include ADOM [6] that relies on software engineering principles or configuration modeling approaches like C-EPC [7] or C-YAWL [3]. Like the most other approaches, they focus solely on the modeling of process configuration. They take different approaches to configuration: ADOM enables the specification of guidelines and constraints, while C-EPC integrates configurable elements into the model. C-YAWL allows for hiding single elements or even blocking whole execution paths. For a qualitative comparison of such configuration approaches, see Ayora et al. [1]. Besides the fact that such approaches only focus on modeling, they also require human interaction for manual application of the configuration in contrast to our approach.

In addition to this, other approaches like [8] target the correctness of process configurations. In contrast to them, our approach encapsulates the minimal set of necessary activities into a base process and other sets of activities in process fragments. Both of these are modeled in a PAIS and can be checked for correctness therein. Furthermore, we limit the complexity of the configurations with the explicit extension points. That way, our approach becomes more lightweight and easy to handle.

Provop [5] constitutes an approach that is more closely-related to our approach. It enables the modeling of base processes and pre-specified configuration options and also the execution of processes configured that way. However, it is more fine-grained and complicated as our approach and it does not allow completely automatic context acquisition, processing, and process configuration. For a broader view on related work, see our paper on challenges and state-of-the-art [4].

4 Conclusion

In this paper, we have introduced an approach for applying configurable processes for complex data collection scenarios like sustainability data exchange in supply chains. The approach is lightweight featuring pre-specified base processes and allows for configuring these with also pre-specified process fragments to adhere to the specifics of various different situations. Moreover, our approach enables such configurations to be applied automatically involving techniques for acquiring, storing, and managing various contextual factors to which the processes must comply.

Our future work will involve the extension of our approach to satisfy all requirements we have specified in [4]. This involves runtime adaptations to processes (e.g., for situations, in which the context changes while a request is already processed) as well as monitoring and quality management capabilities. Furthermore, we have already started to implement our approach on top of the

AristaFlow BPM suite [2] and have also begun to evaluate it using 66 sustainability indicators we have collected from industry surveys in the SustainHub project.

Acknowledgement

The project SustainHub (Project No. 283130) is sponsored by the EU in the 7th Framework Programme of the European Commission (Topic ENV.2011.3.1.9-1, Eco-innovation).

References

1. Ayora, C., Torres, V., Weber, B., Reichert, M., Pelechano, V.: Enhancing modeling and change support for process families through change patterns. In: 14th Int'l Working Conference on Business Process Modeling, Development, and Support (BPMDS'13). pp. 246–260. No. 147 in LNBIP, Springer (June 2013), <http://dbis.eprints.uni-ulm.de/930/>
2. Dadam, P., Reichert, M.: The ADEPT project: A decade of research and development for robust and flexible process support - challenges and achievements. *Computer Science - Research and Development* 23(2), 81–97 (2009), <http://dbis.eprints.uni-ulm.de/486/>
3. Gottschalk, F., van der Aalst, W.M.P., Jansen-Vullers, M.H., La Rosa, M.: Configurable workflow models. *Int. J. Cooperative Inf. Syst.* 17(2), 177–221 (2008)
4. 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. pp. 74–88 (2013)
5. Hallerbach, A., Bauer, T., Reichert, M.: Capturing variability in business process models: The provop approach. *Journal of Software Maintenance and Evolution: Research and Practice* 22(6-7), 519–546 (November 2010), <http://dbis.eprints.uni-ulm.de/628/>
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), 1045–1056 (2010)
7. Rosemann, M., van der Aalst, W.M.P.: A configurable reference modelling language. *Information Systems* 32(1), 1–23 (2005)
8. Van Der Aalst, W., Lohmann, N., La Rosa, M., Xu, J.: Correctness ensuring process configuration: An approach based on partner synthesis. In: *Business Process Management*, pp. 95–111. Springer (2010)