Towards Collecting Sustainability Data in Supply Chains with Flexible Data Collection Processes

Gregor Grambow[(✉)], Nicolas Mundbrod, Jens Kolb, and Manfred Reichert

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

Abstract. Nowadays, OEMs from many domains (e.g., electronics and automotive) face rising pressure from customers and legal regulations to produce more sustainable products. This involves the reporting and publishing of various sustainability indicators. However, the demands of legal entities and customers constitute a tremendous challenge as products in these domains comprise various components and sub-components provided by suppliers. Hence, sustainability data collection must be executed along the entire supply chain. In turn, this involves a myriad of different automated and manual tasks as well as quickly changing situations. In combination with potentially long-running processes, these issues result in great process variability that cannot be predicted at design time. In the SustainHub project, a dedicated information system for supporting data collection processes is developed. This paper provides three contributions: (1) it identifies core challenges for sustainable supply chain communication, (2) it reviews state-of-the-art technical solutions for such challenges, and (3) it gives a first overview of the approach we are developing in the SustainHub project to address the challenges. By achieving that, this comprehensive approach has the potential to unify and simplify supply chain communication in the future.

Keywords: Process variability · Data collection · Sustainability · Supply chain

1 Introduction

Companies of the electronics and automotive industry face steadily growing demands for sustainability compliance triggered by authorities, customers and the public opinion. As products often consist of numerous individual components, which, in turn, comprise sub-components, heterogeneous sustainability data need to be collected along intertwined and intransparent supply chains. As consequence, highly complex, cross-organizational data collection processes are required that feature a high variability. Further issues include incompleteness and varying quality of provided data, heterogeneity of data formats, or changing situations and requirements. So far, there has been no dedicated information system

(IS) supporting companies in creating, managing and optimizing such data collection processes. In the SustainHub[1] project, such a dedicated information system is being developed. In the context of this project, we have intensively studied use cases, which were delivered by industry partners from the automotive and electronics domain in order to elaborate core challenges and requirements regarding the IT support of adaptive data collection processes. To assess whether existing approaches and solutions satisfy the requirements, state-of-the-art has also been thoroughly studied as well. This paper presents core challenges with respect complex sustainability data collection processes along today's supply chains and presents the state-of-the-art in this context. Supply chains are well suited for eliciting these challenges because of the complexity on one hand and the requirements imposed by emerging laws and regulations on the other. However, the core challenges identified apply to many other domains as well.

Altogether, this paper reveals seven core challenges for data exchange and collection in complex distributed environments and evaluates existing approaches to contribute to tackle these challenges. Besides the clear focus on challenges and requirements, this paper also gives a first abstract outlook on a system we are currently developing to tackle the challenges. Thereupon, future research on adaptive business process management technology can be aligned to support more variability and dynamics in today's data collection processes.

Fundamentals of sustainable supply chains as well as an illustrating example are introduced in Sect. 2. Then, seven data collection challenges are unveiled in Sect. 3, exposing concrete findings, identified problems and derived requirements. In Sect. 4, the current state-of-the-art is presented. Following this, we briefly discuss the approach we are developing to solve the reported issues in Sect. 5. Finally, Sect. 6 rounds out this paper giving a conclusion and an outlook.

2   Sustainable Supply Chains

This section gives insights into sustainable supply chains and provides an illustrating example.

2.1   Fundamentals

The development and production of products is often based on complex supply chains involving dozens of interconnected companies distributed around the globe. In order to ensure competitiveness, complex communication tasks must be effectively and efficiently managed for in the context of cross-organizational processes. Generally, such a cross-organizational collaboration consists of a variety of both manual and automated tasks. Moreover, involved companies significantly differ in size and industry background and use heterogeneous ISs. Due to this heterogeneity, neither federated data schemes nor unifying tools or other concepts can be adopted in this context [1].

---

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

As sustainability constitutes an emerging trend, manufacturers face new challenges in their supply chains: sustainable development and production. The incentives are given by two parties: On one hand, legal regulations, increasingly issued by authorities, force companies to publish more and more sustainability indicators on an obligatory basis. Examples include greenhouse gas emissions in production and gender issues. On the other, public opinion and customers compel manufacturers to provide sustainability information (e.g., organic food) as an important base for their purchase decisions.

Prevalent examples of standards and regulations are the ISO 14000 standard for environmental factors in production, GRI[2] covering sustainability factors or regulations like REACH[3] and RoHS[4]. Overall, sustainability information involve a myriad of indicators, relating to social issues (e.g., employment conditions or gender issues), to environmental issues (e.g., hazardous substances or greenhouse gas (GHG) emissions), or managerial issues (e.g., compliance).

There already exist tools providing support for the management and transfer of sustainability data: IMDS[5] (International Material Data System), for instance, is used in the automotive industry. IMDS allows for material declaration by creating and sharing bills of materials (BOM) among different companies. A similar system exists for the electronics industry (i.e., Environ BOMcheck[6]). Despite some useful support regarding basic data declarations and exchange tasks, these tools fail in providing dedicated support for sustainability data collection and exchange along supply chain.

2.2 Illustrating Example

To illustrate the complexity of sustainability data collection processes in a distributed supply chain, we provide an example. The latter exposes requirements gathered from companies from the automotive and electronics industry based on surveys and interviews. Note that data collection in such a complex environment does not have the characteristics of a simple query. It is rather a varying, long-running process incorporating various activities and techniques for gathering distributed data, and involving different participants.

The example illustrated in Fig. 1 depicts the following scenario: Imposed by regulations, an automotive manufacturer (requester) must provide sustainability data relating to its production. This data is captured by two sustainability indicators, one dealing with the greenhouse gas emissions regarding the production of a certain product, the other addressing the REACH regulation. The latter concerns the whole company as companies usually declare compliance to that regulation as a whole.

---

[2] Global Reporting Initiative: https://www.globalreporting.org.
[3] Regulation (EC) No. 1907/2006: Registration, Evaluation, Authorisation and Restriction of Chemicals.
[4] Directive 2002/95/EC: Restriction of (the use of certain) Hazardous Substances.
[5] http://www.mdsystem.com.
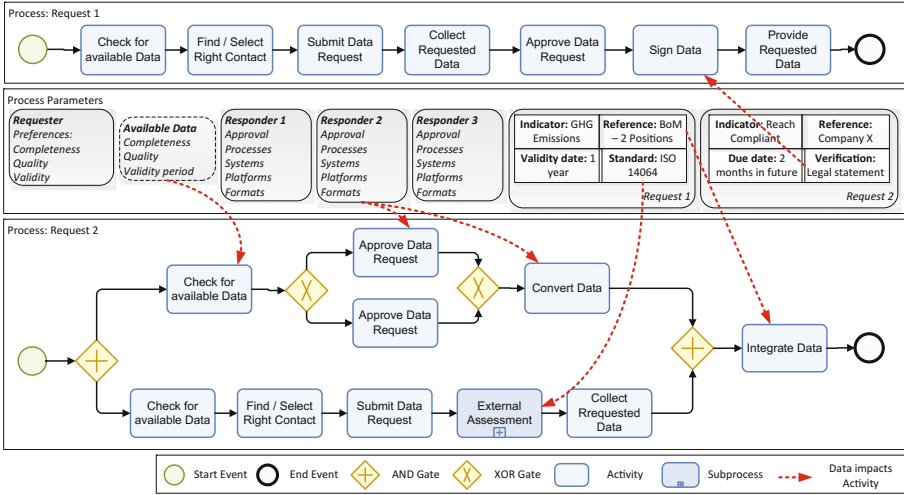[6] https://www.bomcheck.net.

Fig. 1. Examples of two data collection processes

To provide data regarding these two indicators, the manufacturer must gather related information from its suppliers (responder). Hence, it requests a REACH compliance statement from one of its suppliers. To obtain the respective the information, the activities shown in the process *Request 1* must be executed. Furthermore, the product for which the greenhouse gas emissions shall be indicated has a BOM with two items coming from external suppliers. Thus, the request, depicted by the second process, has to be split up into two requests, one for each supplier.

The basic scenario involves a set of activities as part of the data collection processes. Some of these are common for both requests; e.g., on the requester side, checking available data that might satisfy the request, selecting the company and contact person, and submitting the request. On the responder side, data must be collected and provided. In turn, other process activities are specifically selected for each case. Thereby, the selection of the activities is strongly driven by data (process parameters) provided by the requester, the responder, the requests and indicators, and data that may already be available.

For example, *Request 1* implies a legally binding statement considering REACH compliance. Therefore, a designated representative (e.g., the CEO) must sign the data. In many cases, companies have special authorization procedures for releasing such data, e.g., one or more responsible persons may have to approve the request (see the parallel approval activities (*Approve Data Request*) in the context of *Request 2* expressing a *four-eyes-principle*). In some cases, data might be already available in a company, i.e., it needs not to be manually gathered (cf. *Request 2*, *Check of available Data*). However, every time the company-internal format of the responder does not match the requester's one, a conversion becomes necessary. Further, some indicators and requests directly

relate to a given standard (e.g., ISO 14064 for greenhouse gases). In turn, this may directly trigger an assessment of the responder if he cannot exhibit the fulfillment of the standard (cf. *Request 2*, *External Assessment*).

Another important aspect of (long-running) data collection processes is that process parameters might change over time and hence exceptional situations might occur. Even in this very simple example, many variations and deviations might happen: for example, if the CEO was not available, activity *Sign Data* could be delayed. In turn, this may become a problem if defined deadlines exist for the query answer.

3   Data Collection Challenges

This section presents seven challenges for an information system supporting sustainability data collection processes along an entire supply chain (IS-DCP). The results are based on findings from case studies conducted with industrial partners in the SustainHub project. Three figures serve for illustration purposes: Fig. 2 illustrates data collection challenges DCC 1 and DCC 2, Fig. 3 illustrates DCC 3 and DCC4, and Fig. 4 illustrates DCC 5-7.
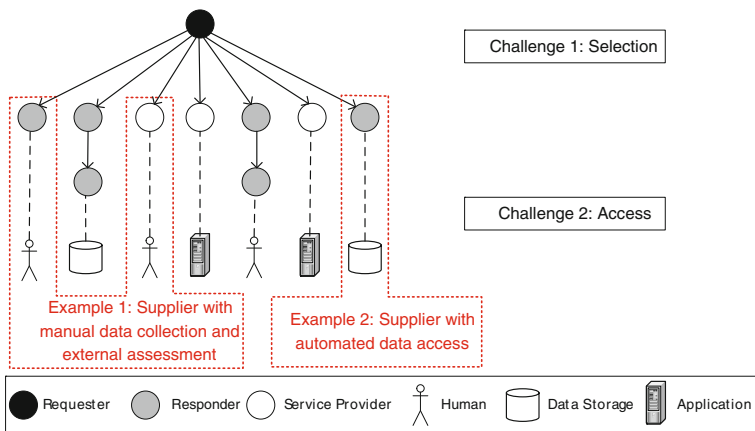

Fig. 2. Data collection challenges DCC 1 and DCC 2

3.1   DCC 1: Dynamic Selection of Involved Parties

Findings. In a supply chain, sustainability data collection involves various parties (cf. Fig. 2). A single request may depend on the timely delivery of data from different companies. For manual tasks, this may have to be accomplished by a specific person with sustainability knowledge or authority. In big companies, in turn, it can be a challenging task to find the right contact person to answer a specific request. In relation, contact persons may change over time. Furthermore, as the requested data is often complex, has to be computed, or relates to

legal requirements, external service providers may be involved in the data collection request as well. Relating to our scenario from Sect. 2.2, Fig. 2 includes two concrete examples: a supplier that applies manual data collection and needs an assessment by an external service provider, and a supplier providing automated access to his data. Finally, regarding the timely answering of a request, many requests may be adjusted and forwarded to further suppliers (cf. Fig. 2); thus, answering times can multiply.

Problems. The contemporary approach to such requests relies on individuals conducting manual tasks and interacting individually. There are tools (e.g., email) which can provide support for some of these tasks and partly automate them. However, much work is still coordinated manually. As a request can be forwarded down the supply chain, it is difficult to predict, who exactly will be involved in its processing. From this we can conclude that answering times of requests can be hardly estimated in a reliable manner as well.

Requirements. An IS-DCP need to enable companies to centrally create and manage data collection requests. Thereby, it must be possible to simplify the dynamic selection process of involved parties and contact persons regarding the request responders as well as potentially needed service providers. This is a basic requirement for enabling efficient request answering, data management, and request monitoring.

3.2 DCC 2: Access to Requested Data

Findings. In a supply chain different parties follow different approaches to data management. While large companies usually have implemented a higher level of automation, SMEs typically rely on the work of individual persons. Furthermore, sustainability reporting is still an emerging area and there exists no unified reporting method along supply chains. In particular, this implies a high degree of variability when it comes to accessing internal data of companies. Some of them have advanced software solutions with respect to data management, some manage their data in databases, some store it in specific files (e.g., Excel), and others have even not started to manage sustainability data yet.

Problems. The contemporary approach to sustainability reporting is managed manually to a large extend. This involves manual requests from one party to another and different data collection tasks on the responder side. This can impose large delays in data collection processes as sustainability data must be manually gathered from systems, databases or specific files before it can be compiled, prepared and authorized in preparation to the delivery to the requester.

Requirements. An IS-DCP must accelerate and facilitate the access to requested sustainability data. On one hand, this requires guiding users in manually collecting data as well as in automizing data-related activities (e.g., data approval, data transformation) where possible. On the other hand, automatic data collection should be enabled whenever possible. This requires accessing the systems containing the data automatically (e.g., via the provision of appropriate interfaces)

and including manual approval activities when needed. Finally, data conversion between different formats ought to be supported as a basis for data aggregation.

3.3   DCC 3: Meta Data Management

Findings. The management and configuration of sustainability data requests in a supply chain relies on a myriad of different data sets. As aforementioned, this data stems from heterogeneous sources. Examples of such parameters include the preferences of the requester as well as the responders (including approval processes and data formats), or the properties of the sustainability indicators (e.g., relations to standards) (cf. Fig. 3). Involving the scenario from Sect. 2.2, concrete examples include the following: a mismatch of the data format configuration of the requester and responder, the need to comply to a specific standard as the ISO 14064, or available data the matches the quality requirements of the requester (also illustrated by Fig. 3). As a result, potentially matching data might be already available in some cases but expose different properties as requested.
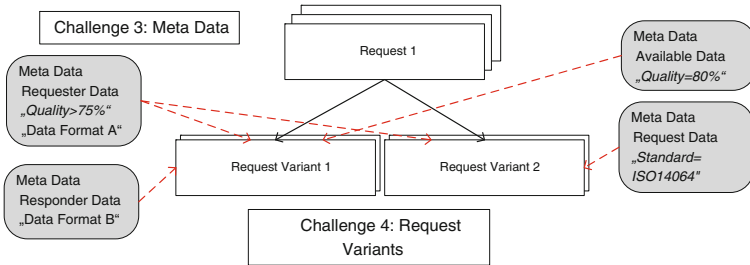

Fig. 3. Data collection challenges DCC 3 and DCC 4

Problems. As requests rely on heterogeneous data, they are difficult to manage. Requirements are partially presumed by the requester and often are implicit. Hence, responders might be unaware of the requirements and deliver data not matching them. Moreover, it is difficult to determine whether data, which has been collected before, matches with a new request. Finally, as a supply chain might involve a large number of requesters and responders, this problem multiplies as crucial request data is scattered along the entire supply chain.

Requirements. To be able to consistently and effectively manage data collection processes, an IS-DCP must centrally implement, manage and provide an understandable meta data schema addressing relevant request parameters. Thereby, instanced data based on the uniform meta data schema can be effectively used to directly derive and adjust variants of data collection processes.

3.4   DCC 4: Request Variants

Findings. As mentioned, sustainability data exchange in a supply chain involves a considerable number of manual and automated tasks aligned to the current

data request. Hence, execution differs greatly among different data requests, highly influenced by parameters and data, and distributed on many sources (cf. DCC 3 and Fig. 3). Moreover, the reuse of provided data is problematic as well as the reuse of knowledge about conducted data requests: persons in charge, managing a data collection, might not be aware of which approach matches the current parameter set best.

Problems. This makes the whole data collection procedure tedious and error prone. Based on the gained insights, initially, to each data request a data collection process is defined manually, and evolves stepwise afterwards. Relying on the various influencing parameters, every request must be treated individually – there is no applicable uniform approach to a data request, instead a high number of variants of data collection processes exist. So far, there is no system or approach in place that allows structuring or even governing such varying processes along a supply chain.

Requirements. An IS-DCP not only needs to be capable of explicitly defining the process of data collection. Due to the great variability in this domain, it must be also capable of managing numerous variants of each data request relating to a given parameter set. This includes the effective and efficient modeling, management, storage, and execution of data collection request processes.

3.5   DCC 5: Incompleteness and Quality

Findings. Sustainability data requests are demanding and their complex data collection processes evolve based on delivered data and forwarded requests to other parties (i.e., suppliers of the suppliers) (cf. Fig. 4). Furthermore, they are often tied to regulative requirements and laws, and also involve mandatory deadlines. Therefore, situations might occur, in which not all needed data is present, but the request answer must still be delivered due to a deadline. As another case, needed data might be available, but on different quality levels and/or in different formats.

Problems. Contemporary sustainability data collection in supply chains is plagued by quality problems relating to the delivered data. Not only that requests are incompletely answered, the requester also has no awareness of the completeness and quality of the data stemming from multiple responders. Moreover, responders have no approach to data delivery in place when being unable to provide the requested data entirely, or their data does not match the request's quality requirements. Missing a unified approach, definitive assertions or statements to the quality of the data of one request can often not be made and requests might even fail due to that fact.

Requirements. An IS-DCP must be able to deal with incomplete data and quality problems. It must be possible that a request can be answered despite missing or low quality data. Furthermore, such a system must be able to make assumptions about the quality of the data that answers a request.
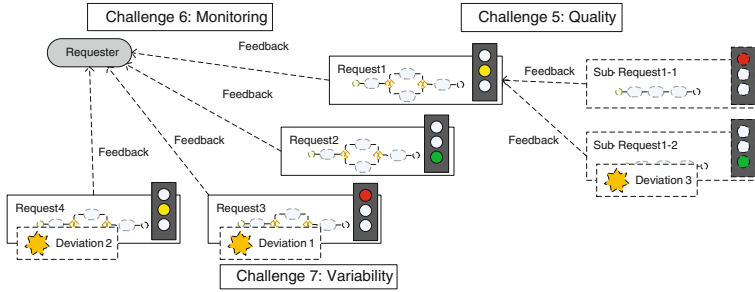
Fig. 4. Data collection challenges DCC 5-7

3.6   DCC 6: Monitoring

Findings. Sustainability data collection along the supply chain involves many parties and logically may take a long time. The requests exist in many variants and the quality and completeness of the provided data differ greatly (cf. DCC 5). The contemporary approach to such requests does not provide any information about the state of the request to requesters before the latter is answered (cf. Fig. 4). This includes missing statements about delivered data as well as the possibly existing recursive requests along the supply chain. Thus, it can be a serious issue for the OEM who issued the initial request to gain an awareness about possible delays and to gather information about their location in the supply chain.

Problems. As a requester has no information about the state of his request and potential data delivery problems, the latter solely become apparent when deadlines are approaching. At that time, however, it might be too late to apply countermeasures to avoid low quality, incomplete data, or responders delivering no data at all.

Requirements. An IS-DCP must be capable of monitoring complex requests spanning multiple responders as well as various manual and automatic activities. Furthermore, a requester should be able to be actively or passively informed about the state of the activities along the data collection process as well as the state of the data delivered.

3.7   DCC 7: Run Time Variability

Findings. The processing of a data collection request might take a long time to answer if the request involves a great number of parties. Further, it exposes manual and automatic activities, different kinds of data and data formats, and unforeseen impacts on the data collection process. This implies that parameters, on which the data collection relies, may change during execution of a data collection process. Exceptional situation handling occurs as a result of expiring deadlines or responders not delivering data.

Problems. The variability relating to sustainability data collection processes constitute a great challenge for companies. Running requests might become invalidated due to the aforementioned issues. However, there is no common sense or standard approach to this. Instead, requesters and responders must manually find solutions to still get requests answered in time. This includes much additional effort and delays. Another issue are external assessments: they could not only be delayed but also completely fail, leaving the responder without a required certification. The final problem touched by this example concerns mostly long-running data collection processes: data, that was available at the beginning of the query, could get invalid during the long-term process (e.g., if it has a defined validity period).

Requirements. An IS-DCP must cope with run-time variability occurring in today's sophisticated sustainability data collection processes. As soon as issues are detected, data collection processes must be timely adapted to the changing situation in order to keep the impact of these issues as considerable as possible. This requests a system which is able to dynamically adapt already running data collection processes without invalidating or breaking the existing process flow.

4 State of the Art

This section gives insights on the state of the art in scientific approaches relating to the issues shown in this paper. It starts with a broader overview and proceeds with more closely related work including three subsections.

Section 3 underlines that exchanging data between different companies along a supply chain in an efficient and effective way has always been a challenge. Nonetheless, this exchange is not only necessary—it is now a crucial success factor and a competitive advantage, these days. However, many influencing factors hamper the realization of a data exchange being automated and homogeneous. In particular for those companies aiming to address holistic sustainability management, the inability to implement automated and consistent data exchange is a big obstacle. Please remind that these companies need to take into account existing and even emerging laws as well as regulations requesting to gather and distribute information about their produced goods. Furthermore, that requested information need be gathered from their suppliers as well. Hence, complex data collection processes, involving a multitude of different companies and systems, have to be designed, conducted, and monitored to ensure compliance. So far, we could not locate any related work that completely addresses the aforementioned challenges (cf. Sect. 3).

For complex data collection processes, IS support in the supply chain is desirable supporting communication and enabling automated data collection. The importance and impact of an IS for supply chain communication has already been highlighted in literature various times. In [2], for instance, a literature review is conducted showing a tremendous influence of ISs on achieving effective SCM. The authors also propose a theoretical framework for implementing ISs in the supply chain. Therefore, they identify the following core areas: strategic planning, virtual

enterprise, e-commerce, infrastructure, knowledge management, and implementation. However, their findings also include that great flexibility in the IS and the companies is necessary and that IS-enabled SCM often requires major changes in the way companies deal with SCM. As another example, [3] presents an empirical study to evaluate alternative technical approaches to support collaboration in SCM. These alternatives are a centralized web platform, classical electronic data interchange (EDI) approaches, and a decentralized, web service based solution. The author assesses the suitability of the different approaches with regard to the complexity of the processes and the exchanged information. Concluding, relating work in this area reveals various approaches to SCM management However, these are mostly theoretic, rather general, and not applicable to the specific use cases of sustainability data collection processes.

As automation can be a way to deal with various issues of sustainability data collection, respective approaches addressing that topic can be found in literature as well. However, none of them applies to the domain of sustainable supply chain communication and its specific requirements. For example, [4] presents an approach to semi-automatic data collection, analysis, and model generation for performance analysis of computer networks. This approach incorporates a graphical user interface and a data pipeline for transforming network data into organized hash tables and spread sheets for its usa in simulation tools. As a specific type of data transformation is considered, it is not suitable in our context. Such approaches deal with automated data collection; yet they are not related to sustainability or SCM and the problems arising in this setting.

There exist several approaches dealing with sustainability reporting (e.g., [5–8]). However, they do not propose technical solutions for automated data collection. Rather they approach the topic theoretically by analysing several relating facts. These include the importance of corporate sustainability reporting, sustainability indicators or the process of sustainability reporting as a whole. Another goal is building a sustainability model by analysing case studies.

Besides approaches targeting generic sustainability, SCM and data collection issues, there exist three areas that are more closely related to our problem context. As discussed, sustainability data collection processes involve numerous tasks to be orchestrated. Data requests may exist in many different variants based on a myriad of different data sources and may be subjected to dynamic changes during run-time (cf. DCC 7). Therefore, this sub-section discusses approaches for process configuration (Sect. 4.1), data- and user-driven processes (Sect. 4.2), and dynamic processes (Sect. 4.3).

4.1 Process Configuration

Behaviour-based configuration approaches enable the process modeller to provide pre-specified adaptations to process behaviour. One option for realizing this is hiding and blocking as described by [9]: blocking allows disabling the occurrence of a single activity/event, whereas hiding allows hiding single activity to be hidden, which is then executed silently; succeeding activities in that path are still accessible.

Another way to enable process model configuration for different situations is to incorporate configurable elements into the process models as described in [10,11]. An example of this approach is a configurable activity, which may be integrated, omitted, or optionally integrated surrounded by XOR gateways. Another approach enabling process model configuration is ADOM [12] that builds on software engineering principles and allows for the specification of guidelines and constraints with the process model. A different approach to process configuration is taken by structural configuration, which is based on the observation that process variants are often created by simply copying a process model and then applying situational adaptations to it. A sophisticated approach dealing with such cases is Provop [13], which realizes a configurable process model by maintaining a base process models and pre-specified adaptations to it. The latter can be related to context variables to enable the application of changes matching to different situations. Finally, [14,15] provide a comprehensive overview of existing approaches targeting process variability.

Process configuration techniques provide a promising approach in our context. Nevertheless, they do not fully match the requirements for flexible data collection processes in a dynamic and heterogeneous environment, as many different data sources must be considered and requests may be subjected to change even during their processing.

4.2   Data- and User-Driven Processes

As opposed to traditional process management approaches focusing on the sequencing of activities, the case handling paradigm [16] is centralized around the 'case'. Similarly, product-based processes focus on the interconnection between product specification and processes [17]. The Business Artifacts approach [18] is a data driven methodology that is centralized around business artifacts rather than activities. These artifacts hold the information about the current situation and thus determine how the process shall be executed. In particular, all executed activities are tied to the life-cycle of the business artifacts. Another data-driven process approach is provided by CorePro [19], which enables process coordination based on objects and their relations. In particular, it provides a means for generating large process structures out of the object life cycles of connected objects and their interactions. The creation of concepts, methods, and tools for object- and process-aware applications is the goal of the PHILharmonic Flows framework [20]. The framework allows for the flexible integration of business data and business processes overcoming many of the limitations known from activity-centered approaches.

The approaches shown in this sub-section facilitate processes that are more user- or data-centric and aware. The creation of processes from certain objects could be interesting for SustainHub as well. However, in dynamic supply chains, processes rather rely on their context than on objects and are continuously influenced by its changes while executing.

4.3   Dynamic Processes

In literature, there exist two main options for enabling flexibility in automatically supported processes: imperative processes being adaptive or constraint based declarative processes being less rigid by design.

Adaptive PAIS have been developed that incorporate the ability to change a running process instance to conform to a changing situation. Examples of such systems are ADEPT2 [21], Breeze [22], and WASA2 [23]. These mainly allow for manual adaptation carried out by a user. In case an exceptional situation leading to an adaptation occurs more than once, knowledge about the previous changes should be exploited to extend effectiveness and efficiency of the current change [24, 25].

In case humans shall apply the adaptations, approaches like ProCycle [26] and CAKE2 [27] aim at supporting them with respective knowledge. In our context, these approaches are not suitable since the creation as well as adaptation of process instances must incorporate various information from other sources. Furthermore, it must be applied before humans are involved or incorporate knowledge the issuer of a process does not possess. Automated creation and adaptation of the data collection processes will thus be favourable. In this area, only a small number of approaches exist, e.g., AgentWork [28] and SmartPM [29] However, these are limited to rule based detection of exceptions and application of countermeasures.

As aforementioned, another way to introduce flexibility to processes is by specifying them in a declarative way, which does not prescribe a rigid activity sequencing [30]. Instead, a number of declarative rules constraints may be used to specify certain facts the process execution must conform to, e.g., mutual exclusion of activities. Based on this, all activities specified can be executed at any time as long as no constraint is violated. Examples are DECLARE [31] and ALASKA [32]. However, declarative approaches have specific shortcomings concerning understandability [30]. Furthermore and even more important in our context, if no clear activity sequencing is specified, all activities relating to monitoring are difficult to satisfy and monitoring is a crucial requirement for the industry in this case.

5   Data Collection with Adaptive Processes

As shown in Sect. 4, none of the approaches present in related work succeeds in satisfying the complex requirements of a domain like sustainable supply chain communication. Even if they provide facilities for complex processes and dynamic behaviour, they mostly fall short regarding human integration and automation. On account of this, in the SustainHub project, we have started developing a process-aware data collection approach that shall satisfy the requirements elicited (see Sect. 3). In this section, we want to give a rough overview of this approach and what it shall be capable of without going too much into detail.

Based on the comprehensive set of challenges, our approach is introduced in four steps: first, we present the basis for handling data exchange in complex environments. Second, we introduce facilities for automatic configuration and variant

management for data requests (cf. Sect. 5.1). Third, we present concepts for auto-mated runtime variability (cf. Sect. 5.2) and, fourth, data quality and monitoring (cf. Sect. 5.3) support.

To build an information system capable of automatically supporting data col-lection along complex supply chains, the basic requirements we elicited in DCC1 and DCC2 must be covered first. In particular, SustainHub must provide central data request management, assistance in terms of selection and integration of the involved parties, and management of access to the latter. To enable this, our app-roach is based on two things: a comprehensive data model and explicit specifica-tions of data exchange processes.
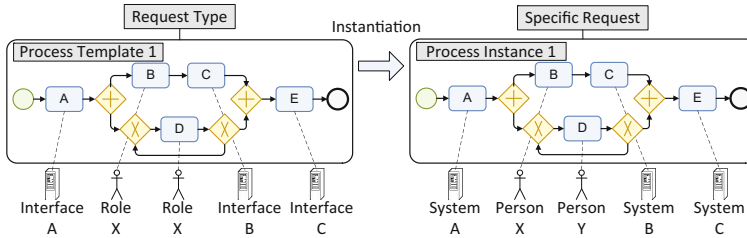
Fig. 5. Processes-based data collection

In our approach, data collection processes are modeled in a Process-Aware Information System (PAIS) integrated into the SustainHub platform providing the domain-related data model. This integration yields a number of advantages: it allows for explicitly specifying the data collection process for one request type through a process template (cf. Fig. 5). Such a request type can be, for example, a sustainability indicator, for which data shall be collected. The process template then governs the activities to be executed at a particular point of time; the activi-ties themselves allow for specifying what exactly is to be done at a particular step of data collection. Further, activities in a process template may be manually exe-cuted by a certain role or may implement an interface to a specific system involved in the data exchange. For a concrete data request relating to a pre-defined request type, a process instance is created to coordinate the data collection process. Via the implemented automatic activities, The PAIS is able to connect to external sys-tems and perform automatic activities concerning the data request. Taking the specified roles in the involved company into account, the PAIS can also automat-ically distribute manual activities to the right persons in charge.

In order to enable an information system to systematically support dynamic data collection processes, it must have access to various kinds of data relating to context, customers, or the collected data. As aforementioned, we integrate a data model uniting different kinds of information that is necessary for manag-ing the data collection. As depicted by Fig. 6, the data model is separated into six sections: first, it comprises customer data like the organizational model of involved
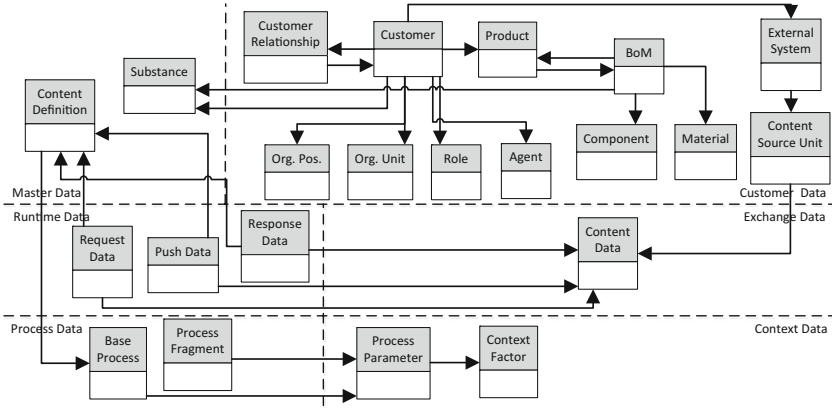
Fig. 6. SustainHub data model

companies, descriptions of their products, BOMs, or systems they employ for sustainability data management (if present). Second, the data model manages a set of master data accessible by all companies connected to the system. This includes, for example, standardized definitions for sustainability indicators or substances widely used by companies in these domains. Third, the data exchange is explicitly managed and stored in the data model by comprising data sets for the data requests, data responses, and, in a separate section, the data collected. Finally, as basis for the advanced features discussed in the following, the data model integrates various data sets covering the data collection processes executed as well as mapping of various contextual influences that may impact the data collection processes during run time.

5.1   Configuration of Data Collection Processes

This section discusses how our approach addresses the challenges DCC3 and DCC4. In particular, it deals with the automated management of data request variants and the meta data leading to the execution of the different variants. Basically, the approach facilitates the automated configuration of pre-defined process templates to match the properties of the given situation. This is enabled by integrating meta data regarding the processes as well as the context of the situation in our data model (cf. Sect. 5). The concrete procedure applied to automated process configuration is shown in Fig. 7.

To incorporate contextual factors influencing the course of the data collection (e.g., if a company executes manual or automated data collection or if, due to a specific regulation, external data validation is necessary), we explicitly model the contextual factors. The latter are processed in a *Context Mapping* component and stored in the data model. In turn, they are utilized in a *Process Configuration* component to determine which process instance may be configured for the current context. In detail, the configuration of data collection processes works as follows:
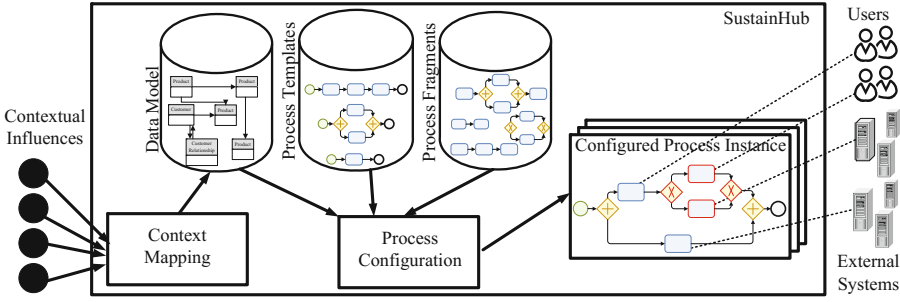
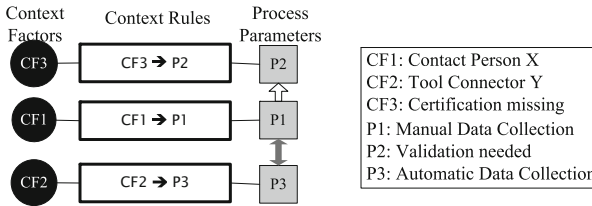**Fig. 7.** Configuration of data collection processes

Fig. 8. Context mapping for configuration of data collection processes

users can specify Process Templates that contain the activities indispensably for a particular data request type. The modeled activities are extended on account of the context factors by *Process Fragments* that may be specified by users as well. In particular, SustainHub selects a set of fragments matching the context of the current situation and automatically integrate them into the process template as illustrated by Fig. 7. After that, a configured process instance is started for the particular data request. In the following, we will exemplarily discuss the context mapping.

As shown in Fig. 8, we distinguish between *Context Factors* and *Process Parameters.* The former capture facts that exist in the environment of Sustain-Hub. As example consider the fact that a company may miss a certain certification necessary to respond to a data request concerning a certain legal regulation. This fact, in turn, may require including additional activities for acquiring the certification. Process Parameters, in turn, capture internal information directly relating to the selection of certain Process Fragments. As the latter do not necessarily correlate with defined Context Factors, we apply a set of configurable *Context Rules* to map Context Factors and Process Parameters. Figure 8 shows a rather simple case. However, complicated cases, where multiple Context Factors relate to one Process Parameter are usual in practice. For example, a company may request a specific four eyes approval procedure in correspondence to different Context Factors: if a certain monetary amount is reached, or the company does not trust the customer, or if the data relates to a certain legal regulation. For a more in depth discussion of this topic, see [33].

5.2   Adaptation of Data Collection Processes

This section discusses our approach for coping with challenge DCC7. In particular, it addresses issues regarding runtime variability. In various situations, it may be required that a data collection process instance has to be changed although the instance is already running. As discussed in DCC7, this could be necessary because of changes to the context or exceptions arising during execution. The first reason constitutes a runtime change to the set of expected situations depicted by the Context Factors. For example, a certification gets invalidated for one company due to a change in a regulation. The second constitutes an error in the execution of the data collection. An example could be that an activity is delayed and exceeds a specific deadline.

Our adaptation approach distinguishes these two cases as depicted by Fig. 9. We apply two different handlings: For erroneous situations a *Compensation Action* is applied to solve the occurred problem or to give users an opportunity to solve the problem on their own. For context changes, a *Context Change Action* is proposed that can influence the set of applied Process Fragments.

In Fig. 9, the different actions SustainHub can perform on account of various dynamic events are illustrated. These are the following:

(1) Various influencing factors dynamically affect SustainHub. Relevant factors are mapped to an internal event.
(2) The type of event determines the way SustainHub addresses the changed situation: a context change induces the change of a Context Parameter whereas an exceptional situation leads to a Compensation Action.
(3) If a Compensation Action is issued, various actions may be governed by it, e.g. resetting a failed activity.
(4) If a Context Parameter changes, the set of integrated Process Fragments will most likely not match the current situation anymore. Therefore, SustainHub estimates whether Process Fragments have to be added, deleted, or replaced.
(5) An issued Context Change action will verify whether an action (e.g., canceling a Process Fragment) is still possible. If not, a corresponding Compensation Action will be created.
(6) A Compensation Action can be used, e.g., to inform the issuer of the data collection process about a failure when adapting to a changing situation.

In order to react to various events and to apply the relating Compensation or Context Change Actions, SustainHub defines a simple event model as illustrated by Fig. 10. An event is composed out of three different parts: (1) a trigger rule that determines, when the event will be fired; (2) the data of the event; (3) an outcome rule governing what action is to be performed due to the event. These three parts are needed for the following reasons: customizable trigger rules enable users to configure what events are important for the data collection process. Further, Fig. 10 shows two examples distinguishing active and passive trigger rules: an event, which contains an active trigger rule is fired due to the change of a certain data set. Instead, an event, which comprises a passive trigger rule is fired by periodic checks, which, e.g., determine, whether a deadline is exceeded.
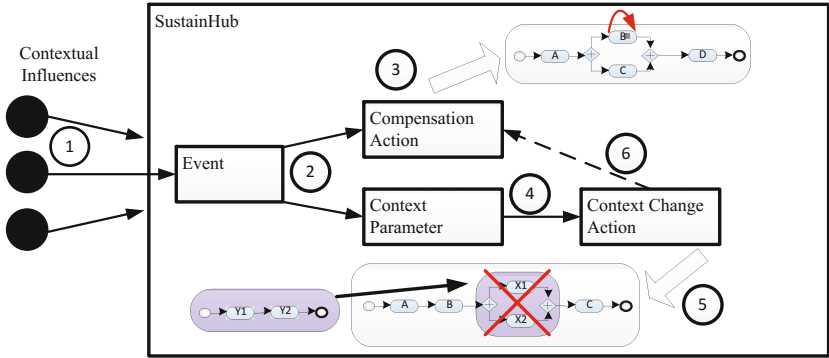
Fig. 9. Adaptation concept of data collection processes

Events can be related to any data or activity in SustainHub. However, not every event necessitates a following action in every situation. Therefore, outcome rules are applied to let users specify, under which circumstances such an action becomes necessary. For example, the introduction of a new regulation may be of utter importance for data collection processes concerning one specific indicator, but have no impact on another one. Finally, the data component stores the information of the event. If an action is carried out based on an event that necessitates human intervention, this information can be delivered to the human.
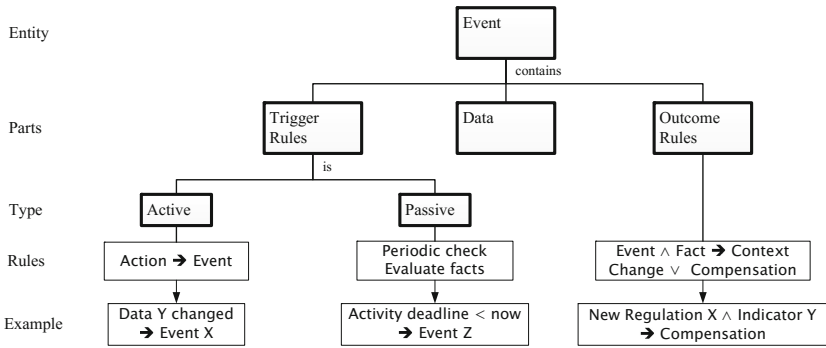

Fig. 10. Event model for adaptation of data collection processes

5.3 Monitoring and Data Quality of Data Collection Processes

This section discusses how our approach addresses the challenges DCC5 and DCC6. In particular, issues relating to incompleteness and quality of data as well as the monitoring of the data collection processes are taken into account. In a complex supply chain, one data request may have dozens of responders. Thus, the

answering time of the request is hardly predictable and some responders might reply with incomplete or low quality data. Our approach, therefore, aims at providing the requester with fine-grained status information about the request and enable SustainHub to handle incomplete data.

As the data collection process is executed in an integrated PAIS, a requester is supposed to perceive request status for basic monitoring. However, this does not suffice for two reasons: first, a request may have an arbitrary number of sub-processes making it cumbersome to check them all. Second, the status of the request might not only depend on activities, but on the transferred data as well. Furthermore, not every activity and data set might have the same importance with respect to the status of a request. Therefore, as first part of our monitoring approach, we introduce a fine-grained, but still comprehensive status object as illustrated in Fig. 11.

Accordingly, a request status is calculated from different activities and data sets involved in the data collection process. These two types of entities can also be annotated with a weight factor to indicate their importance. An example for such a calculation is shown in Fig. 11. In this example, four activities and three data items with varying importance are involved. A particular activity, which gathers data from an IHS, might be very important for the data collection (having a weight = 2) while another one has no importance (a simple administrative task with weight = 0). The values of the weight factor are summed up and combined to indicate the percentage of completeness of the relating request.
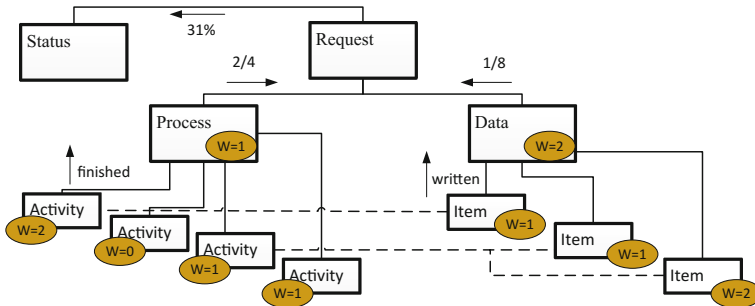

Fig. 11. Status monitoring for adaptive data collection

This extended status is a first improvement for monitoring data collection processes. However, it does not address issues related to incomplete and low quality data. In order to measure such problems and incorporate such meta information into the monitoring process, we apply the following concepts:

– Process and Data Metric: To explicitly specify what is supposed to be measured, we propose a *Process and Data Metric*. The latter may be used for evaluating various facts related to a data collection request. It can be used for various entities and properties, e.g., the status of a process or a SustainHub customer. Furthermore, it may incorporate a mathematical function like a sum or an average. Two examples of metrics are as follows:

Metric X: Average rating of responders who have not yet executed an activity X.

Metric Y: Average precision deviation of responses of a request.

– Dynamic Recalculation. The data collection process and the data it relates to are subject to changes. Therefore, metrics applied to one of these may have to be recalculated frequently. To automate this, we propose a dynamic recalculation defining what has to be done with a particular metric if a change to the data collection process is conducted. It allows for specifying the targeted metric, the trigger for action, and a description. Examples of such actions include full recalculation or discarding the metric.

– Monitoring Annotation. As aforementioned, responders might reply incomplete or not at all. In practice, companies often finish a data collection process without receiving responses from all suppliers as some of them are not even capable of answering properly. Thus, the requester waits until a number of important suppliers has replied and finishes the request based on the available data. To support such advanced data collection behavior, we propose a *Monitoring Annotation*. The latter can be added to a request in order to automatically trigger various actions related to reporting and monitoring. It allows specifying a target entity, a trigger event, and a set of facts (Context Factors or metrics) that will be evaluated when the trigger event is fired to determine, whether the rule will be executed. For the latter, various actions can be defined, ranging from recalculating the metric to canceling the entire data collection request. In the following, we will give two concrete examples of such rules:

Annotation A1: Target: Data Collection Process, Trigger Event: Status > 60 %, Facts: none, Action: Calculate preliminary Results

Annotation A2: Target: Data Collection Process, Trigger Event: Status > 80 %, Facts: Metric X > 80 %, Action: Cancel Request Processing.

The combination of the concepts introduced in this section enables SustainHub to deal with incompletely answered requests. Furthermore, based on the status and the active Monitoring Annotations, the requester can be actively informed about the status of his requests.

6   Conclusion

This paper motivated the topic of sustainability data exchange along supply chains to subsequently present core challenges as well as state of the art in this area. We have identified seven core challenges for today's data collection processes based on intensive interaction with our SustainHub partners most of them relating to variability issues. Especially, both design and run time flexibility are major requirements for any approach supporting sustainable development and production. The presented challenges can serve as starting point for further developments to support today's complicated supply chain communication. The challenges are expressed in terms of sustainability data collection, however they describe generic problems that may occur in many other domains involving cross-organizational communication. Thus the results can be transferred and used in other domains.

There exists a substantial amount of related work in different areas touching these topics. Yet, none of these approaches or tools has succeeded in providing holistic support for the process of sustainability data exchange in a supply chain. The support of data collection requests and processes along today's complex supply chains is a challenge in the literal sense. Nonetheless, the SustainHub project is actively working on a process-based solution to deal with, and successfully manage the high variability occurring during design and run time. Thus, we provide a first outlook on the approach we are developing to tackle the challenges identified in this paper in the future. Future work will describe the exact approach, combination of technologies, and the architecture of the system to systematically address the presented data collection challenges.

References

1. Fawcett, S.E., Osterhaus, P., Magnan, G.M., Brau, J.C., McCarter, M.W.: Information sharing and supply chain performance: the role of connectivity and willingness. Supply Chain Manage. Int. J. **12**(5), 358–368 (2007)
2. Gunasekaran, A., Ngai, E.W.T.: Information systems in supply chain integration and management. Eur. J. Oper. Res. **159**(2), 269–295 (2004)
3. Pramatari, K.: Collaborative supply chain practices and evolving technological approaches. Supply Chain Manage. Int. J. **12**(3), 210–220 (2007)
4. Barnett, P.T., Braddock, D.M., Clarke, A.D., DuPré, D.L., Gimarc, R., Lehr, T.F., Palmer, A., Ramachandran, R., Renyolds, J., Spellman, A.C.: Method of semi-automatic data collection, data analysis, and model generation for the performance analysis of enterprise applications (2007)
5. Singh, R.K., Murty, H.R., Gupta, S.K., Dikshit, A.K.: An overview of sustainability assessment methodologies. Ecol. Ind. **9**(2), 189–212 (2009)
6. Ballou, B., Heitger, D.L., Landes, C.E.: The future of corporate sustainability reporting: a rapidly growing assurance opportunity. J. Account. **202**(6), 65–74 (2006)
7. Adams, C.A., McNicholas, P.: Making a difference: sustainability reporting, accountability and organisational change. Account. Auditing Account. J. **20**(3), 382–402 (2007)
8. Pagell, M., Wu, Z.: Building a more complete theory of sustainable supply chain management using case studies of 10 exemplars. J. Supply Chain Manage. **45**(2), 37–56 (2009)
9. Gottschalk, F., van der Aalst, W.M.P., Jansen-Vullers, M.H., La Rosa, M.: Configurable workflow models. Int. J. Coop. Inf. Syst. **17**(2), 177–221 (2008)
10. Rosemann, M., van der Aalst, W.M.P.: A configurable reference modelling language. Inf. Syst. **32**(1), 1–23 (2005)
11. La Rosa, M., van der Aalst, W.M.P., Dumas, M., ter Hofstede, A.H.M.: Questionnaire-based variability modeling for system configuration. Softw. Syst. Model. **8**(2), 251–274 (2009)

12. Reinhartz-Berger, I., Soffer, P., Sturm, A.: Extending the adaptability of reference models. IEEE Trans. Syst. Man Cybern. Part A **40**(5), 1045–1056 (2010)
13. Hallerbach, A., Bauer, T., Reichert, M.: Configuration and management of process variants. In: vom Brocke, J., Rosemann, M. (eds.) Handbook on Business Process Management I, pp. 237–255. Springer, Heidelberg (2010)
14. Torres, V., Zugal, S., Weber, B., Reichert, M., Ayora, C., Pelechano, V.: A qualitative comparison of approaches supporting business process variability. In: La Rosa, M., Soffer, P. (eds.) BPM Workshops 2012. LNBIP, vol. 132, pp. 560–572. Springer, Heidelberg (2013)
15. Ayora, C., Torres, V., Weber, B., Reichert, M., Pelechano, V.: Vivace: a framework for the systematic evaluation of variability support in process-aware information systems. Information and Software Technology (2014, to appear)
16. van der Aalst, W.M.P., Weske, M., Grünbauer, D.: Case handling: a new paradigm for business process support. Data Knowl. Eng. **53**(2), 129–162 (2004)
17. Reijers, H.A., Liman, S.: Product-based workflow design. Manage. Inf. Syst. **20**(1), 229–262 (2003)
18. Bhattacharya, K., Hull, R., Su, J.: A data-centric design methodology for business processes. In: Cardoso, J., van der Aaalst, W.M.P. (eds.) Handbook of Research on Business Process Management, pp. 503–531. IGI, Hershey (2009)
19. Müller, D., Reichert, M., Herbst, J.: A new paradigm for the enactment and dynamic adaptation of data-driven process structures. In: Bellahsène, Z., Léonard, M. (eds.) CAiSE 2008. LNCS, vol. 5074, pp. 48–63. Springer, Heidelberg (2008)
20. Künzle, V., Reichert, M.: PHILharmonicFlows: towards a framework for object-aware process management. J. Softw. Maint. Evol. Res. Pract. **23**(4), 205–244 (2011)
21. Dadam, P., Reichert, M.: The ADEPT project: a decade of research and development for robust and flexible process support - challenges and achievements. Comput. Sci. - Res. Dev. **23**(2), 81–97 (2009)
22. Sadiq, S., Marjanovic, O., Orlowska, M.: Managing change and time in dynamic workflow processes. Int. J. Coop. Inf. Syst. **9**(1&2), 93–116 (2000)
23. Weske, M.: Formal foundation and conceptual design of dynamic adaptations in a workflow management system. In: Proceedings of Hawaii International Conference on System Sciences (HICSS-34) (2001)
24. Lenz, R., Reichert, M.: IT support for healthcare processes - premises, challenges, perspectives. Data Knowl. Eng. **61**(1), 39–58 (2007)
25. Minor, M., Tartakovski, A., Bergmann, R.: Representation and structure-based similarity assessment for agile workflows. In: Weber, R.O., Richter, M.M. (eds.) ICCBR 2007. LNCS (LNAI), vol. 4626, pp. 224–238. Springer, Heidelberg (2007)
26. Weber, B., Reichert, M., Wild, W., Rinderle-Ma, S.: Providing integrated life cycle support in process-aware information systems. Int. J. Coop. Inf. Syst. **18**(1), 115–165 (2009)
27. Minor, M., Tartakovski, A., Schmalen, D., Bergmann, R.: Agile workflow technology and case-based change reuse for long-term processes. Int. J. Intell. Inf. Technol. **4**(1), 80–98 (2008)
28. Müller, R., Greiner, U., Rahm, E.: AgentWork: a workflow system supporting rule-based workflow adaptation. Data Knowl. Eng. **51**(2), 223–256 (2004)
29. Lerner, B.S., Christov, S., Osterweil, L.J., Bendraou, R., Kannengiesser, U., Wise, A.E.: Exception handling patterns for process modeling. IEEE Trans. Softw. Eng. **36**(2), 162–183 (2010)

30. Zugal, S., Soffer, P., Haisjackl, C., Pinggera, J., Reichert, M., Weber, B.: Investigating expressiveness and understandability of hierarchy in declarative business process models. Softw. Syst. Model. (2013). doi:10.1007/s10270-013-0356-2
31. Pesic, M., Schonenberg, H., van der Aalst, W.M.P.: Declare: full support for loosely-structured processes. In: 11th IEEE International Enterprise Distributed Object Computing Conference, 2007, EDOC 2007, pp. 287–287. IEEE (2007)
32. Weber, B., Pinggera, J., Zugal, S., Wild, W.: Alaska simulator toolset for conducting controlled experiments on process flexibility. In: Soffer, P., Proper, E. (eds.) CAiSE Forum 2010. LNBIP, vol. 72, pp. 205–221. Springer, Heidelberg (2011)
33. Grambow, G., Mundbrod, N., Steller, V., Reichert, M.: Towards process-based composition of activities for collecting data in supply chains. In: 6th Central European Workshop on Services and their Composition (ZEUS 2014), February 2014