TOWARDS A DATA-INTEGRATION APPROACH FOR ENTERPRISE SUSTAINABILITY RISK INFORMATION SYSTEMS

Andreas Thöni[1], Lisa Madlberger[1], Alexander Schatten[2]

[1]Institute of Software Technology and Interactive Systems,
Vienna University of Technology {andreas.thoeni, lisa.madlberger}@tuwien.ac.at
[2]Sophisystems, Austria, alexander.schatten@sophisystems.com

Keywords

Sustainability Risk Management, Sustainability Risk Information System, Enterprise Integration

Abstract

Unsustainable practices of (large) corporations damage the environment, harm society, and can cause severe reputational as well as economic damage to the company. An enterprise sustainability risk information system can help in order to manage respective risks proactively. For the design of such a system, several domain-specific requirements need to be considered such as distinctive source heterogeneity, the need for push & pull information retrieval, the integration of different temporal scopes and the provision of stakeholder-specific data views. This paper evaluates four enterprise integration patterns with regard to their applicability in the sustainability domain. Finally, it proposes an architecture which combines both the message oriented and the shared database approach in order to meet the domain-specific requirements.

1. Introduction

In recent years, sustainability has become a topic that has reached broad attention and many companies have started to consider next to economic also social and environmental factors in their decisions. Numerous companies start to understand, that mid- and long-term economic success is not possible following unsustainable environmental and social practices – both aspects are covered by the term "sustainability" as used in this paper. Awareness of the own practices is important within the company but also in the interaction with customers and other stakeholders. One immediate motivation for companies is to enhance or protect their company reputation. A positive "sustainable" image provides several beneficial effects such as an increased customer base, raised employee motivation or increased attractiveness to potential employees or partners (Hansen & Schrader, 2005, p. 384). On the other side, particularly issues with sustainability can cause severe public damage. Compelling is the case of Nike, which faced an extensive consumer boycott after the New York Times and other newspapers published reports about abusive labor practices at some of its Indonesian suppliers in the early 1990s (Porter & Kramer, 2006).

To ensure prosperity, corporate risk management has to include a thorough assessment of the environmental and social impacts to avoid harmful practices and mitigate risks early on. Although present in visions, missions and goals, most of the time the strategic sustainability considerations are hardly integrated on an operative level (Petrini & Pozzebon, 2009, p. 119). This task is especially challenging, because in a world of competing supply chains, sustainability risks affect not only one organization, but the whole supply chain. Hence, in large companies, a profound risk evaluation likely involves numerous suppliers worldwide, additionally to hundreds of internal company locations (Kogg & Mont, 2012, p. 9). Companies need tools which support them to collect and analyze sustainability data in a fast and efficient way. Therefore, sustainability risk management is seen as an important field for future research (Ghadge, 2012, p. 328).

The level of complexity connected to sustainability risk management is not only influenced by the number of locations and suppliers, but also by the type and quantity of information necessary. Various frameworks define a range of sustainability-relevant factors including economic, social and environmental aspects (e.g. Global Reporting Initiative, 2011). Unlike economic data, environmental and social data is harder to manage automatically. Sustainability information systems need to integrate new forms of automated and semi-automated mechanisms to collect heterogeneous sustainability information.

This paper proposes an architecture for a corporate sustainability risk management system that supports enterprises to establish a consistent, integrated view on environmental and social risks.

From a technical perspective, this challenge can be primarily seen as an enterprise application and data integration task, with the goal to combine heterogeneous data from different sources and application components. In this paper, we will evaluate existing design patterns for enterprise integration tasks according to their applicability in the enterprise sustainability domain and suggest a hybrid architecture, which meets the domain-specific requirements.

2. Related work

Previously, information technology has been an important driver of progress and increase of efficiency for many company functions. However, the role of corporate IT Systems with regard to sustainability was often mentioned only in the context of Green IT, focusing on the reduction of energy consumption of IT systems. More recently, IT has been identified as mean for improved sustainability through its potential to transform information and business processes (The Climate Group, 2008, p. 16)

Sustainability risk management systems have already been addressed in academic literature. Iakovou (2001, p. 25) proposed a decision support system to improve maritime crude transport and the related environmental risks. Peng, Zhang, Tang, & Li (2011; p.316) propose a framework for data integration, data mining and decision support for evaluating the risks and selecting an appropriate alternative during natural and manmade incidents. In non peer-reviewed journals, i.e. in the commercial environment, some risk management systems are available, in a supply chain context. Sedex (2012) allows suppliers to input sustainability data and sourcing deputies to see supplier inputs as well as to determine the basic risks associated with a supplier. However, the integration of other data, especially real-time outside-in data and uncertainty, is limited. The same applies for other software like Maplecroft (2012), allowing country and sector specific risk assessments, and Earthster (2012-project status unclear), which focuses on life cycle assessments including sustainability. Other offerings such as Dun & Bradstreet (D&B, 2012), Thomson Reuters (2013) or Spotter (Spotter Europe, 2012) integrate data retrieved through near real-time mining (e.g. news feeds or social media), but are not specialized in sustainability specific visualization of risk. Generally they are limited in the extent they cover local and site specific risk aspects.

3. Specific requirements for the integration of enterprise sustainability data

This paper, when referring to sustainability data, essentially focuses on data about environmental or social aspects (e.g. information about air quality, labour practises) related to a company's business practises. This specific domain and the related information needs pose additional requirements towards a data integration architecture.

3.1. Source Heterogeneity

As outlined above, sustainability data potentially originates in multiple locations or sources. Internal operational IT Systems, for example Environmental Information, Facility Management or Human Resource Systems, can potentially provide information about energy consumption, or labour practises (Kemper, Baars, & Mehanna, 2009, p. 15). Hardware-sensors, e.g. air quality sensors, can be used to collect environmental measurements. Additionally to internal sources also external sources have to be integrated. Information available on the web, e.g. news or forums, can be automatically evaluated using text-mining methods for the sensing either of public events (Gluchowski, Gabriel, & Dittmar, 2008, p. 110) or reported concerns about certain suppliers and sustainability issues. A further external input channel is given by the possibility to use Linked Open Data sources. Local weather data or open data about country-level indicators, e.g. national corruption index are examples for data retrievable from the Linked Open Data-cloud. Moreover, not all information can be entered automatically. Instead, manual input forms or surveys sent out to multiple stakeholders can deliver valuable information. Recently evolved methods, like "participatory sensing", can represent a further semi-automatic input channel, where stakeholders (e.g. employees or suppliers) can use their mobile phones as an input device to report subjective observations (see e.g. Mun et al., 2009).

As illustrated, one faces a heterogeneous landscape of potential input sources. Furthermore, the system has to account for the fact, that new suppliers might offer further input sources. A SRIS data integration architecture therefore has to be extensible, adaptive and has to provide mechanisms to integrate various heterogeneous sources (RQ 1).

3.2. Push & Pull Data Retrieval

In the sustainability risk domain, information about social or environmental incidents might be critical and demand immediate actions from the management. Data that is actively sent by different sources to the SRIS e.g. by sensors, has to be analyzed on the flow, in order to trigger according mitigation activities. Other systems do not "push" information to the SRIS, but only provide access to certain information e.g. webpages. In this case, the SRIS needs to "pull" the information from the sources. Both ways need to be considered in an SRIS's architecture (RQ2).

3.3. Temporal Scope

Depending on the specific business use case, data from distinct time horizons is required. On the one hand, data has to be persistent to allow long-term sustainability reporting and analysis and on the other hand (near) real-time data is needed to allow for ad-hoc responses and alerts. Therefore, the architecture has to provide mechanisms such that it can provide data in real-time to the business user (RQ3). Furthermore, it has to provide functionality to maintain historic information, and depending on the use case-keep a versioned view on the data (RQ4).

3.4. Update frequency

As some of the described data sources update frequently e.g. air quality sensors, others update in less frequent periods, e.g. a corruption index might be updated just once a year. An SRIS has to account for different update frequencies of various source systems (RQ5).

3.5. Stakeholder-specific data views

A sustainability risk management targets different stakeholders as users. These stakeholders have various other requirements that also affect data integration. First of all, stakeholder needs regarding the scope of to the data (environmental or social sustainability, company or supplier specific, etc.) differ significantly. Second, whereas governments and suppliers will want to be able to access specific company related data (e.g. Hofman, Bastiaansen, van den Berg, & Pruksasri, 2012), management will have an interest to include both, company and supply chain internal as well as external data for analysis. Third, possibly government, but also other stakeholders, could request standardized formats for their data view and an international standardization could be a facilitator (compare with e.g. World Economic Forum, 2012, p. 19). This leads to the need for stakeholder-specific views on subsets of the integrated data base (RQ6).

3.6. Summary

Given these requirements, a SRIS will have to include functional components to integrate push and pull based data with quantitative or qualitative content from internal or external sources. Moreover, different output and analysis components are needed to respect the diverse stakeholder needs including persistence for historical analysis. Finally, data conversion and cleaning has to be implemented to allow for different kinds of sustainability data. Figure 1 presents a basic overview of the general system components.

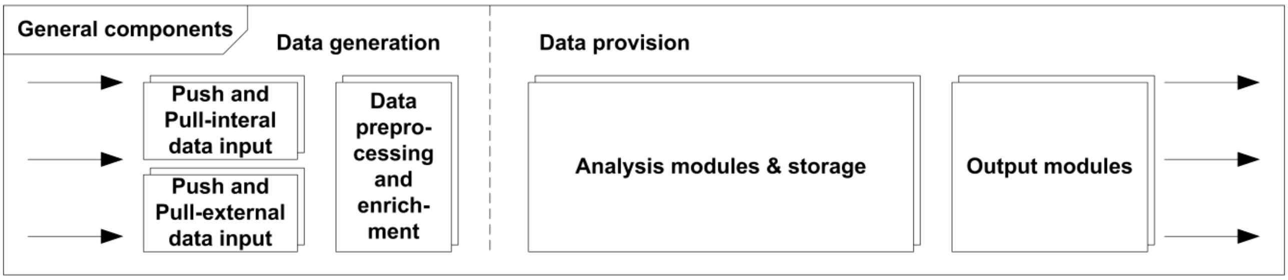

Figure 1: General component overview of sustainability risk information system

Logically, the system can be split into two parts (1) retrieving the data from source systems and processing it for persistence and (2) analyzing the data and finally presenting it to different user circles. These scopes will be referenced below as “Data generation” and “Data provision”.

4. Evaluation of Integration Patterns

A sustainability risk information system has to implicitly account for the requirements described above in its basic architectural approach. Hohpe and Woolf (2003, Chapter 2) have outlined four application integrations variants, which have proved to be successfully applied and can be used as a baseline for the architecture. The following section will evaluate the different approaches with respect to their applicability in the specific problem domain.

4.1. File transfer

File transfer has been a popular integration pattern since decades, but is still used to a large extent today. Key advantage of file sharing is that different applications or application components are loosely coupled as long as they agree on the file's format as well as about the rules on how to create, read, update and delete files (Hohpe & Woolf, 2003, p. 43).

A loose coupling mechanism seems beneficial considering the heterogeneity of the data sources in the sustainability domain. Using common directories, different application components can publish files in a standardized format (e.g. XML) to feed data into the system ("push") or to initiate data collection procedures ("pull"). Since file processing actions, are resource intensive processes, this approach does not seem promising to suit the requirement for real-time information. In general, files are easy to archive, however, as data potentially resides in a large number of files, the risk for arising inconsistencies increases over time. Through loose coupling, different update-frequencies are inherently supported by this integration pattern. The provision of stakeholder-specific views seems hard, since several files have to be read and filtered to obtain specific subsets of the data.

4.2. Shared databases

Shared databases require the existence of a central database managing data related activities, which can be accessed by multiple systems or components. By accessing the same data, it is exchanged "nearly ad-hoc". However, this requires the different system elements to agree on a common data model and respect latencies in dispersed settings (Hohpe & Woolf, 2003, p. 47).

In a SRIS, several input components, which extract data from heterogeneous source systems can store the data in a shared data base. This data storage could also serve as the basis for several analysis and output modules. This requires a common data scheme used by all input and analysis components and resulting conversions. Since the applications are only coupled through a shared database, input components cannot directly call application functions and thereby invoke analysis operations automatically ("push"). As no data transmission between application components is necessary, data can be processed fast and historization can be handled centrally. The direct connection from each component to the database supports different update frequencies. As the data is stored centrally, stakeholder-specific views can be easily realized e.g. using SQL views.

4.3. Remote procedure invocations (RPI)

Remote procedure invocations have been designed to enable system components to remotely invoke processes and retrieve according results. Hence, it is more than a simple data request as it allows for synchronous actions by system elements to reach an overall goal. However, strong coupling with complex settings and multi-step dependencies can cause problems (Hohpe & Woolf, 2003, p. 50).

With regard to a SRIS, analysis modules could be coupled with various input components through RPIs. The heterogeneity of input sources could be abstracted through standardized application interfaces, which keeps the system extensible. Since remote procedures can be invoked both by the input component and analysis components, "Push" as well as "Pull" data retrieval is supported. RPIs are executed synchronously, therefore a high processing speed is possible. However, when integrating different data sources into an analysis, potentially numerous procedures have to be called, which requires vast amounts of data to be transmitted. Data can be historized in every application component separately. Due to the synchronous nature of RPIs, different update frequencies can cause conflicts and slow down the system. Stakeholder-specific data views are possible, but have to be compiled by specific application components.

4.4. Messaging

Messaging is an asynchronous version of integration, resulting in a very loose coupling of applications. Messages transfer data or procedure requests and a messaging system controls the flow of them. Asynchronous communication increases speed on the one hand, but on the other hand also increases complexity (Hohpe & Woolf, 2003, p. 53).

The loose decoupling of input sources can account for the high heterogeneity and flexibility of potential sources for sustainability data. Messages can be invoked both by analysis and by input components, therefore both "push" and "pull" data retrieval is facilitated by this architecture. As messages are not manifested in physical files, data transmission is much faster compared to the file transfer architecture. Since messages allow for asynchronous operations, different update frequencies are well supported by this architecture. The historization as well as the provision of different views has to be explicitly handled by different application components.

4.5. Architectural discussion and conclusion

RQ#	Requirement	File Transfer	Shared Database	RPI	Messaging
RQ1	Heterogeneous input sources	x	-	x	x
RQ2	Push/Pull data retrieval	x	-	x	x
RQ3	Real-time data processing	-	x	x	x
RQ4	Historization	-	x	-	-
RQ5	Varying update frequencies of sources	x	x	-	x
RQ6	Stakeholder-specific data views	-	x	-	-

Table 1: Comparison of design variants with regard to requirements (x...supported -...not inherently supported)

As illustrated in Table 1, no single architectural archetype fulfills all requirements inherently. However, on the one hand, a messaging-based approach appears to meet the most important needs at the "Data generation" side, such as high speed, easy integration of heterogeneous sources and flexibility while maintaining loose coupling and varying update frequencies. On the other hand, a shared database can fulfill the analytical needs posed by the "Data provision" side which requires persistence and different data views. As a result of this evaluation, we will propose a combined approach, which integrates both the message-based and the shared database approach.

5. General Architecture

A messaging architecture can be decomposed into a collection of different patterns, each representing a generic description of one part of the system. Hohpe and Woolf (2003, p. 57) presented an overview of different general patterns, which allow for example, the handling of inbound messages, their structured preprocessing or content enrichment. These patterns build a basic set of building blocks for an initial architectural structuring.

As the shared data approach mainly relates to the "Data provision" – side of the system, a specific type of shared database, namely a Data Warehouse, seems an appropriate design option. A Data

Warehouse (DWH) replicates data from different sources in a central database to allow for extensive analysis and central access to results (Leser & Naumann, 2007, p. 371).

By combining the requirements, respecting the two architectural paradigms of shared database and messaging, we compiled the architecture outlined in Figure 2.

In the proposed architecture, input sources can be coupled with the system using standardized message formats. In this way, new input sources can be added easily, conversion and transformation procedures from heterogeneous schemes and formats into one common scheme are encapsulated in message endpoints of input components (RQ1). Sources "pushing" information into the system, e.g. sensors can be connected through "Channel Adapters", or "Message Translators" to convert external messages into an internal canonical data scheme (RQ2). Based on their content, messages can be routed through the pre-processing and enrichment component, which is responsible for supplementing the extracted data with additional contextual information, e.g. location data. Analytical modules can subscribe a specific message type and thereby receive information in real-time (RQ3). Analytic processing components can send results to the message broker, which can route them to the target output systems for presentation purposes or to trigger further actions, e.g. the posting of notifications. The data warehouse component subscribes all message types and stores all extracted information in a common scheme for long-term analyses (RQ4). The event-based nature of the architecture accounts for different update frequencies of input sources (RQ5). Since the DWH holds all data in a central location, output components can easily access extract subsets of the data to compile stakeholder-specific views (RQ6).

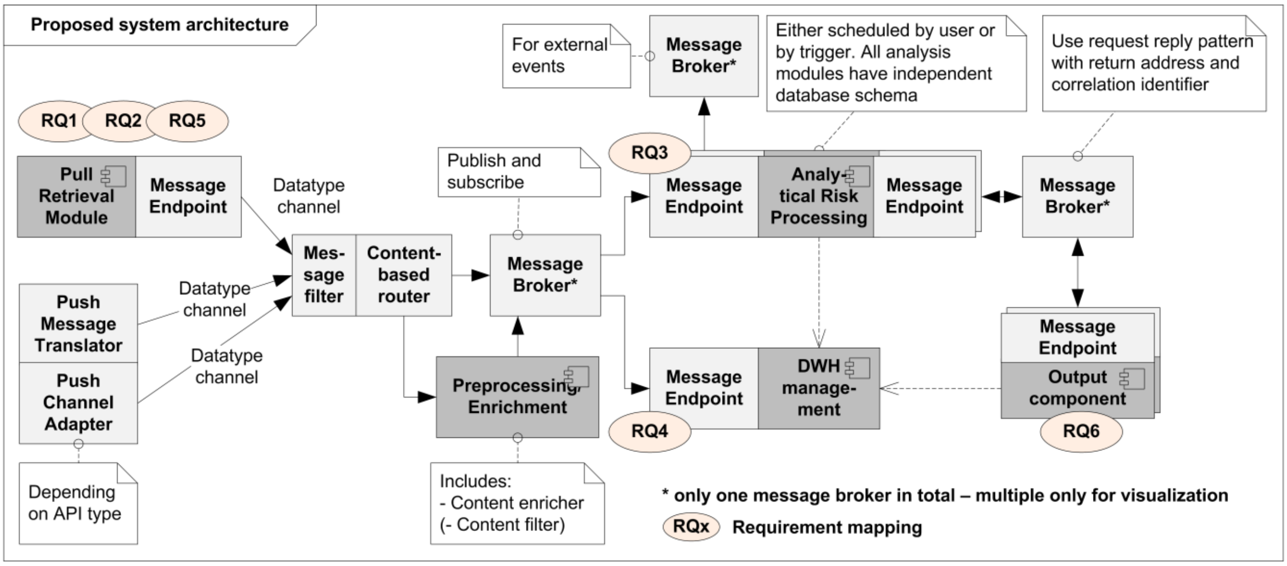

Figure 2: General component overview of a sustainability risk information system

6. Conclusion & Outlook

Sustainability has become an important issue for companies around the globe. An IT-system supporting companies to manage sustainability risk holds specific requirements that need to be matched by its architecture. By evaluating different archetypes, this paper identified a compound of a messaging-based together with a shared-database approach as a promising combination. Finally, we propose an architecture that unifies both variants and fulfils the domain-specific requirements.

In a next step, we will examine mechanisms to overcome the potential semantic heterogeneity of the data extracted from different source systems, e.g. by the use of ontologies and implement the suggested architecture. In further stages, we plan to develop new input modules for sustainability data collection, e.g. based on participatory sensing methods and analytical modules e.g. for the prioritization of social sustainability auditing activities.

Acknowledgements

The 2nd author of this paper is financially supported by the Vienna PhD School of Informatics (http://www.informatik.tuwien.ac.at/teaching/phdschool).

7. References

D&B. (2012). D&B Direct - Supply Management. Retrieved December 11, 2012, from http://www.dnb.com/

Earthster. (2012). Welcome to Earthster. Retrieved May 17, 2013, from http://www.earthster.org/

Ghadge, A. (2012). Supply chain risk management: present and future scope. *International Journal of Logistics Management, The*, *23*, 313–339.

Global Reporting Initiative. (2011). *G3.1 Sustainability Reporting Guidelines*. Amsterdam: Global Reporting Initiative. Retrieved from https://www.globalreporting.org/reporting/latest-guidelines/g3-1-guidelines/Pages/default.aspx

Gluchowski, P., Gabriel, R., & Dittmar, C. (2008). *Management Support Systeme und Business Intelligence. Computergestützte Informationssysteme für Fach- und Führungskräfte* (2., vollst. überarb. Aufl. 2008.). Springer.

Hansen, U., & Schrader, U. (2005). Corporate social responsibility als aktuelles Thema der Betriebswirtschaftslehre. *BETRIEBSWIRTSCHAFT-STUTTGART-*, *65*, 373.

Hofman, W., Bastiaansen, H., van den Berg, J., & Pruksasri, P. (2012). A platform for secure, safe, and sustainable logistics. Retrieved from http://www.efreightconference.com/uploadfiles/papers/efreight2012_submission_5.pdf

Hohpe, G., & Woolf, B. (2003). *Enterprise Integration Patterns: Designing, Building, and Deploying Messaging Solutions* (1st ed.). Addison-Wesley Professional.

Iakovou, E. T. (2001). An interactive multiobjective model for the strategic maritime transportation of petroleum products: risk analysis and routing. *Safety Science*, *39*, 19–29.

Kemper, H.-G., Baars, H., & Mehanna, W. (2009). *Business Intelligence : Grundlagen und praktische Anwendungen: Eine Einführung in die IT-basierte Managementunterstützung*. Vieweg +Teubner.

Kogg, B., & Mont, O. (2012). Environmental and social responsibility in supply chains: The practise of choice and inter-organisational management. *Ecological Economics*, *83*, 154–163.

Maplecroft. (2012). Sector risk | agro communities. Retrieved December 9, 2012, from http://maplecroft.com/

Mun, M., Reddy, S., Shilton, K., Yau, N., Burke, J., Estrin, D., Boda, P. (2009). PEIR, the personal environmental impact report, as a platform for participatory sensing systems research. In *Proceedings of the 7th international conference on Mobile systems, applications, and services*. New York, NY, USA: ACM. 55-68.

Peng, Y., Zhang, Y., Tang, Y., & Li, S. (2011). An incident information management framework based on data integration, data mining, and multi-criteria decision making. *Decision Support Systems*, *51*, 316–327.

Petrini, M., & Pozzebon, M. (2009). Managing sustainability with the support of business intelligence: Integrating socio-environmental indicators and organisational context. *The Journal of Strategic Information Systems*, *18*, 178–191.

Porter, M. E., & Kramer, M. R. (2006). The link between competitive advantage and corporate social responsibility. *Harvard Business Review*, *11*.

Sedex. (2012). Enabling ethical and responsible supply chains. Retrieved December 9, 2012, from http://www.sedexglobal.com/member-services/risk-assessment/

Spotter Europe. (2012). Risk Assessment. Retrieved December 11, 2012, from http://www.spotter-europe.com/

The Climate Group. (2008). *SMART 2020: Enabling the low carbon economy in the information age*. Retrieved from www.smart2020.org/_assets/files/02_Smart2020Report.pdf

Thomson Reuters. (2013). Thomson Reuters | News Analytics | Financial. Retrieved February 11, 2013, from http://thomsonreuters.com/products_services/financial/financial_products/a-z/news_analytics/

World Economic Forum. (2012). *New Models for Addressing Supply Chain and Transport Risk*. Cologny/Geneva: World Economic Forum. Retrieved from http://www3.weforum.org/