

## Data-Centric Risk Management for Business Processes

Markus Michael Mützel  
 Evonik Industries AG  
 Germany  
[markus.muettel@evonik.com](mailto:markus.muettel@evonik.com)

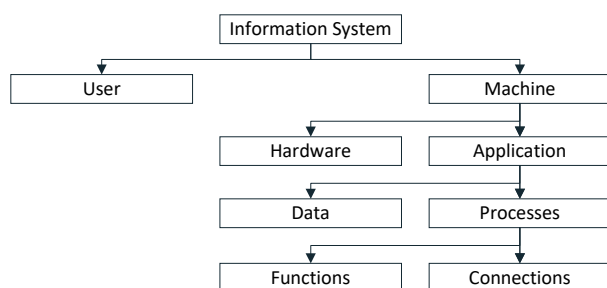
Omid Tafreschi  
 Darmstadt University of Applied Sciences  
 Germany  
[omid.tafreschi@h-da.de](mailto:omid.tafreschi@h-da.de)

### Abstract

*Data quality is a key factor enabling the digital transformation, as automated business processes demand it for value creation. This paper explains the relationship between data quality and business processes. In this context, the paper presents two Key Performance Indicators (KPIs) to analyze this relationship and to prioritize actions to improve data quality. The first KPI quantifies the risk of a single process belonging to a process hierarchy based on the quality of its data. The second KPI considers the risk for all analyzed processes based on the quality of a specific data type. Based on these results, the paper presents an approach for prioritizing measures to improve data quality. The applicability of the KPIs is demonstrated with a real application.*

### 1. Introduction

According to Legner et al., the digital transformation is affecting all areas of private and professional life [1]. Despite the advents of the information and communication technology, such as autonomous systems or peer-to-peer applications, its core is still the processing of data as depicted by Figure 1.



**Figure 1. Generic Model of a Information System Proposed by Krcmar [2]**

With this figure, Krcmar [2] shows the relation of

data and processes as parts of an application. The quality of data has a direct impact on applications and their processes as reported by Hildebrand et al. [3]. According to Ge and Helfert [4], data quality problems often lead to high costs. Analogously, the case study conducted by Hüner et al. [5] also confirms the negative effects of poor data quality on cross-company collaborations. We explain the extent of poor data quality in the Section 4 using practical examples from the purchase-to-pay process. Among other things, the importance of the quality of supplier data and the resulting process risks, e.g. delayed delivery or low net working capital, are explained.

Otto and Österle [6] provide an overview on numerous approaches for solving the mentioned data quality problems and classifies them into the distinct categories preventive and reactive approaches. Preventive approaches aim at avoiding data quality problems while generating data. However, these approaches cannot completely solve the problem, as process changes and technological developments can create new requirements for the quality of existing data. Therefore, reactive approaches are still necessary to solve quality problems of existing data. According to Otto and Österle [6], these approaches need an efficient concept for assessing data quality based on a process-oriented method to analyze the real impact of data quality from a company's perspective. This paper presents the mentioned concept to answer following questions:

1. How critical is data quality to the risk of a process?
2. How must data improvement be prioritized so that the risks of the affected processes are efficiently reduced?

To answer these questions, the paper presents two key performance indicators (KPIs). The first KPI reflects the risk of a certain process due to the quality of all data used in that process. In other words, the first KPI is process specific. In contrast to that, the second KPI

is data specific. It quantifies the impact of quality of certain data on all processes using that data.

The paper's contribution includes four artifacts. These are an extended data model for representing the ingredients of a business process including the aspect of data quality, two processes for calculating the mentioned KPIs, and a matrix for those application.

This paper is structured as follows: After this introduction, Section 2 describes the relation between data and processes, the aspect of data quality, and its impact on process performance. Section 3 presents the calculation of the KPIs which are evaluated in Section 4. We discuss related work in Section 5. Section 6 provides the conclusions and future work.

## 2. Data-oriented Process Management

According to the International Organization for Standardization (ISO) [7], "Understanding and managing interrelated processes as a system contributes to the organization's effectiveness and efficiency in achieving its intended results." ISO [8] defines a process as a "set of interrelated or interacting activities that use inputs to deliver an intended result." A more sophisticated view on business processes is presented by Dumas et al. [9] which is depicted by Figure 2. It shows the ingredients of a business processes from a generic perspective. The logical flow is determined with the events, activities and decisions points. The value is represented by the outcome for the customer and the input parameters are the objects.

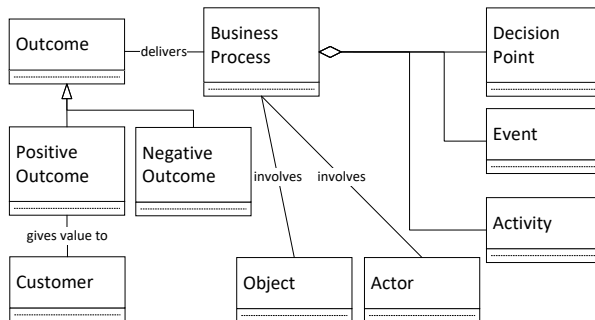


Figure 2. The Ingredients of a Business Process Proposed by Dumas et al. [9]

In order to analyze the relationship between a business process and the quality of its data in more detail, we have extended the model proposed by Dumas [9]. It is depicted by Figure 3.

The classes and attributes in italics are extensions and are explained in the following:

**Data Type** is a subtype of object and reflects data that can be used by the activities of a business process.

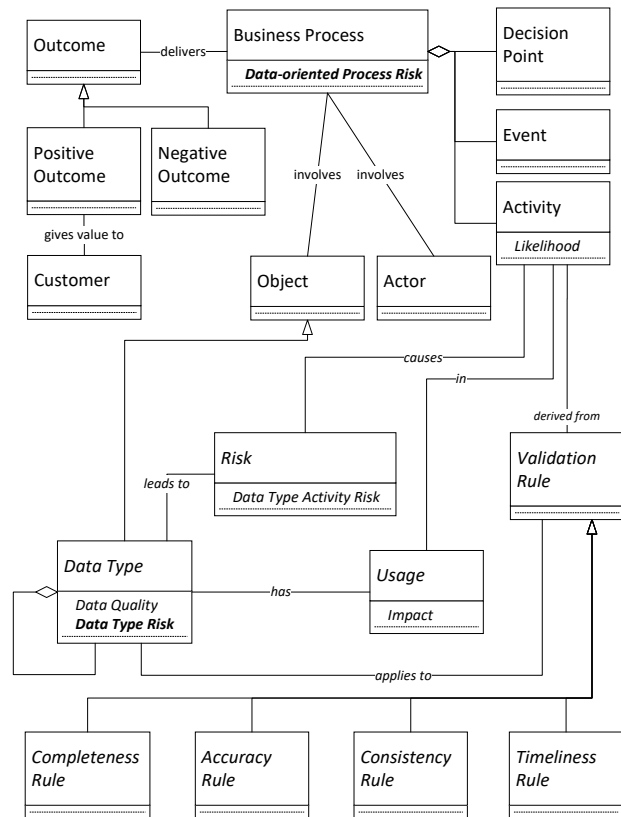


Figure 3. Data Risk-oriented Extension of a Business Process

Data type can be understood as a template for a set of data objects used by business processes. Each data object contains values for the set of attributes according to its data type. Data type can contain other data types. For example the data type supplier contains the address of the supplier, which consists of single attributes like a postal code. In addition to these attributes, we define two attributes, i.e., data type risk and data quality. The first is a KPI reflecting the risk caused by a specific data type for all activities using that data type. The calculation of the KPI is presented in Section 3.2. The latter represents the current quality of the data type based on validation rules. Strong et al. [10] define "high-quality data as data that is fit for use by data consumers." Data consumers use data in context of processes. Therefore, we consider data quality from the process perspective.

**Validation Rule** According to Hüner et al. [5], a validation rule checks whether data objects belonging to a data type fulfill characteristics defined by the rule. A validation rule has four

subtypes for checking the different dimensions of data quality, i.e., consistency, accuracy, timeliness, and completeness. These four reflect the common dimensions of data quality as reported by Batini et al. [11]. Validation rules are required to determine the value for the attribute data quality. This value can range between 0 (minimum data quality) and 1 (maximum data quality) which is the result if all existing data objects fulfill all requirements of corresponding rules. In other words, the data is fit for usage in the related processes. An example for the application of the validation rules would be checking the completeness of the postal codes of suppliers. If all data records of all suppliers had a postal code, the quality value 1 would be delivered as a result of applying the rule. Further examples are discussed by Hüner et al. [5].

**Usage** reflects the aforementioned usage of a data type by an activity and is required, since a specific data type can be used by different activities, while one activity can use different data types. For example, the address of a supplier is used by an activity in the ordering process to send an order to the supplier. The attribute impact quantifies the possible negative influence of data with poor quality on the activities concerned. This is estimated by experts based on their experience. The value range for this attribute is between 0 (no impact) and 1 (maximum impact). An example of a small impact is the wrong title. In contrast, incorrect payment terms or bank details have monetary effects.

**Risk** is comparable to usage, since it also depends on a data type and activity. However, due to the semantic difference between usage and risk, a distinction is necessary. Risk has one attribute, i.e., Data Type Activity Risk (DTAR), which considers the likelihood of the corresponding activity, the quality of the corresponding data type and the possible impact of its usage. For example, the risk of a digital ordering process due to incorrect postal addresses of suppliers is lower than the risk of a postal ordering process. The calculation of DTAR is explained in Section 3.1.

**Likelihood** is an attribute of an activity and represents the probability of performing that activity during a business process. This can be quantified by means of process mining as proposed by van der Aalst [12]. The value range for this attribute is between 0 (no usage of the activity in the process) and 1 (the activity is used in all process runs).

**Data-oriented Process Risk** is an attribute of a business process and another KPI representing the risk of a certain business process caused by the quality of data required in that process. The calculation of the KPI is presented in the next chapter.

### 3. Key Performance Indicators

We propose two KPIs in order to quantify the risk of a business process caused by data quality. The first KPI, i.e., Data-oriented Process Risk, is specific for a single process and reflects its risk caused by the quality of data used by its activities. The second KPI, i.e., Data Type Risk, provides an holistic view to all risks caused by a single data type with regard to all processes using that data type. We describe the calculation of both KPIs using Business Process Model and Notation (BPMN) defined by the Object Management Group [13] in the following.

#### 3.1. Data-oriented Process Risk

Strong et al. [10] define “high-quality data as data that is fit for use by data consumers.” Business processes can be understood as data consumers. Along this understanding, we propose a KPI for assessing the aspect fit for use. Specifically, the KPI determines the fitness of data based on the requirements of the business processes. The approach is to analyze the data used by a process to evaluate the process risk caused by data quality. According to Panayiotou et al. [14], it is reasonable to assess a risk by its impact and likelihood of occurrence. Therefore, we consider all activities of a process with regard to the required data and the likelihood of occurrence. In addition, we examine the potential impact of usage of data with poor quality in the activities. The calculation of the KPI is depicted by Figure 4. This process requires following four inputs:

**Process Model** reflecting a flow of connected activities and decision points resulting in a value generation. According to Reijers and Mendling [15], process modularization is a proven concept for managing complexity of business processes and increasing their efficiency. Therefore, we consider process hierarchies while calculating the KPI. With respect to BPMN [13], the process hierarchy is supported by means of sub-processes.

**Event Logs** include cases which consist of events as described by van der Aalst [12]. A case is a possible flow of a process. Event logs are

generated during the execution of processes and can be understood as traces.

**Expert Estimations** include the negative impact that can arise from poor data quality in execution activities. These are assessed by experts, e.g. process owners.

**Data Quality Scores** are results of the application of validation rules to existing data. According to Zaveri and Rula [16], these rules target at assessing the quality of existing data within a company and are derived from use cases, i.e., the activities processing the data to be checked.

The first input is the process model, which can contain a process hierarchy. After the start, indicated by the start event KPI Needed, the first activity of the top-level process is determined and analyzed. The result of this analysis can be either a Task or a Sub-Process. According to BPMN [13], a task reflects an atomic activity and a sub-process is a process that is included within another process.

In case of a task, the required Data Types for the corresponding activity are determined based on the considered process model. Then, for each data type its risk for the activity in question is calculated. For this, we multiply the likelihood derived from event logs and the impact assessed by experts. We refer to the result as Data Type Activity Risk (DTAR), which is calculated using the formula

$$DTAR_{DataType} = Likelihood_{DataType} \times Impact_{DataType} \quad (1)$$

In case of a sub-process, a recursive process call is performed. In other words, we calculate the same KPI for the process located one level below. This is repeated until the next analyzed activity is a task. After the recursive process call, the results are used to calculate the DTAR for the activity reflecting the sub-process with the help of the formula

$$DTAR_{DataType} = Likelihood_{DataType} \times \sum_{i=1}^n DTAR_i \quad (2)$$

$n$  denotes the number of all DTARs of the sub-process. In both cases, the result, i.e., DTAR, is stored for later calculations. Next, we analyze whether the process model has other activities. In case of other activities, the process repeats with the next activity. After all activities are analyzed, all DTARs of the

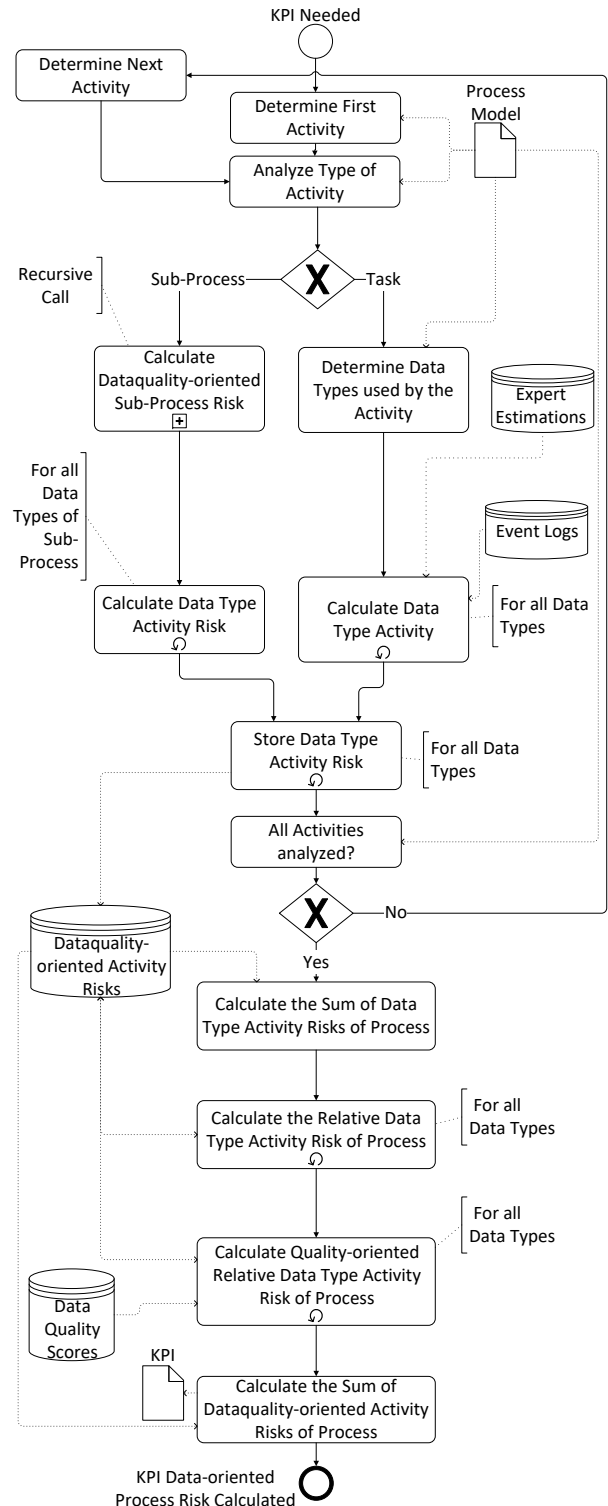


Figure 4. Calculation of KPI - Data-oriented Process Risk

considered process are summed up. Afterwards, to enable comparability this sum is used to calculate the Relative Data Type Activity Risks (RDTARs) of the considered process model for all data types using the formula

$$RDTAR_{Data\ Type} = \frac{DTAR_{Data\ Type}}{\sum_{i=1}^n DTAR_i} \quad (3)$$

$n$  denotes the number of all DTARs. During the next step, we use data quality scores to calculate the Quality-oriented Relative Data Type Activity Risks (QRDTAR) for all data types of the considered process model using the formula

$$QRDTAR_{Data\ Type} = RDTAR_{Data\ Type} \times DataQuality_{Data\ Type} \quad (4)$$

This is required to incorporate the aspect of data quality into the KPI. The results are stored in the data source dataquality-oriented activity risks for later calculations. This data source is a table containing the attributes activity, data type, DTAR, RDTAR, and QRDTAR. The last step of calculation produces the KPI, i.e., the sum of QRDTARs. We refer to this KPI as Data-oriented Process Risk.

This KPI is a dataquality-based risk assessment for a specific process which can be part of a process hierarchy. The KPI quantifies the fitness for use of existing data for processes requiring that data.

### 3.2. Data Type Risk

Since a data type can be used by different independent processes, a holistic analysis of the data type is necessary to prioritize a data quality improvement in a systematic way. This prioritization is necessary in order to develop a comprehensive understanding of the relevance of a data type for the processes under consideration and to derive appropriate measures, e.g., to carry out data cleansing. For this purpose and in order to complement the previous KPI, we propose a second KPI which focuses on a specific data type rather than on a specific process model.

The aforementioned KPI calculation process delivers intermediates results reflecting the risk associated with data types and according activities along the process hierarchy. In order to calculate holistic and concise KPIs, we consider the process at the highest hierarchy level, i.e., the top process. This approach is sound, as all process levels are considered due to the recursion mentioned above. The calculation process is depicted by Figure 5. The process consists of two activities. The first

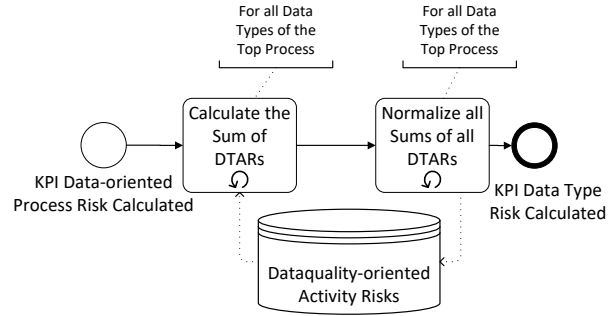


Figure 5. Calculation of KPI - Data Type Risk

activity calculates the sum of all DTARs of all data types along the top process. For this, the data produced during the calculation of the first KPI is used. The result of the first activity, i.e., the sums, are normalized in order to ensure their comparability. The normalization is based on the following proven formula:

$$Normalized\ Sum\ of\ DTAR = \frac{Sum\ of\ DTAR - DTAR_{min}}{DTAR_{max} - DTAR_{min}} \quad (5)$$

$DTAR_{min}$  represent a data type without any risks for the activities. This is either the result of no impact of that data type, e.g., vendor search terms, or no usage of the data type. Therefore, we can assume that the value of  $DTAR_{min}$  equals 0 and can simplify the formula as follows:

$$Normalized\ Sum\ of\ DTAR = \frac{Sum\ of\ DTAR}{DTAR_{max}} \quad (6)$$

## 4. Evaluation

In order to evaluate the proposed KPIs, we deployed them for an existing business process hierarchy of a large European company in the chemical industry. The analyzed business process hierarchy includes a set of typical processes belonging to the domain purchase-to-pay and is depicted by Figure 6. The process consists of seven activities. One of them, i.e., *pre-purchase required?*, is a task, whereas the remaining six activities are sub-processes. In order to analyze the relevance of data we examine the sub-process *purchase* in detail. It is depicted by Figure 7. For the sake of simplicity and without constraining generality, we focus on master data in the analyzed process. With respect to the extended generic model of a business process depicted by Figure 3, the process model depicted by Figure 7 includes several data usages, i.e., data flows

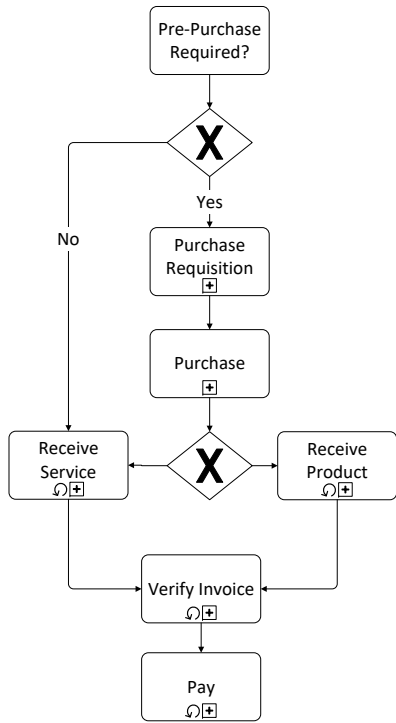


Figure 6. Overall Purchase-To-Pay Process

to activities. E.g., the master data *supplier* is required for the first activity *assign source of supply*. This is represented by a data flow to the mentioned activity. As mentioned above, a risk is linked to both a certain activity and a certain data type. We discuss some risks than can be caused by the analyzed process model.

1. Delivery delays because of incomplete contact data of suppliers used by the activity *send purchase order by mail*
2. Disclosure of confidential data to unauthorized parties because of incorrect contact data of suppliers used by the activity *send purchase order by mail*
3. Financial loses because of outdated payment terms of suppliers used by the activity *assign source of supply*
4. Communication overhead because of inaccurate general ledger account used by the activity *create service purchase order*
5. Unusable surplus products because of incorrect units of products used by the activity *create product purchase order*

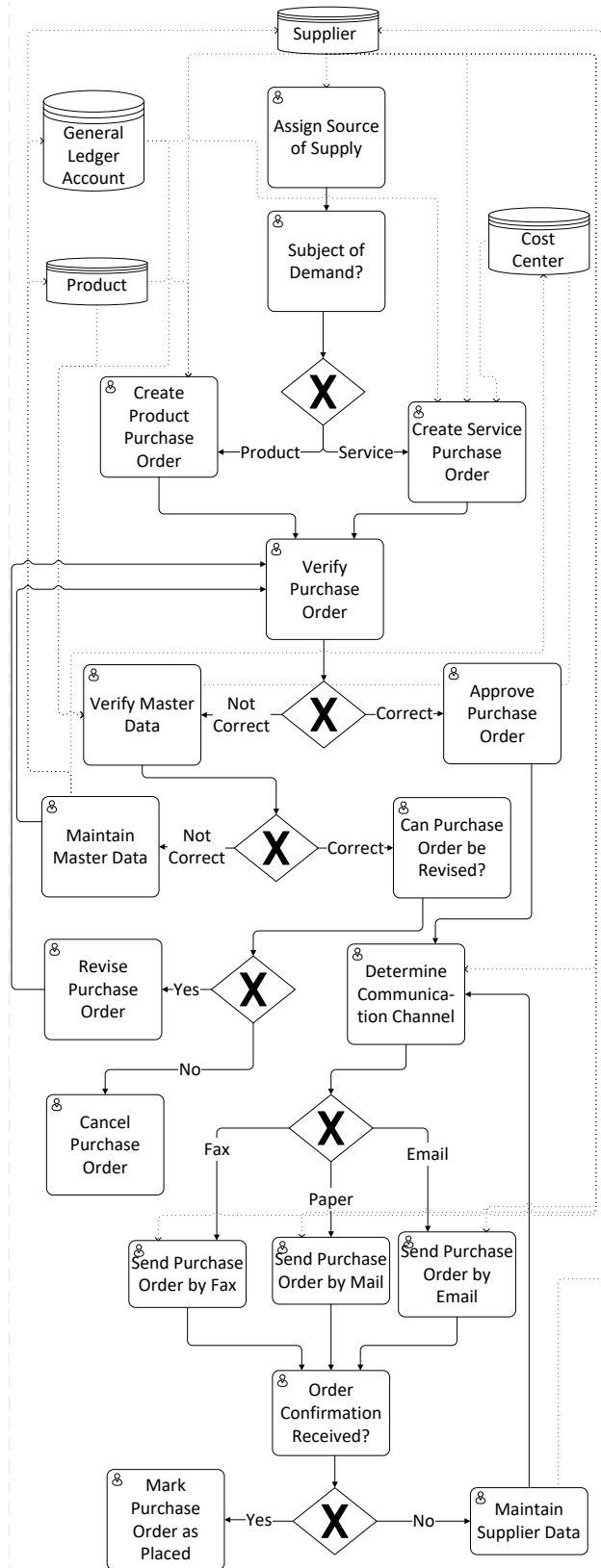


Figure 7. The Purchase Sub-Process

6. Legal consequences because of incorrect incoterms of suppliers used by the activity *assign source of supply*
7. Non-compliant reporting because of inconsistent supplier data, i.e., duplicates, used by the activity *assign source of supply*
8. Delays in production because of inappropriate purity or packaging of product used by the activity *create product purchase order*
9. Safety issues because of missing safety regulations of products using by the activity *receive product*

Please note that the presented risks are interdependent, i.e., they influence and overlap each other. In the following we describe the calculation of the proposed KPIs based on the presented process models depicted by Figures 6 and 7.

#### 4.1. Data-oriented Process Risk

The process model depicted by Figure 6 is used as input for the calculation of the first KPI, i.e., Data-oriented Process Risk. Since the first activity, i.e., *pre-Purchase required?*, is a task, data types used by the activity are determined. Since the result is empty, the types of the following activities are analyzed. Following this logic, the calculation process analyzes the activity *purchase*. This process step would lead to the calculation of the KPI, which is a recursive process execution following the process hierarchy. In other words, the sub-process depicted by Figure 7 is used as input for the recursive KPI calculation. The first analyzed activity is *assign source of supply*. Since this activity is a task, data types used by the activity are determined. The result contains the supplier master data for which DTAR, i.e., the product of likelihood and impact, is calculated. The result is stored and the next activities are analyzed. In case of the activity *create product purchase order* two data types, i.e., supplier and product master data, are considered. For each of them a specific DTAR is calculated.

Once the last activity of the sub-process *purchase*, i.e., *mark purchase order as placed*, is analyzed, the sum of all DTARs is calculated which is used in the following step to calculate the RDTARs for all data types used by the sub-process *purchase*. The result is used by the next activity to calculate QRDTARs for all data types used by the sub-process which requires data quality scores. These are measured using the validation rules.

Summing the QRDTARs leads to the first KPI for the sub-process *purchase* depicted by Figure 7. This is

the end of the recursive call which is executed for all following sub-processes of the overall process depicted by 6. Table 1 shows some results of the calculation. The highest process risk based on the KPIs exists for the process pay. This risk takes into account the current quality of the data used, its usage frequency and the estimated impact.

**Table 1. First KPI Representing Data-oriented Process Risks**

Process	Data-oriented Process Risk
Purchase-To-Pay	0.76
Purchase	0.84
Verify invoice	0.95
Pay	0.62

The process for the calculation of the first KPIs leads to some intermediate results which are necessary for the second KPIs. For this purpose, the intermediate results are stored in the data source dataquality-oriented activity risks. Some example results are shown by Table 2. Please note that the process names are given to improve the clarity.

#### 4.2. Data Type Risk

The second KPI is calculated based on the intermediate calculation results shown in Table 2. The calculation process is depicted by Figure 5 and starts with summing up all DTARs of a specific data type used by the top process. This task is repeated for all data types. Then, the results of the previous task are normalized using the Formula 6. Results reflect the second KPIs and are stored for further analysis. Some of them are shown in the Table 3.

The data type payment method of supplier constitutes the highest risk for all analyzed processes. Due to the current poor quality of the data type mentioned, it will have a critical effect on the affected processes. In order to better evaluate this correlation and to derive recommendations for action, such as data cleansing, a method for applying the second key figure is required. For this purpose, we propose a two-dimensional matrix as depicted by Figure 8.

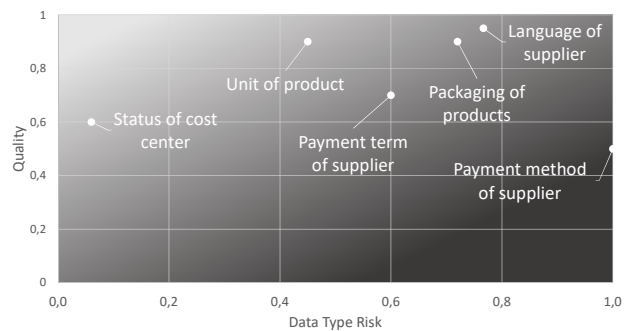
The matrix allows a segmentation of the data types based on their risk assessment and their quality. Thus, targeted actions for data management can be derived and prioritized. In the scenario considered, the data type status of cost center is to be treated with low priority despite its medium quality due to the low risk. In comparison, the data type payment term of supplier is to be treated with higher priority due to the higher risk.

**Table 2. Calculation Results for the Process Purchase-To-Pay**

Process	Activity	Data Type	DTAR	R DTAR	QR DTAR
Purchase	Create service purchase order	Payment term of supplier	0.12	0.08	0.06
Purchase	Create product purchase order	Payment term of supplier	0.28	0.20	0.14
Purchase	Create product purchase order	Unit of product	0.30	0.21	0.19
Purchase	Send purchase order	Language key of supplier	0.20	0.14	0.13
Purchase	Create service purchase order	Status of cost center	0.04	0.03	0.02
Purchase	Create product purchase order	Packing of product	0.48	0.34	0.30
Verify invoice	Create rejection letter	Language key of supplier	0.14	1.00	0.95
Pay	Check payment method	Payment method of supplier	0.60	0.68	0.34
Pay	Send payment confirmation	Language key of supplier	0.14	0.16	0.15
Pay	Send payment confirmation	Contact data of supplier	0.14	0.16	0.13
Purchase -To-Pay	Purchase	Payment term of supplier	0.36	0.16	0.11
Purchase -To-Pay	Purchase	Unit of product	0.27	0.12	0.11
Purchase -To-Pay	Purchase	Language key of supplier	0.18	0.08	0.07
Purchase -To-Pay	Purchase	Status of cost center	0.04	0.02	0.01
Purchase -To-Pay	Purchase	Packaging of product	0.43	0.19	0.17
Purchase -To-Pay	Verify invoice	Language key of supplier	0.14	0.06	0.05
Purchase -To-Pay	Pay	Payment method of supplier	0.60	0.26	0.13
Purchase -To-Pay	Pay	Language key of supplier	0.14	0.06	0.06
Purchase -To-Pay	Pay	Contact data of supplier	0.14	0.06	0.05

**Table 3. Second KPI Representing Data Type Risks**

Data Type	Data Type Risk	Data Quality Score
Payment term of supplier	0.60	0.70
Unit of product	0.45	0.90
Language key of supplier	0.77	0.95
Status of cost center	0.06	0.60
Packaging of product	0.72	0.90
Payment method of supplier	1.00	0.50
Contact data of supplier	0.23	0.80



**Figure 8. Matrix Showing the Correlation of Data Type Risk and Data Quality**

## 5. Related Work

Data quality has often been addressed in research. Batini et al. [11] provide a comprehensive overview of existing approaches and distinguishes between data-driven and process-driven strategies for improvement of data quality. The methodology presented in this paper is data-driven but follows a process-oriented approach. In contrast to our work, the following papers are process-driven.

Ofner et al. [17] present a data quality oriented process modeling approach. Similar to our paper, the authors provide sound arguments for joint analysis of data quality and business processes. However, they focus on the design of business processes with respect to data quality.

Similar to the previous mentioned work, Rodríguez et al. [18] present an approach for including data quality requirements in business process modeling. For



this, the authors propose an extension to BPMN to specify data quality requirements in business processes. Similar to our work, they consider the probability of execution of different activities and the impact of data usage. However, their contribution does not analyze the quality of existing data. In another paper, the authors expand their approach and present a methodology for selecting measures to improve data quality during process modeling [19]. However, the paper does not consider the quality of existing data.

Hüner et al. [5] present a case study analyzing product data quality in supply chains. A result of that case study includes data clusters, which are critical in the concrete case. These are similar to the risks we described above. A further comparability is based on the proposed metrics, which are also calculated considering frequency and impact. However, in contrast to our paper, the metrics are calculated without considering business processes and process hierarchies.

Glowalla and Sunyaev [20] provide another process-driven approach to integrate data quality into existing process models. Compared to our contribution, the holistic view is missing, which allows the analysis of process hierarchies and the evaluation of single data types.

## 6. Conclusions and Outlook

This paper explains the relationship between data quality and business processes and presents two KPIs to analyze this relationship and prioritize actions to improve data quality. Simplified, the paper answers the question: Which data must be treated with priority based on its quality and usage? As a result of our evaluation, we found out that the quality of the process model analyzed, the availability of event-logs, and the subjectivity of impacts estimated by experts have an implication on the KPIs. Analyzing those factors and their implications is the next research step in addition to extending the results to derive concrete actions. In this context, the usage of the KPIs for automatic data improvement by machines is a requirement. Furthermore, we will evaluate the effort required for actions, such as data cleansing or consolidation, in order to identify efficient actions. Simplified, the following research step must answer the question: Which actions are efficient for the prioritized data? In addition, our further research will address the question: How to analyze and interpret variations of KPIs over time? The aim is to identify the causes of the data quality problems.

## References

- [1] C. Legner, T. Eymann, T. Hess, C. Matt, T. Böhmman, P. Drews, A. Mädche, N. Urbach, and F. Ahlemann, "Digitalization: Opportunity and challenge for the business and information systems engineering community," *Business & Information Systems Engineering*, vol. 59, pp. 301–308, Aug 2017.
- [2] H. Krcmar, *Informationsmanagement*. Springer, 6 ed., 2015.
- [3] K. Hildebrand, M. Gebauer, H. Hinrichs, and M. Mielke, *Daten- und Informationsqualität Auf dem Weg zur Information Excellence*. Springer Vieweg, 2018.
- [4] M. Ge and M. Helfert, *Handbook of Data Quality*, ch. Cost and Value Management for Data Quality, pp. 75–92. Berlin, Heidelberg: Springer Berlin Heidelberg, 2013.
- [5] K. M. Hüner, A. Schierning, B. Otto, and H. Österle, "Product data quality in supply chains: the case of beiersdorf," *Electronic Markets*, vol. 21, p. 141, May 2011.
- [6] B. Otto and H. Österle, *Corporate Data Quality*. 2016.
- [7] International Organization for Standardization, "Quality management systems – Requirements (ISO 9001:2015)," 2015.
- [8] International Organization for Standardization, "Quality management systems - Fundamentals and vocabulary (ISO 9000:2015)," 2015.
- [9] M. Dumas, M. L. Rosa, J. Mendling, and H. A. Reijers, *Fundamentals of Business Process Management*. Springer, 2013.
- [10] D. M. Strong, Y. W. Lee, and R. Y. Wang, "Data quality in context," *Communications of the ACM*, vol. 40, pp. 103–110, May 1997.
- [11] C. Batini, C. Cappiello, C. Francalanci, and A. Maurino, "Methodologies for data quality assessment and improvement," *ACM Computing Surveys*, vol. 41, pp. 16:1–16:52, July 2009.
- [12] W. van der Aalst, *Process Mining Data Science in Action*. Springer, Berlin, Heidelberg, 2 ed., 2016.
- [13] Object Management Group - (OMG), "Business Process Model and Notation (BPMN)," 2011. <http://www.omg.org/spec/BPMN/2.0/>.
- [14] N. Panayiotou, S. Oikonomitsios, C. Athanasiadou, and S. Gayialis, *Risk Assessment in Virtual Enterprise Networks: A Process-Driven Internal Audit Approach*, pp. 289–311. 01 2010.
- [15] H. Reijers and J. Mendling, "Modularity in process models: Review and effects," in *Business Process Management* (M. Dumas, M. Reichert, and M.-C. Shan, eds.), (Berlin, Heidelberg), pp. 20–35, Springer Berlin Heidelberg, 2008.
- [16] A. Zaveri and A. Rula, *Data Quality and Data Cleansing of Semantic Data*, pp. 573–579. Cham: Springer International Publishing, 2019.
- [17] M. Ofner, B. Otto, and H. Österle, "Integrating a data quality perspective into business process management," *Business Proc. Manag. Journal*, vol. 18, no. 6, pp. 1036–1067, 2012.

- [18] A. Rodríguez, A. Caro, C. Cappiello, and I. Caballero, "A BPMN Extension for Including Data Quality Requirements in Business Process Modeling," in *Business Process Model and Notation* (J. Mendling and M. Weidlich, eds.), (Berlin, Heidelberg), pp. 116–125, Springer Berlin Heidelberg, 2012.
- [19] C. Cappiello, A. Caro, A. Rodríguez, and I. Caballero, "An Approach To Design Business Processes Addressing Data Quality Issues," in *Proceedings of 21st European Conference on Information Systems, ECIS 2013, Utrecht, The Netherlands, June 5-8, 2013*, 2013.
- [20] P. Glowalla and A. Sunyaev, "Process-Driven Data Quality Management Through Integration of Data Quality into Existing Process Models - Application of Complexity-Reducing Patterns and the Impact on Complexity Metrics," *Business & Information Systems Engineering*, vol. 5, no. 6, pp. 433–448, 2013.