IDENTIFICATION OF BUSINESS ORIENTED DATA QUALITY METRICS

(Research Paper)

Boris Otto
University of St. Gallen, Switzerland
boris.otto@unisg.ch

Kai M. Hüner
University of St. Gallen, Switzerland
kai.huener@unisg.ch

Hubert Österle
University of St. Gallen, Switzerland
hubert.oesterle@unisg.ch

Abstract: Corporate data of poor quality can have a negative impact on the performance of business processes and thereby the success of companies. Similar to machine tools corporate data show signs of wear (imaging a moving customer with a new address, for example) and have to be monitored continuously for quality defects. Effective quality control of corporate data requires metrics that monitor potential data defects with the most significant impact on the performance of a company's processes. However, due to company specific success factors and IT landscapes, it is hardly possible to provide generic metrics that can be implemented without any adjustment. This paper presents a method for the identification of business oriented data quality metrics. The presented approach takes into account company specific requirements from both a business and an IT perspective. The method's design and evaluation process is discussed in the context of three real-world cases.

Key Words: data quality controlling, data quality metrics, design science research, method engineering

INTRODUCTION

Poor corporate data can have a negative impact on the performance and the success of companies [2, 43]. Corporate data are generated in business processes (when entering design parameters for manufacturing a watch movement, for example), and they are used in business processes (when manufacturing a watch movement using design drawings, for example). If critical data are missing, a business process may be severely impeded. And if critical data are flawed (e.g. inaccurate values), undesired results are very likely to occur. With regard to the example of watch movement manufacturing, missing design drawings would increase process cycle time, and inaccurate data in the design drawings would result in the production of scrap. Apart from impairing business process performance, poor data quality (DQ) may result in inappropriate or even wrong decision-making regarding a company's overall business strategy [39]. For example, consolidating procurement volumes for achieving optimized procurement conditions is possible only if a supplier or several subdivisions of one supplier can be uniquely identified. If the quality of supplier master data is poor, the decision in favor of or against a certain supplier will be based on poor information.

While the fact that DQ does affect business process performance becomes obvious by simple examples like the one mentioned above, the question still remains as to what DQ aspects have what impact on what aspects of business process performance. The definition of DQ as ‚fitness for use' [42, 53] leads to a set of requirements on data as determined by a specific business context. While a specific set of corporate data can be highly appropriate for one purpose, the same set of data can be totally inappropriate for another [2]. Both theoretical research and practical experiences made in business projects have shown that knowing about some generic interrelations (e.g. missing data usually contribute to decreased process cycle time) provides no sufficient basis for sustainable data quality controlling (e.g. identifying critical business problems caused by data defects as a first step). While literature offers a number of dimensions and frameworks for identifying various DQ aspects [12, 44, 52], there has hardly been any presentation of concrete DQ metrics or procedures to identify such metrics. As a matter of fact, this is what has been explicitly declared further theoretical work to be done [2, 29]. Also, the interrelatedness between data defects and their impact on business process performance usually is illustrated by means of specific cases [14, 43, 50], what suggests that company specific aspects seem to be of crucial importance.

As taking into consideration the importance of company specific aspects for identifying the interrelations between DQ and business process performance, the development of a method supposed to collect company specific requirements, include findings known from previous research, and provide DQ metrics in relation to process metrics, seems to be appropriate. The research question to be answered is: How can companies identify those DQ metrics which are relevant to their business process metrics?

The paper discusses two central components of the method suggested, namely the procedure model and the involved roles. A part of the method's meta-model is also presented in order to clarify the semantics of the concepts proposed. The main part of the paper starts with information on the background as well as related work of the topic to be examined, and outlines design science research and method engineering as the research approaches the work is based on. Following these approaches, the next section describes how the method was applied iteratively in three business cases and how it was adapted accordingly. Finally, the last section gives a brief conclusion of work at hand and offers an outlook on future research to be conducted.

BACKGROUND

Basics

Information systems provide data in a certain business context (a business process in which customer master data are used, for example). When data are used by human beings, they turn into information, and information finally turns into knowledge. A detailed discussion on the differentiation of data, information, and knowledge is offered by [3, 4, 10, 47, 48], for example. This paper postulates that data contain pieces of information, and information can be extracted by interpreting data, e.g. by subordinating data under existing categories or schemas. Data interpretation usually is done by computer programs, e.g. by interpreting a certain sequence of characters as a date of birth. So, while any transformation of data into information usually (though not always) is independent of the user of the data, it is by any means dependent on the context the data are used in (the computer program, in the example) [51]. Knowledge, finally, is generated by linking pieces of information with one another, or by linking pieces of information with existing knowledge. This transformation does depend on the user of the information, and on their specific situation. The result of any such transformation process can be seen in the effect generated by any concrete action of the information consumer in a certain context. For example: In a DQ monitoring process, a figure (piece of data) is interpreted as a certain value of a metric (information) triggering a maintenance activity (effect) because a certain threshold value (existing knowledge) has been exceeded (linking of information and knowledge).

Regardless of such clear theoretical differentiation between data and information, practitioners use the term ‚data' in a broader sense. Master data (e.g. customer or materials master data) are not just values (e.g. 0721) but comprise also the act of interpreting by means of certain schemas (e.g. a telephone area code) or in a certain context (e.g. area code plus a customer's phone number). As the method to be presented in this article does not so much aim at a theoretical differentiation of certain terms but rather focuses on the practical use of data in business processes, the paper favors broader semantics of the term ‚data' (cf. Figure 1).

DQ is defined by the degree of benefit (or value) perceived by a user using certain data in a certain context ('fitness for use') [42, 53]. The benefit the data bring about (i.e. their quality) is described by a set of quality dimensions from the user's point of view, such as timeliness, completeness, and believability. In order to obtain concrete values for DQ, these DQ dimensions need to be operationalized by DQ metrics [1, 2, 8, 29]. For any DQ metric a measuring method has to be specified that determines a point of measuring (where does the measurement take place?), the object (what is to be measured?), a measuring tool (how is the measuring done?), and a measuring scale [2].

Related Work

There have been numerous theoretical studies on the identification of DQ dimensions [11, 13, 29, 42, 52, 53]. Some of these studies even include a definition of metrics for DQ measurement. However, the metrics proposed are either generic and do not include a description of possible measuring techniques [11], or they refer to certain domains or even single specific cases only [19]. Also, the impact of DQ on companies' business process performance or on companies' capabilities in general has been examined by many experts [2, 12, 43, 45, 46]. However, what has rarely been provided yet are concrete measurements of DQ or any attempts of quantification of any stated impact on business process performance.

While provision of generally applicable DQ metrics is broadly desired by practitioners and explicitly mentioned to be in the focus of further scientific research [2, 29], the necessity to adapt metrics to specific contexts (e.g. a specific database) is given by definition due to context sensitivity of DQ ('fitness for use'). Therefore, some scientific studies dealt with procedures as to how measuring techniques could be specified in certain contexts [6, 15, 17, 37]. An overview on various techniques is offered by Batini et al. [1], who also describe a method by which DQ metrics can be specified. A similar approach comes from Caballero et al. [7], who describe a method for specifying techniques for DQ measurement on the basis of a specific DQ related information model [8].

Research Methodology

The paper offers a description of the design process of the proposed method for the identification of business oriented DQ metrics. Thus, the paper's main topic is a method as an artifact of design science research (DSR). As a consequence, the paper does not present and discuss any concrete DQ metrics, but the method itself and the process of its design. The paper uses DSR as a methodological framework for the general design process, and method engineering (ME) as a concrete guideline for designing the method. The method is designed and evaluated together with various companies of an applied research program.

DSR is a framework for design oriented research, aiming at the design of solutions to practical problems [20, 31]. Outcomes of DSR are artifacts, i.e. constructs, models, methods, or instantiations. In this context, a method is defined as a procedure applied to solve a problem (e.g. an algorithm or best practices). Regarding several requirements [20], the design process of a DSR artifact comprises phases of constructing, evaluating, and adapting the artifact [16, 31].

A method in the context of ME represents a form of integrated, systematic procedure for developing information systems [21, 38]. ME deals with the design of methods [5], and applies method principles (i.e. an engineering design approach of information systems) to the design of methods (i.e. the design of a method's components *design activities*, *design results*, a *meta-model*, *roles*, and *techniques*) [18]. Design activities aim at producing one or more defined design results. Activities can be structured hierarchically and can be part of a process sequence. The total of process sequences constitute the procedure model as one central component of a method [30]. Being a formal language, the meta-model defines the syntax and the semantics of the results of the activities. Roles aggregate several activities, which are executed by individuals or boards assigned to a certain role. Roles are always involved in activities in one way or another (e.g. 'responsible'). Techniques describe how results or groups of logically associated results are produced 'within' activities [5].

METHOD DESIGN PROCESS

Overview

The paper presents three cases in which, together with the companies involved, business problems and causing data defects are identified and related DQ metrics are specified. The following subsections describe the use and the design process of the proposed method in the context of the three presented cases. Table 4 summarizes the cases' contribution to the components of the method. Table 5 provides a description of the activities of the procedure model (i.e. the design state after Case C) which are already mentioned in the case descriptions. The companies mentioned in the three cases are members of the consortium research project Competence Center Corporate Data Quality (CC CDQ) which is part of the research program Business Engineering (BE HSG) of the University of St. Gallen.

Case A: Metrics for Customer Data Used in Customer Service Processes

Company A is a German telecommunications provider with 234,123 employees worldwide, about 20 million customers, and an annual revenue of 61.12 billion euros (in 2008). Company A was the first company that evaluated the method. The search for DQ metrics in this company was basically triggered by problems occurring in business processes (invoices were sent to wrong addresses due to flawed address data, for example) and by increasing inconsistencies in customer master data (missing dates of birth, dummy entries, etc.). As these data usually were collected and used by company employees (in call centers or retail stores, for example), a questionnaire was developed to conduct interviews with employees in order to identify business problems (cf. activity I.3 of the procedure model). The questionnaire was structured according to the fifteen DQ dimensions [53], and for each dimension the questionnaire contained a number of (partly company specific) questions. This kind of structured interviewing made it quite easy to assign business problems that could be identified to DQ dimensions (cf. activity I.3). However, no specification of concrete DQ metrics (cf. phase II of the procedure model) has been done so far. Primary objective was to identify DQ problems. Other, more complex tasks (e.g. the specification of measuring points, which must be considered quite time consuming due to large data volumes and a complex system land-

	Procedure Model	Roles
Confirmation	The use of a questionnaire for identifying cause-effect chains between business problems and causing data defects has proven to be very helpful, and so has the questionnaire's structuring according to the fifteen DQ dimensions. Only few problems occurred in interviews regarding interviewees' understanding of DQ.	Involving data users and collectors has proven to be very reasonable. A lot of business problems could be identified which the chief data steward had not been aware of before.
New Findings	At the beginning a concrete benefit (referring to the business process in which problems are occurring) should be specified. The specification of the measurement approach is a complex task that should be described in detail.	Persons conducting interviews should really be interested not only in identifying business problems but also in designing concrete DQ metrics including measurement devices and procedures. To be able to provide concrete specifications for DQ metrics also persons from IT departments (i.e. the persons responsible for data providing systems) should be involved.

Table 1: Findings from Case A

	Procedure Model	Roles
Confirmation	Detailing the documentation of activity II.2 (i.e. defining the four given steps for specifying a DQ metric) has proven to be very reasonable, as that made it possible to determine single aspects very specifically (e.g. specification of measuring points in focus group interviews with technical data stewards). As expected, the addition of activities I.1 and I.2 helped to focus on 'relevant' cause-effect chains and so improved the result of I.3.	Involving both business and technical data stewards in interviews for identification of DQ problems has proven to be very helpful, as, for example, points of measuring could be identified already during the interviews.
New Findings	Upon specification of DQ metrics it was discussed whether there are any general requirements on DQ metrics, and if so, if these requirements are met. Appropriate activities should be included.	No new findings could be identified.

Table 2: Findings from Case B

scape) will be done at a later stage. Therefore, upon documenting the business problems identified and classifying them according to the DQ dimensions the project was completed.

Case B: Metrics for Design Data Used in Manufacturing Processes

Company B is a Swiss producer of watch movements. Unlike in Company A, in Company B the responsibility for designing DQ metrics was clearly assigned to a certain role, the chief data steward. Moreover, at the beginning a concrete benefit referring to a certain process (i.e. a manufacture process in a certain production line) plus all IT systems involved in this process were identified. Thus, interviewees were both process users and technical data stewards.

Like with Company A, the findings of the interviews at Company B were documented and structured. In addition, the documentation for Company B included an informal description of potential DQ metrics (cf. activity I.3) as agreed upon by the interviewees. This documentation then was iteratively developed into a specification of data to be measured (e.g. all customer data of a particular database), of a measuring device, of a specification of measuring scales, and of measurement procedures (cf. activity II.2). This project led to a requirements specification for the implementation of eight DQ metrics, which meanwhile have been implemented and are being used in the manufacturing process.

Case C: Metrics for Material Data used in Maintenance Processes

Company C is a German machine manufacturer and automotive industry supplier with 282,123 employees worldwide, 290 manufacturing sites worldwide, and an annual revenue of 45.12 billion euros (in 2008). For the project with Company C, a machine maintenance process was identified for being in the focus of identifying DQ problems. The primary objective was to achieve improved conditions for the procurement of spare parts and accessories by means of cross-site procurement, with identical parts to be identified over technical characteristics (e.g. weight, dimensions, material). In the course of the project it turned out that both general requirements on DQ metrics and company specific requirements (e.g. caused by the system landscape or internal reporting requirements) have to be regarded, so activity II.1 and activity III.1 were added to the procedure model.

	Procedure Model	Roles
Confirmation	Examining requirements on DQ metrics has proven to be very helpful in order to keep in mind overall design objectives. By taking into account requirements it is possible to adjust expectations regarding the use of DQ metrics at an early stage.	No new findings from our project with Company B had to be evaluated. The role model basically has proven to be applicable for Company C too.
New Findings	To evaluate the quality of the DQ metrics developed, it is necessary to examine the degree of mitigation of the DQ problems identified. Therefore it is necessary to identify not only business processes but also KPIs allowing to monitor these processes.	To evaluate the quality of DQ metrics the respective process owners should be involved (e.g. for identifying KPIs whose values are supposed to be manipulated by improved DQ metrics).

Table 3: Findings from Case C

Method Composition

Table 4 shows how each case contributed to the development of the method's components. The line labels represent a qualitative statement on the degree of contribution for the design of the documentation model, the procedure model, the roles, and the techniques. The meta-model is left out here, as it is primarily deducted from the documentation model.

The documentation model and the techniques will not be treated further here. Basically, in all three projects similar techniques (interviews, check lists etc.) were applied. However, these techniques have not been standardized yet, what makes them hardly comparable. The development, testing, and adaptation of concrete techniques and result documents will constitute the next steps in the research process.

As far as the procedure model is concerned, which was developed in an iterative process, no activity was eliminated and at least one activity was added as we moved from one project to the next. So we ranked all three projects with 'high' regarding their contribution to the development of the procedure model. The same thing applies for the role model, apart from case C where no further role was added.

METHOD COMPONENTS

Meta-model

Figure 1 shows entities and relations that are used to describe the activities of the procedure model and the results. The six entities and eight relations form just a subset of the method's meta-model. In the context of the paper at hand, their presentation aims at clarifying dependencies between data defects, business problems, and lower process metric values, for example. The following listing provides a brief description for each entity illustrated in Figure 1.

- *Business Problem*. State (e.g. delivery of goods not possible) or incident (e.g. production of scrap parts) leading to decreased process performance and hence to poorer values of process metrics. A business problem poses a risk (in terms of probability of occurrence and intensity of impact [23], both of which are represented by a random variable) to a business process.
- *Business Process*. Sequence of chronologically and typologically linked tasks intended to generate a clearly defined output bringing about customer benefit. Transforms a given input (e.g. a certain material) into a desired output (e.g. a watch movement) under consideration of certain rules and by using certain resources (e.g. data). Is controlled and designed by means of metrics defined as part of an overall business strategy [9, 22, 40].

Contribution	Documentation Model	Procedure Model	Roles	Techniques
High	*Case C*. A template for documenting and ranking requirements for DQ metrics (cf. activity II.1) that can be used as a checklist (cf. activity III.1) was designed.	*Case A*. The procedure model comprised three activities comparable to current activities I.3, II.2 and II.3. Activity II.2 was just performed on a very high level compared to cases B and C. *Case B*. Activities I.1 (yet without focusing process metrics as well) and I.2 were added. Activity II.2 was described in more detail. *Case C*. Focusing process metrics was added to activity I.1. Activities II.1 and III.1 were added as well.	*Case A*. The interviews were conducted by a chief data steward with process users. The order was placed by a sponsor. *Case B*. Both business and technical data stewards were included into the interviews. The role of a process owner was intended as well but not yet involved.	*Case C*. A list of generic requirements for DQ metrics to be used in activity II.1 was defined in a focus group interview with subject matter experts.
Medium	*Case B*. Templates for the informal documentation of the results of I.3 and II.2 were designed.			*Case A*. A questionnaire with generic but adaptable questions to be used as a guideline for focus group interviews (cf. activity I.3) was designed.
Low	*Case A*. No further standardized result templates.		*Case C*. No further roles.	*Case B*. No further techniques.

Table 4: Contribution of cases to the design process of method components

- *Data*. Representations of objects (e.g. customers) and relations between objects (e.g. Customer A orders Article X) based on the description of certain object characteristics [35]. The process of choosing characteristics from the set of all existing characteristics of an object is an abstraction process, synonyms are *conceptualization*, defining an *abstract syntax* [33], or *data modeling* [35]. A data element (or attribute [32]) is a component of a data model which cannot be further subdivided in logical terms in a given or presupposed setting [49]. The paper considers corporate master data (e.g. customer address data, material data, or parts lists), with a focus not on data models (i.e. data classes and data elements) but on values assigned to data elements.

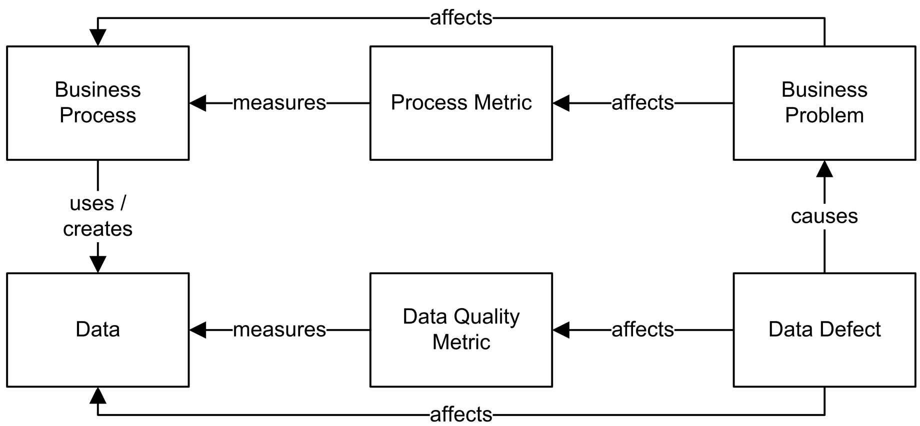

Figure 1: Entities and relations of a business oriented data quality metric

- *Data Defect*. Incident (e.g. a customer's address data is entered wrongly into a CRM system) leading to poorer values of data quality metrics. A data defect poses a risk (in terms of probability of occurrence and intensity of impact [23], both of which are represented by a random variable) to data.
- *Data Quality Metric*. Quantitative measure (cf. *Metric*, as well denoted by *data quality measure* [11, 44]) of the degree to which data possess a given quality attribute (e.g. completeness, timeliness, see [29, 41] for examples). For a data quality metric, one or more measuring methods need to be provided (i.e. where the measurement is made, what data are included, what measuring device is used, and what scale is used for measuring).
- *Metric*. Quantitative measure of the degree to which an item (in the following denoted as 'entity', according to [26]) possesses a given quality attribute (i.e. a feature or characteristic that affects an entity's quality) [25]. A measuring system (i.e. the measuring procedure that is implemented by a measuring device) measures a metric value at a certain measuring point and at a certain measuring frequency. The magnitude measured is mapped to a value on a scale to which a measuring unit is assigned [26, 36].
- *Process Metric*. Quantitative measure (cf. *Metric*, as well denoted by operational measures [28]) of the degree to which a business process possesses a given quality attribute (e.g. lead time or scrap rate). Results directly from process performance, e.g. as the average lead time for processing of an order, or as the number of order cancelations per month [34]. Provides information on a process' state indicating weakpoints and allowing immediate reaction [40].

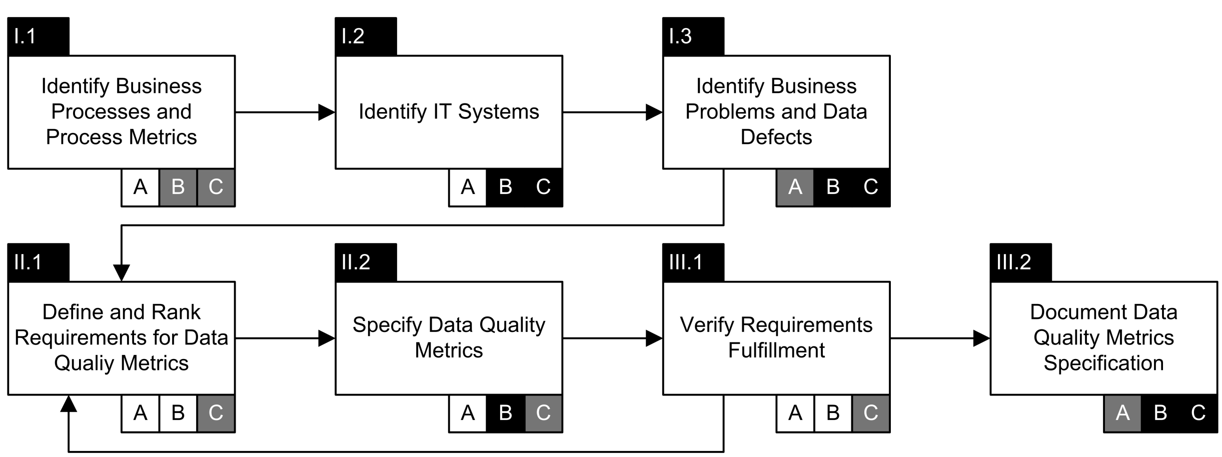

Figure 2: Procedure model and degree of usage of activities in each case

Procedure Model

Figure 2 shows the procedure model as it was specified upon completion of the project with Company C. The color codes of the letters under each activity box indicate whether this activity was fully used (black), partly used (gray), or not used (white) in the respective project (cf. Table 4 for a brief description of the activities' design process). Table 5 gives a brief description for each activity.

Name	Description
Phase I: Collect Information	
I.1. Identify Business Processes and Process Metrics	I.1 is a preliminary activity for I.3 and aims at identifying business processes and process metrics to focus on during the remaining identification process. Criteria for the selection of a particular business process might be its importance for the company's business success, for example. However, an important selection constraint is the availability (i.e. existence and access) of metrics and measurement values for the selected business processes to enable a comparison of process metric values and measurement values of the subsequently designed DQ metrics. Beside an informal documentation of the rationales for choosing the particular business processes and metrics, the result of activity I.1 comprises a list of contacts (i.e. process owner and process user) which might by invited to the focus group interviews (cf. activity I.3).
I.2. Identify IT Systems	I.2 is a preliminary activity for I.3 and aims at identifying IT systems (e.g. ERP systems, CRM systems, or databases) that are supporting the identified business processes (cf. I.1). The overall objective of activity I.2. is to identify IT experts for the focus group interviews of activity I.3).
I.3. Identify Business Problems and Data Defects	I.3 is the main activity of phase I and aims at identifying cause-effect chains between business problems and data defects (i.e. data defects that cause business problems). The top-down search direction (i.e. first identifying critical business problems and then indentifying causing data defects) has proven to be effective in the discussed cases, but indentifying potential business problems for already known data defects might be might be useful as well. Subject matter experts from both business and IT departments should be involved in collaborative focus group interviews to enable discussions with different perspectives. [24] contains interview guidelines and exemplary cause-effect chains to support the focus group interviews. The result of activity I.3 is an informal documentation of cause-effect chains (i.e. business problems and data defects) and likely affected business processes, process metrics, and data classes.
Phase II: Specify Requirements and Design Data Quality Metrics	
II.1. Define and Rank Requirements for Data Quality Metrics	The result of activity II.1 is a ranked list of requirements a DQ metric must meet. The list comprises both generic (e.g. a specified measurement device and measurement points) and company specific requirements (e.g. facility to visualize metric measurements in a specific manner). [24] contains a list of generic requirements that have been identified in a focus group interview with subject matter experts of different companies. The list is used as a checklist in activity III.1 to verify the requirements' fulfillment.
II.2. Specify Data Quality Metrics	The result of activity II.2 is a specification of at least one DQ metric. The basis for the metrics' design process are the results of activities I.3 and II.1. The activity comprises for each DQ metric a) the specification of a (subset of a) data class (e.g. customer data) that is measured by the metric,

Name	Description
	b) the specification of a measurement device (e.g. a database with data analysis functionality) and a measurement point where the measurement is to be done, c) the specification of a measurement scale to which a measured magnitude is to be mapped, and d) the specification of a measurement procedure that is implemented by the measurement device and performs the measurement at the given measurement point at a measurement frequency.
Phase III: Approve and Document Results	
III.1. Verify Requirements Fulfillment	Activity III.1 verifies the requirements defined in activity II.1. If a requirement is not met, the process starts again with II.1 in order to check the requirements' content and ranking.
III.2. Document Data Quality Metrics Specification	The result of III.2 is a documentation of the specification of the DQ metrics (cf. activity II.2) including the identified cause-effect chains (I.3), and the requirements (activity II.1). This documentation might be used as a requirements documentation for the implementation of the DQ metrics, for example.

Table 5: Description of activities of the procedure model

Roles

Table 6 shows the assignment of roles to activities of the procedure model by using the RACI notation [27] that declares a role to be responsible, accountable, consulted, or informed regarding a particular activity. The role definitions partially base on the roles described by [54]. The whole identification and specification process is managed by the chief data steward who is involved in each activity.

- *Chief Data Steward*. The person which puts strategic DQ objectives into reality and leads the process of identifying DQ metrics. Directs business data stewards and technical data stewards.
- *Business Data Steward*. Cooperates with company departments and business units on DQ issues. One business data steward is proposed as being responsible for one specific data class (e.g. material master data, customer master data etc.). Yet other concepts are possible too (e.g. responsible for a certain department, a certain process etc.).
- *Technical Data Steward*. Has detailed knowledge about the IT systems managing and providing corporate data. One technical data steward is proposed as being responsible for one specific data class (cf. business data steward).
- *Process Owner*. Plans and monitors process objectives, and plans and initiates measures for improvement. The process owner might grant the permission to use metric measurements of the monitored processes.
- *Process User*. Execute concrete tasks belonging to a certain process. Regarding the creation and usage of date in business processes, data defects have impact on daily work of these individuals.
- *Sponsor*. A single person (e.g. the chief executive officer) or a board (e.g. management board) that has sufficient resources (both money and power) to support or prohibit an initiative like the identification and implementation of DQ metrics and that might place the order with the chief data steward for designing and implementing DQ metrics.

	Chief Data Steward	Business Data Steward	Technical Data Steward	Process Owner	Process User	Sponsor
I.1. Identify Business Processes and Process Metrics	Responsible, Accountable			Consulted		
I.2. Identify IT Systems	Responsible, Accountable			Consulted		
I.3. Identify Business Problems and Data Defects	Responsible, Accountable	Consulted	Consulted		Consulted	
II.1. Define and Rank Requirements for Data Quality Metrics	Responsible, Accountable	Consulted	Consulted			Informed
II.2. Specify Data Quality Metrics	Responsible, Accountable	Consulted	Consulted			
III.1. Verify Requirements Fulfillment	Responsible, Accountable	Consulted	Consulted			Informed
III.2. Document Data Quality Metrics Specification	Responsible, Accountable					Informed

Table 6: Assignment of roles to activities of the procedure model

CONCLUSION AND FURTHER RESEARCH

The article at hand describes a method for identifying business oriented DQ metrics and outlines two of the method's central components (i.e. a procedure model and a role model). The method takes up a number of input variables (e.g. critical success factors for a company) and provides a specification of concrete DQ metrics that can be used as a basis for implementation of such metrics.

During the process of designing and adapting the method it became obvious that general findings from previous research (e.g. DQ metrics requirements) and company specific findings (e.g. concrete data defects identified) need to be combined. Generic approaches (e.g. DQ dimensions) can be used here as supporting tools but do not offer any applicable results for concrete business contexts without prior operationalization.

Despite the presented description of seven activities and six roles to perform these activities, the process of identifying 'DQ metrics that matter' remains a complex task that has to be supported not only by a guiding procedure, but also by a repository of already implemented DQ metrics which serve as best practices. In particular for the process of identifying 'relevant' cause-effect chains between business problems and data defects within focus group interviews (cf. activity I.3) illustrative examples of correlated measurements of process and DQ metrics enable discussions and thereby support the identification process. Therefore the design and documentation of DQ metrics in real-world cases, the analysis of the identified cause-effect chains, and the derivation of generic cause-effect patterns between data defects (e.g. grouped by DQ dimensions) and business problems (e.g. grouped by commercial sectors, or supply chain reference models) constitute multiple areas for further research.

REFERENCES

[1] Batini, C., Barone, D., Mastrella, M., Maurino, A., Ruffini, C. “A Framework and a Methodology for Data Quality Assessment and Monitoring.” in Proceedings of the 12th International Conference on Information Quality, Cambridge, 2007, pp. 333-346
[2] Batini, C., Scannapieco, M. Data Quality. Concepts, Methodologies and Techniques. Springer. Berlin, 2006
[3] Boisot, M., Canals, A. “Data, information and knowledge: have we got it right?” Journal of Evolutionary Economics, 14 (1). 2004. pp. 43-67
[4] Bourdreau, A., Couillard, G. “Systems Integration and Knowledge Management.” Information Systems Management, 16 (4). 1999. pp. 24-32
[5] Brinkkemper, S. “Method Engineering: Engineering of Information Systems Development Methods and Tools.” Information and Software Technology, 38 (4). 1996. pp. 275-280
[6] Burgess, M. S. E., Gray, W. A., Fiddian, N. J. “Quality Measures and the Information Consumer,” in Proceedings of the 9th International Conference on Information Quality, Combridge, 2004, pp. 373-388
[7] Caballero, I., Calero, C., Piattini, M., Verbo, E. “MMPro: A Methodology based on ISO/IEC 15939 to Draw up Data Quality Measurement Processes,” in Proceedings of the 13th International Conference on Information Quality, Cambridge, 2008
[8] Caballero, I., Verbo, E., Calero, C., Piattini, M. “A Data Quality Measurement Information Model based on ISO/IEC 15939,” in Proceedings of the 12th International Conference on Information Quality, Cambridge, 2007, pp. 393-408
[9] Davenport, T. H. Process Innovation: Reengineering Work through Information Technology. HBS Press. Boston, 1993
[10] Davenport, T. H., Prusak, L. Working Knowledge: How Organizations Manage What They Know. HBS, 1998
[11] English, L. P. Improving Data Warehouse and Business Information Quality. Wiley. New York, 1999
[12] English, L. P. “Process Management and Information Quality: How Improving Information Production Processes Improves Information (Product) Quality,” in Proceedings of the 7th International Conference on Information Quality, Cambridge, 2002, pp. 206-209
[13] Eppler, M. J. Managing Information Quality, 2 ed. Springer, 2006
[14] Fisher, C. W., Kingma, B. R. “Criticality of data quality as exemplified in two disasters.” Information & Management, 39 (2). 2001. pp. 109-116
[15] Gebauer, M., Caspers, P., Weigel, N. “Reproducible Measurement of Data Field Quality,” in Proceedings of the 10th International Conference on Information Quality, Cambridge, 2005
[16] Gregor, S. “The Nature of Theory in Information Systems.” MIS Quarterly, 30 (3). 2006. pp. 611-642
[17] Gustavsson, M. “Information Quality Measurement. Considerations when defining Measures supporting MPC decision making,” in Proceedings of the 11th International Conference on Information Quality, Cambridge, 2006
[18] Gutzwiller, T. A. Das CC RIM-Referenzmodell für den Entwurf von betrieblichen, transaktionsorientierten Informationssystemen. Physica. Heidelberg, 1994
[19] Heinrich, B., Kaiser, M., Klier, M. “Does the EU Insurance Mediation Directive Help to Improve Data Quality? A Metric-Based Analysis,” in Proceedings of 16th European Conference on Information Systems, Galway, 2008
[20] Hevner, A. R., March, S. T., Park, J., Ram, S. “Design Science in Information Systems Research.” Management Information Systems Quarterly, 28 (1). 2004. pp. 75-105
[21] Heym, M., Österle, H. “Computer-Aided Methodology Engineering.” Information and Software Technology, 35 (6/7). 1993. pp. 345-354
[22] Höning, F. Methodenkern des Business Engineering. Metamodell, Vorgehensmodell, Techniken, Ergebnisdokumente und Rollen. Ph.D. thesis, Institute of Information Management, University of St. Gallen, 2009
[23] Hubard, D. W. The Failure of Risk Management: Why It's Broken and How to Fix It. Wiley. Hoboken, 2009
[24] Hüner, K. M. “Methode zur Identifikation geschäftsorientierter Datenqualitätskennzahlen,” working paper BE HSG / CC CDQ / 13, Institute of Information Management, University of St. Gallen, 2009
[25] IEEE “IEEE Standard Glossary of Software Engineering Terminology,” The Institute of Electrical and Electronics Engineers IEEE Std 610.121990, 1990
[26] ISO/IEC “ISO/IEC 15939. Systems and software engineering – Measurement process,” International Organization for Standardization ISO/IEC 15939:2007(E), 2007
[27] ITGI “CobiT 4.1,” IT Governance Institute 2007
[28] Kaplan, R. S., Norton, D. P. “The Balanced Scorecard: Measures that Drive Performance.” Harvard Business Review, 70 (1). 1992. pp. 71-79
[29] Lee, Y. W., Pipino, L. L., Funk, J. D., Wang, R. Y. Journey to Data Quality. MIT Press. Boston, 2006
[30] Leist, S., Zellner, G. “Evaluation of Current Architecture Frameworks,” in Proceedings of the 21st Annual ACM Symposium on Applied Computing, Dijon, 2006, pp. 1546-1553

[31] March, S. T., Smith, G. F. “Design and natural science research on information technology.” Decision Support Systems, 15 (4). 1995. pp. 251-266
[32] Martin, J. Computer Data-Base Organization. Prentice-Hall. Englewood Cliffs, 1975
[33] McCarthy, J. “Towards a Mathematical Science of Computation,” in Proceedings of the IFIP Congress 62, Munich, 1962, pp. 21-28
[34] Mende, M. Ein Führungssystem für Geschäftsprozesse. Ph.D. thesis, Institute of Information Management, University of St. Gallen, 1995
[35] Mertens, P., Bodendorf, F., König, W., Picot, A., Schumann, M., Hess, T. Grundzüge der Wirtschaftsinformatik, 9 ed. Springer. Berlin, 2005
[36] Muschter, S. IS-gestütztes Prozessmanagement. Ph.D. thesis, Institute of Information Management, University of St. Gallen, 1998
[37] Naumann, F., Rolker, C. “Assessment Methods for Information Quality Criteria,” in Proceedings of the 5th Conference on Information Quality, Cambridge, 2000, pp. 148-162
[38] Nuseibeh, B. A., Finkelstein, Anthony, Kramer, J. “Method Engineering for Multi-Perspective Software Development.” Information and Software Technology, 38 (4). 1996. pp. 267-274
[39] Orr, K. “Data Quality and Systems.” Communications of the ACM, 41 (2). 1998. pp. 66-71
[40] Österle, H. Business Engineering: Prozeß- und Systementwicklung, Band 1: Entwurfstechniken, 2 ed. Springer. Berlin, 1995
[41] Pipino, L. L., Lee, Y. W., Wang, R. Y. “Data Quality Assessment.” Communications of the ACM, 45 (4). 2002. pp. 211-218
[42] Redman, T. C. Data Quality for the Information Age. Artech House. Boston, 1996
[43] Redman, T. C. “The Impact of Poor Data Quality on the Typical Enterprise.” Communications of the ACM, 41 (3). 1998. pp. 79-82
[44] Redman, T. C. Data Quality. The Field Guide. Digital Press. Boston, 2001
[45] Sheng, Y. P. “Exploring the Mediating and Moderating Effects of Information Quality on Firms? Endeavor on Information Systems,” in Proceedings of the 8th International Conference on Information Quality, Cambridge, 2003, pp. 344-352
[46] Sheng, Y. P., Mykytyn, P. P., Jr. “Information Technology Investment and Firm Performance: A Perspective of Data Quality,” in Proceedings of the 7th International Conference on Information Quality, Cambridge, 2002, pp. 132-141
[47] Spiegler, I. “Knowledge management: a new idea or a recycled concept?” Communications of the AIS, 3 (4). 2000. pp. 1-24
[48] Spiegler, I. “Technology and knowledge: bridging a ‘generating’ gap.” Information & Management, 40 (6). 2003. pp. 533-539
[49] Stahlknecht, P., Hasenkamp, U. Einführung in die Wirtschaftsinformatik, 11 ed. Springer. Berlin, 2005
[50] Strong, D. M., Lee, Y. W., Wang, R. Y. “10 Potholes in the Road to Information Quality.” IEEE Computer, 30 (8). 1997. pp. 38-46
[51] Tuomi, I. “Data Is More Than Knowledge: Implications of the Reversed Knowledge Hierarchy for Knowledge Management and Organizational Memory.” Journal of Management Information Systems, 16 (3). 1999. pp. 103-117
[52] Wang, R. Y., Storey, V. C., Firth, C. P. “A Framework for Analysis of Data Quality Research.” IEEE Transactions on Knowledge and Data Engineering, 7 (4). 1995. pp. 623-640
[53] Wang, R. Y., Strong, D. M. “Beyond Accuracy: What Data Quality Means to Data Consumers.” Journal of Management Information Systems, 12 (4). 1996. pp. 5-34
[54] Wende, K. “A Model for Data Governance – Organising Accountabilities for Data Quality Management,” in Proceedings of 18th Australasian Conference on Information Systems, Toowoomba, 2007, pp. 417-425