Process-Driven Data Quality Management: A Critical Review on the Application of Process Modeling Languages

PAUL GLOWALLA and ALI SUNYAEV, University of Cologne

Data quality is critical to organizational success. In order to improve and sustain data quality in the long term, process-driven data quality management (PDDQM) seeks to redesign processes that create or modify data. Consequently, process modeling is mandatory for PDDQM. Current research examines process modeling languages with respect to representational capabilities. However, there is a gap, since process modeling languages for PDDQM are not considered. We address this research gap by providing a synthesis of the varying applications of process modeling languages for PDDQM. We conducted a keyword-based literature review in conferences as well as 74 highranked information systems and computer science journals, reviewing 1,555 articles from 1995 onwards. For practitioners, it is possible to integrate the quality perspective within broadly applied process models. For further research, we derive representational requirements for PDDQM that should be integrated within existing process modeling languages. However, there is a need for further representational analysis to examine the adequacy of upcoming process modeling languages. New or enhanced process modeling languages may substitute for PDDQM-specific process modeling languages and facilitate development of a broadly applicable and accepted process modeling language for PDDQM.

Categories and Subject Descriptors: D.2.2 [**Software engineering**]: Design Tools and Techniques—*Flow charts*; K.6.3 [**Management of Computing and Information Systems**]: Software Management—*Software process*; J.1 [**Computer Applications**]: Administrative Data Processing—*Business*

General Terms: Design, Documentation

Additional Key Words and Phrases: Information quality, data quality, process modeling, conceptual modeling, data and knowledge visualization.

ACM Reference Format:
Glowalla, P. and Sunyaev, A. 2014. Process-driven data quality management: A critical review on the application of process modeling languages. *ACM J. Data Inform. Quality* 5, 1–2, Article 7 (August 2014), 30 pages.
DOI:http://dx.doi.org/10.1145/2629568

1. INTRODUCTION

Data quality is critical to organizational success, since organizations process vast amounts of data [Gorla et al. 2010; Madnick et al. 2009; Shankaranarayanan and Wang 2007; Tallon 2010] and poor data quality leads to high costs [Redman 1998, 2004]. Additionally, missing data quality management inhibits identifying unknown data quality issues, thus, negatively affecting decision making and strategic planning. To improve data quality provided by information systems, many methodologies exist, addressing data quality from different perspectives. Methodologies are "guidelines and techniques that [...] define a rational process to assess and improve the quality of data" [Batini et al. 2009]. Methodologies recommend various techniques. Process-driven techniques and accordingly process-driven data quality management (PDDQM)

Author's address: A. Sunyaev (corresponding author), University of Cologne; email: sunyaev@wiso.uni-koein.de.

DOI:http://dx.doi.org/10.1145/2629568

seek to assess and improve data quality by redesigning processes that create or modify data. PDDQM seeks to identify root causes of errors, eliminate them, and sustain the improvements in the long term [English 1999; Redman 1996].

Process modeling is mandatory for conducting process control activities or process redesign [Batini et al. 2009]. A prominent process-driven perspective is to treat data as a product that is processed from raw data to the final information product (IP) (cf. [Lee 2006; Wang 1998]). This perspective is applied in the total data quality management (TDQM) methodology. To represent processing of an IP, TDQM suggests the modeling of an information manufacturing system (IMS) [Wang 1998]. The IMS describes the process of how an IP is produced and the interactions among the process stakeholders (i.e., information suppliers, manufacturers, consumers, and IP managers). An extension of the IMS is the information product map (IP-MAP) [Lee 2006; Shankaranarayanan et al. 2003; Shankaranarayanan and Wang 2007]. These IP-centric modeling languages are used to model the manufacturing process of an IP and complement other modeling languages, for instance, process flow charts (PFCs) or dataflow diagrams (DFDs) [Shankaranarayanan and Wang 2007].

However, a wide range of process modeling languages emphasize different aspects of processes [Recker et al. 2009], and organizations apply process modeling languages in different ways. Extant research mostly neglects integrating data quality into business processes [Cappiello et al. 2013]. Moreover, the degree to which data quality and the production of IPs are integrated depend on the process (e.g., product manufacturing vs. IP production) and the process modeling language used (e.g., PFC vs. IP-MAP). Furthermore, IP-MAPs still are under development to fulfill further requirements [Shankaranarayanan and Wang 2007]. An inadequate use of process models may impede communication and understanding of business processes and underlying data quality requirements as well as data-quality issues. Resulting poor data quality may lead to direct and hidden costs [Haug et al. 2011]. Consequently, we are interested in providing an overview of the varying use of process models for PDDQM. This basic objective leads to our first research question (RQ).

RQ1. What varying application of process modeling languages for PDDQM in organizations can be derived from literature?

To advance IS research in terms of an adequate application of process modeling languages, we are interested in organizations' requirements that should be fulfilled by customizing process models. Since extant research lacks a structured examination of PDDQM (cf. Section 2.3), building on RQ1, we state RQ2.

RQ2. What relevant requirements for PDDQM modeling can be derived from the identified process models?

Answering these RQs would support further research for developing a broadly applicable process modeling language for PDDQM or at least guide the development and use of process models for specific purposes, additionally satisfying practitioners' needs. In turn, answering the RQs would support long-term improvement of PDDQM and thus data quality in organizations. Based on the approach from Webster and Watson [2002], we conducted a keyword-based literature review.

Summarizing published research, this article provides a review of which process modeling languages are applied to represent data and its quality in organizational contexts. Furthermore, we present an overview of how the methods are embedded in existing data-quality methodologies. Building on this review, we finally derive process modeling requirements related to PDDQM and present their potential relevance depending on the modeling purpose. The main contribution of our study is twofold. First, the synthesis provides examples of how organizations can apply well-known and mature process modeling languages, enhancing them with information about data quality fitting their particular organizational context, instead of switching to new process modeling languages. Second, the integration of data quality into different process modeling languages and the context-specific application of these languages show existing needs and opportunities for further differentiated research in PDDQM.

The remainder of this article is organized as follows. In the next section, we give an overview of data-quality research. We focus on PDDQM, corresponding process modeling languages, and present existing gaps. In Section 3, we provide our methodology regarding the literature review and present our identified primary studies. The primary studies are structured with respect to (1) the organizations in which the data-quality assessment or improvement took place, (2) the process types the efforts are aimed at, (3) the applied methodology, and (4) the applied process modeling language. For answering RQ1, each of these four aspects of the primary studies is discussed in more detail in Section 4, which is concerned with the application of process modeling languages for PDDQM. Section 5 deals with RQ2, providing the identified representational requirements for PDDQM based on the primary studies. Finally, the key issues are discussed in Section 6, and implications for practice and research are provided in the concluding Section 7.

2. DATA AND INFORMATION QUALITY

2.1. Distinguishing Data, Information, and Knowledge

Our research considers data and information quality. However, we apply the term data quality, referring to data and information quality (cf. [Madnick et al. 2009]). To specify the scope of our research, we provide the differences between data, information, and knowledge.

We refer to data as the representation of facts [English 1999]—describable states of the physical world [Boisot and Canals 2004]. These facts are represented by symbols inscribed by human hands or by instruments [Spiegler 2003]. Furthermore, data is the raw material for information [English 1999]. Information is derived from data by putting data into context, that is, giving data a meaning [Boisot and Canals 2004; English 1999; Zack 1999]. This meaning and the information that can be extracted from data depend on the receiver of the data. To add value, information and its real potential has to be understood by the receiver, turning it into knowledge [English 1999]. Knowledge is also referred to as the capacity for effective action [Spiegler 2003]. This capacity depends on the consumer of information, their mental models and values [Boisot and Canals 2004], and their specific situation [Otto et al. 2009]. Hence, knowledge is personalized information [Alavi and Leidner 2001] that can be extended from an individual to an organizational level through interaction [Škerlavaj et al. 2010].

Our article deals with the management of data quality for use in a specific context, that is, for a given task. However, how delivered information is actually used by the consumer is outside our scope. Knowledge management is beyond our scope, as it deals with this personalized knowledge, for example, its creation, sharing, and distribution [Alavi and Leidner 2001]. Since data quality is defined as data's fitness for use and this fitness is defined by the data consumer (cf. next section), we need to consider consumers' requirements on data. Based on their knowledge, the consumer has expectations on the data and information [Boisot and Canals 2004] which can be formulated as requirements on the data, improving the quality of the data as it is set into context.

2.2. Data-Quality Research

The broad definition of data quality as fitness for use (e.g., [Madnick et al. 2009]) is difficult to measure [Kahn et al. 2002]. An important milestone to data quality research is an empirical study, adopting the view of the data consumer and treating data as a product with attributes important to the consumer [Wang and Strong 1996]. This study extended previous intuitive and theoretical approaches for defining data quality. Classifications of data-quality dimensions and a discussion of the mostly referred to dimensions (namely, accuracy, completeness, consistency, and timeliness) can be found in Batini et al. [2009]. To extend the product view of data quality with the service view, the product and service performance model for information quality (PSP/IQ) was developed [Kahn et al. 2002]. The product view includes dimensions related to product features and involves tangible measures. The service view includes quality dimensions that are related to the delivery process and intangible measures.

No general agreement exists about which set of dimensions defines data quality, nor about the exact meaning of each dimension [Batini et al. 2009; Haug et al. 2009; Lee et al. 2002], since data quality needs to be assessed within the business and task context [Even and Shankaranarayanan 2007]. Furthermore, with new types of information systems, it will be necessary to match data-quality dimensions to new technological contexts [Batini et al. 2009; Kim et al. 2005; Madnick et al. 2009].

The field of data-quality research is a crosssectional issue with several fields of research (e.g., [Glowalla and Sunyaev 2013a; Madnick et al. 2009; Sunyaev and Chornyi 2012]). Extant research in data quality provides a comprehensive overview of the research field. A pragmatic framework allows for categorizing data-quality research related to research topics and research methods [Madnick et al. 2009]. This framework allows for the classification of each research project dealing with data quality in the field of management information systems and computer science. Madnick et al. use this framework to give an overview of the landscape of data-quality research and existing literature that addresses the topics and applies the research methods. Furthermore, a comparative description of methodologies dealing with the assessment and improvement of data quality is provided [Batini et al. 2009]. Nine perspectives for analyzing and comparing data-quality methodologies are presented, including the differentiation between process- and data-driven techniques. We are interested in the long-term improvement of data quality and therefore in PDDQM.

2.3. Representational Research on Process Modeling Languages

Processes are logical sequences of tasks, where goods and services are created or where the creation is coordinated using resources [Buhl et al. 2011]. To emphasize the involvement of business stakeholders as process model users, we focus on “business and manufacturing processes that create, update, and delete data, distribute or disseminate information, and retrieve or present information to information producers and knowledge workers” [English 1999]. For simplicity, we stick to the term process.

Process models allow for understanding and communicating processes and thus are mandatory for conducting process-control activities or process redesign [Batini et al. 2009]. Process models are instantiations from process modeling languages. Process modeling languages provide a vocabulary of model elements and compositional rules, which define legal compositions of the vocabulary. A general meaning of the vocabulary’s elements is given as well but should not be confused with the semantics and meaning of the instantiation, which relate to a specific (problem) domain [Lindland et al. 1994; Moody 2009].

Current research examines several process modeling languages with respect to their representational capabilities. The Bunge-Wand-Weber representation model [Wand

and Weber 1993, 1995] is applied for a structured examination and comparability of several process modeling languages, that is, to what extent process modeling languages and their elements are able to represent desired types of real-world phenomena [Recker et al. 2010]. With the examination of the business process model and notation (also referred to as business process modeling notation) (BPMN) [Recker et al. 2009; Recker and Rosemann 2010], the ontological analysis of process modeling languages was updated. Additionally, extant research compares the examined process modeling languages, for instance, petri nets, DFDs, and BPMN [Rosemann et al. 2009]. We are not aware of such an analysis for process modeling languages developed for PDDQM. Since different representational capabilities are relevant for PDDQM, process modeling languages, their application, and potential requirements for PDDQM should be examined first.

We see the necessity to include process modeling languages that focus on data quality within this research stream, since several modeling languages emphasize different aspects of processes. Extant literature shows the need to consider the information flow in processes, regardless of whether the IP is the final product (cf. Section 4.2) or information is needed for tangible products (e.g., [Lee et al. 2007a]). Building on a comparative description of process-driven methodologies [Batini et al. 2009], we find two different process modeling languages focusing on data quality. First, the information chain maps embedded within the cost-effect of low data quality (COLDQ) methodology [Loshin 2001] to model strategic and operational dataflows. Second, the already mentioned IP-MAP, which extends the IMS that is embedded within the TDQM methodology.

Information chain maps provide generic steps to represent the conversion of raw input data into usable information. Strategic and operational dataflows are set together from generic steps that are replaced by specific processing steps (e.g., data entry, credit card processing, data collection and merging). Annotations can be added to the arcs (the information flow channel), but no further information about the IP or its quality are integrated into the models. Information chain maps show the directed information flow between different locations. Similar to DFDs [Shankaranarayanan and Wang 2007], no explicit processing sequence can be derived. However, the information flow is not represented between processes but between stakeholders and systems. Alternatively, the processing stages of the dataflow, from data supply to data consumption, can be presented in a process sequence, thus, resembling IP-MAPs.

The IMS is introduced by Ballou et al. [1998], applying concepts from product quality in manufacturing systems. The IP attributes and data units can be tracked systematically from the source to the final IP that is delivered to the consumer. Furthermore, the impact of system modifications on the attributes can be analyzed. The IP-MAP was introduced as an extension of the IMS [Shankaranarayanan et al. 2000]. The design of the IP-MAP is driven by requirements of the final IP. Therefore, the final IP provides the basis for the specification of necessary raw or component data. A major change (with respect to the IMS) is the definition of additional modeling elements, namely, the business boundary block, the information system boundary block, and the decision block [Lee 2006; Shankaranarayanan et al. 2003]. A comprehensive description of the IP-MAP can be found in Shankaranarayanan et al. [2003]. Figure 1 shows a simple example of the IP-MAP.

Both models focus on information as a product. However, the representation of data differs, as do the models. The IP-MAP focuses on the delivery of a specific IP and on the necessary sequential steps to manufacturing such an IP. Additionally, the necessary data and its sources are presented. 'Necessary' means that the presented dataflow is limited to the purpose of producing the IP. As process modeling is a subtask of process management [Buhl et al. 2011], several other process modeling languages,

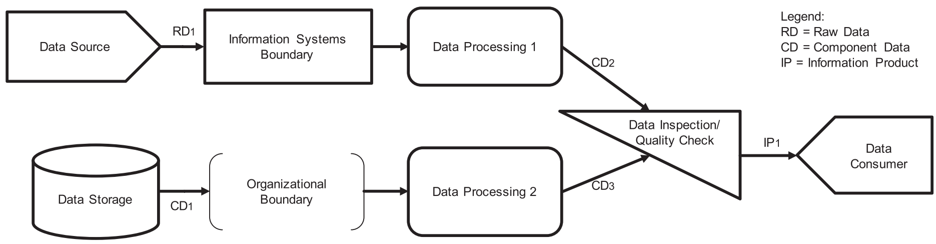

Fig. 1. Example of an IP-MAP.

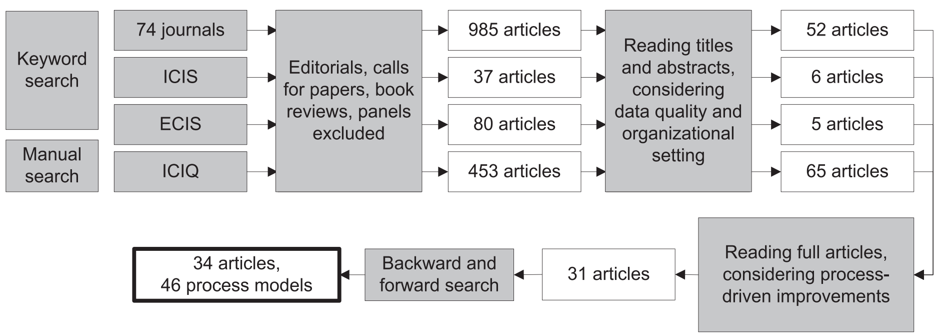

Fig. 2. Article selection method.

that is, activity-centric modeling languages [Recker et al. 2009], have to be considered [Ko et al. 2009; Recker et al. 2009], for example, petri nets, DFDs, and BPMN [Rosemann et al. 2009]. Different process modeling languages emphasize different aspects of processes [Recker et al. 2009]. Therefore, research addresses the integration of different languages and their focus (e.g., [Thi and Helfert 2007]). However, integrating different languages increases the language's complexity [Glowalla and Sunyaev 2013b; Thi and Helfert 2007]. To identify different applications of process modeling languages, we conducted a structured literature review that is presented in the next section.

3. REVIEWING DATA QUALITY LITERATURE

3.1. Keyword and Manual Search

To identify the varying applications of process modeling languages and provide the basis for answering our RQs, we conducted a structured literature review. We followed the approach proposed by Webster and Watson [2002]. The methodology is described in the following and shown in Figure 2.

According to the first phase, we based our keyword search on the Senior Scholars' Basket[1] and the 50 highest-ranked journals, applying the AIS/MIS journal ranking.[2] We additionally included the *ACM Journal of Data and Information Quality*, the *International Journal of Information Quality*, and *Data & Knowledge Engineering* due to their high relevance for this research topic. Regarding these included ACM and IEEE Transactions and further journals, we conducted a manual selection, considering our inclusion and exclusion criteria. The selection led to 74 journals, provided in the Appendix. As the already mentioned prominent perspective—to view data as a

[1] `http://aisnet.org/general/custom.asp?page=SeniorScholarBasket`.
[2] `http://aisnet.org/general/custom.asp?page=JournalRankings`.

product—was proposed in 1995, we included articles from 1995 onwards. To allow a view on the latest developments and broaden the research on more practice-oriented articles, we also reviewed three conferences. First, we included the International Conference on Information Systems (ICIS) and the European Conference on Information Systems (ECIS). Second, we included the International Conference on Information Quality (ICIQ) due to its relevance to our topic. As the ICIQ proceedings are not accessible for a keyword-based search, they were searched manually. The ECIS proceedings before 2000 needed to be searched manually as well. We derived the keywords based on an explorative search, especially considering review and overview articles. The keywords were consolidated, counted, and supplemented (e.g., information product was supplemented by data product), leading to the following list: *data quality, information quality, data product(s), information product(s), data production, information production, data manufacturing, information manufacturing, data management, information management, data flow(s), information flow(s)*. The keywords were searched in the title, abstract, and keywords/subject terms, excluding, for example, editorials and calls for papers, the search yielded 1,555 articles. We conducted the search in February 2013 and included all articles available until then.

3.2. Inclusion and Exclusion Criteria

First, we read the titles and abstracts of the 1,555 articles, excluding editorials, calls for papers, book reviews, and panels. If no abstract was available, we read the article in more detail. We considered an article relevant and therefore included it if it focuses on data-quality or at least on one of its quality dimensions. According to RQ1, the data quality aspects have to be set into an organizational setting, dealing with measures to assess or improve the quality of the organization's data. We considered an article within an organizational setting if the measures to assess or improve data quality were conducted in the field and are described in the context of the particular organizational setting. That means we included, for instance, case studies and case descriptions. We excluded articles in which results are presented isolated from the organizational setting (e.g., the presentation of lessons learned with short examples from conducted case studies or personal experience for corroboration). In case of the inclusion criterion dealing with organizational data, we explicitly considered data stored and processed by information systems. After this review, 128 articles were selected.

We only selected articles which included process-driven strategies. Employing the aforementioned definition of processes, we excluded information systems development processes [English 1999] that do not provide business processes. Since we are interested in the organizational context and models are mainly used for communicating processes [Bandara et al. 2005; Dehnert and van der Aalst 2004], including stakeholders that are non-modeling experts (cf. [Rosemann 2006]), we explicitly excluded improvements of processes that are inherent to information systems (e.g., the optimization of data warehouse internal processes). This review led to 31 articles.

We did not identify further articles in the backward search that constitutes the second phase [Webster and Watson 2002]. The forward search, constituting the third phase, led to three more articles. We chose 'Google Scholar', since it indexes conference papers in addition to journal papers.

Of 18 conference articles, 14 are from the ICIQ and three from the ECIS. The 15 Journal articles are from several different journals. Four articles are from the *International Journal of Information Quality*. The remaining journal articles are scattered across nine different journals. Finally, one working paper is included [Wang et al. 2002], which was found in the backward and forward search. Overall, the distribution of the articles throughout the outlets corroborate the crosssectional character of data quality.

As in some cases, our differentiation is rather soft and the inclusion or exclusion of articles was discussed, we cannot claim that we captured all relevant articles dealing with process-driven data-quality techniques. More importantly, we aim at a detailed description and the presentation of the broad employment of these techniques.

3.3. Primary Studies

This section gives an overview of the identified articles. According to RQ1, we examine the application of process modeling languages for PDDQM across organizations. Since some process modeling languages are rooted in PDDQM methodologies, the methodologies might bridge organizations' PDDQM efforts and adequate process modeling languages (cf. Section 2.3). Therefore, we additionally examine if and what methodologies are applied. Finally, PDDQM focuses on process improvements, and thus we examine the specific processes provided in the primary studies as well.

For a structured view on this context, we differentiate between organizations that manufacture or deliver tangible products (P) and service organizations (S). However, it is difficult to clearly dichotomize products from services, as virtually "all tangible products have intangible attributes, and all services possess some properties of tangibility" [Pitt et al. 1995]. Since this article focuses on IPs, we additionally transfer the differentiation between services and products to the intangible domain. From an output perspective, an organization might deliver a service or an IP. Again, a clear differentiation might be difficult, since services are necessary to produce IPs [Kahn et al. 2002]. Moreover, IPs have tangible properties as well. In contrast to tangible products, a specific physical form might not be part of the IP, but an IP (e.g., a report) will be delivered as a tangible print-out or on a tangible storage medium. With the distinction of tangible products, services, and information products, we are able to examine the organizational context in which process-driven improvements take place in more detail. Additionally, data have to be managed actively, that is, as a product and not as a by-product. Hence, different approaches may be appropriate for different types of organizations. We structure the processes in a similar way, differentiating manufacturing, service, and IP processes. This is necessary, as an organization typically has several consumers—internal and external—and also within a manufacturing organization, data quality improvement may take place, focusing on information production processes (e.g., reporting).

These four aspects—organization, process, methodology, and process modeling language—will be discussed in the next sections in more detail. Table I provides an overview of the primary studies. If an article contains more than one relevant case—for instance, several process—models within different organizations or representing different processes, the cases are separated by an "m" slash.

4. APPLICATION OF PROCESS MODELING LANGUAGES FOR PDDQM

This section forms the main part for answering RQ1, as we present the primary studies with respect to the structure in Table I. We provide the context of the application of process modeling languages within PDDQM with regard to the organizations, the processes, and the methodologies. Then we present the process modeling languages in more detail.

4.1. Organizations and Processes

Table II shows the organizations and the related processes. Regarding our primary studies, most cases deal with service organizations (Table II, column 1–2). However, at the process level, most cases consider IP processes (Table II, column 3–4). If we examine the relations between organizations and processes in more detail, we see that only 14 out of 46 cases deal with the same type of process and organization (Table II,

Table I. Overview of Primary Studies

Article	Organizations (P/S/IP)	Processes (P/S/IP)	Methodologies	Process modeling languages
[Balka et al. 2012]	S	IP	—	PFC
[Sulong et al. 2012]	S / S / S / P /P / S / S	—	4-step service oriented architecture implementation	—
[Ofner et al. 2012]	P	S	Data quality process redesign	PFC
[Wamba 2012]	S	P	—	—
[Dejaeger et al. 2010]	S	IP	Process model driven survey	PFC
[Xie and Helfert 2010]	S / S / S	S / S / S	—	PFC
[Gaynor and Shankaranarayanan 2008]	S	IP	—	IP-MAP
[Hakim 2008]	S	S	IDEF0, Design Structure Matrix	PFC
[Laumann and Rosenkranz 2008]	P	P / P	Viable System Model	PFC, DFD
[Lee et al. 2007b]	S / S / S / S / S / S	S / IP / IP / IP / IP / IP	—	Context-embedded IP-MAP
[Tee et al. 2007]	S	—	Total Quality Service	—
[Thi and Helfert 2007]	P	IP or P	—	IP-MAP, PFC, DFD
[Wijnhoven et al. 2007]	P	—	Total Data Quality Management (TDQM)	(only generic IP-MAP in the context of TDQM)
[Shankaranarayanan and Cai 2006]	P	IP	—	IP-MAP
[Dravis 2005]	P	—	IQ Solution Cycle	—
[Keenan and Simmons 2005]	P	S	Customer Support Data Quality (CSDQ), extension of TDQM	DFD
[Mielke 2005]	S	—	7-step cycle, based on TDQM	DFD, PFC, other
[Davidson et al. 2004]	S	IP / IP / IP / IP	TDQM, Plan-Do-Study-Act (PDSA)	DFD
[Klesse et al. 2004]	S	IP	Data Evolution Life Cycle (DELC)	PFC, CED
[Lee et al. 2004]	P	—	TDQM	—
[Shankaranarayanan et al. 2003]	S	IP / IP/ IP	—	IP-MAP
[Katz-Haas and Lee 2002]	S	S / S	Root-Cause Analysis	DFD
[Kovac and Weickert 2002]	IP	— / S	TDQM, based on Plan-Do-Check-Act (PDCA)	PFC
[Wang et al. 2002] and [Lee et al. 2001] (IP-MAPs cf. [Shankaranarayanan et al. 2003])	S	P	Total Information Awareness with Quality (TIAQ), based on TDQM	PFC

Table I. Continued

Article	Organizations (P/S/IP)	Processes (P/S/IP)	Methodologies	Process modeling languages
[Eppler 2001]	IP / IP	IP / IP	Information quality framework for knowledge-intensive processes	—
[Helfert and von Maur 2001]	S	IP	Data Quality Management (DQM) for Data-Warehouse-System, based on TDQM, and total quality management	PFC
[Kahn et al. 2001]	S	S	Six Sigma	DFD
[Millard and Lavoie 2000]	S	IP	—	use case diagram, state diagram
[Ballou et al. 1998]	P	IP	—	IMS
[Kovac et al. 1997]	IP	IP	TDQM, PDCA	PFC
[Harkness et al. 1996]	P	— / S or IP	Seven-Step Reactive Problem Solving, PDCA, Process Discovery	CED, PFC
[Meyer and Zack 1996]	IP	IP	Product and process platform	PFC
[Zack 1996]	IP	IP / IP	Product and process platform	PFC, workflow diagram
Deliverables of organization or process: P = tangible product; S = service; IP = information product DFD = Data-Flow Diagram; PFC = Process Flow Chart; CED = Cause-and-Effect Diagram				

Table II. Number of Cases by Organization and Process Type

Organization type	No. of cases	Process type	No. of cases	Same process and organization type	No. of cases
Tangible product	11	Tangible product	3	Tangible product	2
Service	29	Service	11	Service	7
IP	6	IP	26	IP	5
Total no. of cases	46		40		14

column 5–6). Only one organization delivering tangible products deals with processes whose deliverables are tangible products. Furthermore, only 7 out of 29 service organizations deal with service processes, whereas 5 out of 6 cases deal with IP organizations and IP processes. We could not classify the remaining case in the IP organization (cf. Table I, [Kovac and Weickert 2002]).

4.2. Processes and Their Deliverables

Based on the different deliverables (tangible product, service, IP) of a process presented in Table I, we categorized the processes for internal or external consumers. However, in contrast to the view applied on the organizations, processes are considered from the input through processing to the deliverable. Therefore, internal or external suppliers and custodians can be explicitly included in this view. We assume the process type having implications on the process modeling languages, that is, what process

Table III. IP Processes and Their Deliverables

Deliverables	No. of cases		No. of cases		No. of cases
Reports	19	External reports	14	End product	3
				Other	11
		Internal reports	5		
Other end products	4				

modeling language is applied and how, depending on the process type and possible differences within the process types.

The number of cases related to a specific process type is presented in Table II. Our literature review provided two articles dealing with the same remanufacturing process of a tangible product [Lee et al. 2001; Wang et al. 2002] and only one further article dealing with a delivery process of a tangible product represented by different models [Laumann and Rosenkranz 2008]. The service processes deal with internal administration or administration of service products (e.g., insurance policies) [Kahn et al. 2001; Katz-Haas and Lee 2002; Ofner et al. 2012], a service conducted on the customer [Hakim 2008], or service requests of customers [Keenan and Simmons 2005; Xie and Helfert 2010]. In the latter case, besides a service (fixing a product), the deliverable can be an IP (product or service offering) as well. In only one case, we classified a process as a service, although it is modeled as an IP-MAP [Lee et al. 2007b]. This case deals with a trading process between two companies, where mismatched information occurred. The IP that is delivered is a response to a service process, which constitutes the main deliverable.

Most of the processes deal with the delivery of IPs. In the literature, the importance of information is often related to decision making (cf. [Loshin 2001; Redman 1998; Shankaranarayanan et al. 2008]). Therefore, it is not surprising that most of the IPs within the identified processes are reports. Considering organizational boundaries, we can differentiate between internal and external reports. The external reports can be further divided into IPs that constitute the end product of the company and other reports (e.g., government reports). Finally, there are end products that do not constitute reports. The number of cases we assigned to each category are presented in Table III.

Internal reports constitute the information basis for decision making within an organization. Such decision making can be a patient's data request process by physicians [Lee et al. 2007b], as well as a process producing business or health reports (e.g., [Shankaranarayanan et al. 2003; Gaynor and Shankaranarayanan 2008]). In a more abstract view (e.g., inter-organizational), *external reports* are used within decision making as well. However, in this case, the decision maker would be beyond the organizational boundary. Examples for external reports that constitute *end products* are market research reports [Zack 1996; Eppler 2001]. In other cases, organizations have to provide external reports, for instance, due to law regulations. Therefore, these reports have to be delivered but do not constitute an end product. Most of these cases deal with healthcare IPs [Davidson et al. 2004; Lee et al. 2007b; Millard and Lavoie 2000; Shankaranarayanan et al. 2003]. Another example is the rating process for banks [Dejaeger et al. 2010] or the customer investigation process within a bank which has to be conducted upon request [Klesse et al. 2004]. The requests can be carried out in connection with prosecutions and arise externally (e.g., by the government) but can also be requested internally. IPs can constitute end products apart from reports as well. These end products are, for example, business solutions [Kovac et al. 1997]. In this case, instead of reporting the gathered data, actionable information is delivered to the customer. Further examples are IPs, such as newsletters, consolidated news,

Table IV. Overview of TDQM-Based Methodologies

	TDQM or PDCA phases	Other phases	Process modeling language derived from IMS, IP-MAP	Process modeling language DFD	PFC
[Wijnhoven et al. 2007]	TDQM	—	X		X
[Keenan and Simmons 2005]	TDQM	—	—	X	—
[Mielke 2005]	—	X	—	X	X
[Davidson et al. 2004]	PDCA	—	X	X	—
[Lee et al. 2004]	TDQM	X	—	—	—
[Kovac and Weickert 2002]	—	X	—	—	X
[Wang et al. 2002] and [Lee et al. 2001]	—	—	—	—	—
[Helfert and von Maur 2001]	—	X	—	—	X
[Kovac et al. 1997]	—	X	—	—	X
TDQM = Total Data Quality Management; PDCA = Plan-Do-Check-Act					

feeds [Meyer and Zack 1996], book abstracts [Eppler 2001], or IPs that are necessary to continue another process [Ballou et al. 1998]. Since a process can have several deliverables, it is not always possible to classify IP processes. In one process model, two IPs seem to be used for specific purposes (uploaded onto website and used to print mailing labels), whereas the other IP obviously can be used for several purposes, implied by the usage to 'run the business' [Lee et al. 2007b]. Applying the differentiation between strategic and operational dataflow [Loshin 2001], the IP for running the business might be considered an IP for decision making. Therefore, this IP might be an internal report IP. However, the other IPs rather constitute operational IPs and are used for further processing. Similarly, another case deals with IPs that are requested for external or internal consumers [Lee et al. 2007b]. The IPs are not described in detail; however, since the information is requested and delivered ad-hoc, we would assume a rather operational usage.

Five cases remain in which we could not assign to a process type, although specific processes were addressed. One reason being processes with several deliverables of different types [Harkness et al. 1996; Kovac and Weickert 2002; Mielke 2005]. Another reason being that flawed processes can be identified by flawed data [Lee et al. 2004]. Since data can be produced by any process, the referred process and therefore its type is not known in advance.

4.3. PDDQM Methodologies

Several methodologies provide process modeling languages to manage data quality. We examine how the methodologies are applied in PDDQM. Since the methodologies are highly heterogeneous regarding their phases and application, we focus on the presentation of the prevalent methodologies that are based on TDQM or the Plan-Do-Check-Act (PDCA) methodology (Table IV). The PDCA methodology for improving processes is prevalent in data-quality management (cf. [English 1999; Redman 1996; Wang 1998]). The TDQM methodology [Wang 1998] and the total information quality management (TIQM) methodology (originally known as the total quality data management (TQdM) methodology) [English 1999] are based on the PDCA cycle [Deming 1989].

The TDQM methodology consists of a four-step cycle aiming at the delivery of high-quality IPs. The TIQM methodology proposes a framework consisting of six processes for information-quality improvement. One of these processes is dedicated to the improvement of the process quality, following the PDCA cycle. The TDQM methodology focuses on the IP, whereas the TIQM methodology focuses on the process, embedding the improvement into a broader context.

Although TDQM is based on the PDCA cycle, three articles refer to both methodologies. However, only in one of these cases are the phases of the PDCA cycle applied, whereas the information-quality survey is based on TDQM [Davidson et al. 2004]. In the other two cases, a six-step methodology is used [Kovac et al. 1997; Kovac and Weickert 2002]. The other methodologies referring to TDQM are customized as well. Only in three cases are the TDQM phases explicitly applied. In the first case, the phases include tools and group sessions [Wijnhoven et al. 2007]. In the second case, the customer support data quality (CSDQ) methodology is presented, emphasizing user focus [Keenan and Simmons 2005]. The TDQM phases are kept, and further methodologies for each phase are proposed. In the third case, the TDQM phases are incorporated into an action research cycle [Lee et al. 2004]. Although this results in a five-step methodology, the TDQM phases are still included. In the other cases, the methodologies are customized as well [Helfert and von Maur 2001; Mielke 2005; Wang et al. 2002]. Although TDQM proposes the IMS for process modeling, in only two articles are the modeling languages at least derived from the IMS or IP-MAP [Davidson et al. 2004; Wijnhoven et al. 2007]. Similar to the methodologies, the process modeling languages—DFDs as well as PFCs—seem to be customized depending on the specific situation and needs. Although TDQM is based on the PDCA cycle and both are based on a four-step cycle, the applied phases vary. TDQM is referred to especially as a general quality concept, without being implemented as a specific methodology [Helfert and von Maur 2001; Kovac et al. 1997; Kovac and Weickert 2002]. In other cases, TDQM is extended regarding the phases [Lee et al. 2004], methods included [Keenan and Simmons 2005], and process modeling languages [Mielke 2005]. Examining the applied process modeling languages provides a similar picture. Even when referring to the IMS or IP-MAP (within the TDQM methodology), heterogeneous process modeling languages can be observed [Davidson et al. 2004; Mielke 2005; Wang et al. 2002]. In the primary studies, the COLDQ methodology and the embedded information chain map (cf. Section 2.3) are not used at all. The application of the process modeling languages will be examined in the next section in detail.

4.4. Process Modeling Languages

We consider process models, that is, activity-centric models [Recker et al. 2009] and differentiate between PFCs and DFDs, referring to Shankaranarayanan and Wang [2007]. They compare IP-MAPs to other modeling languages with respect to a possible substitution or complementation of the IP-MAP. PFCs represent the sequence of process steps without the dataflow, and DFDs represent the flow of data without the sequence of the process steps. We use this simplifying categorization, since organizations apply enhanced models and the use of the process modeling languages varies. Even differentiating between DFDs and PFCs is not always clear-cut. For example, some PFCs contain further elements such as databases or repositories without presenting the flow of data [Helfert and von Maur 2001; Meyer and Zack 1996; Zack 1996]. Other process models specifically focus on other aspects, although providing activities and data or control flows as well (e.g., the Viable System Model specifies functional criteria within organizational systems [Laumann and Rosenkranz 2008]).

The cause-and-effect diagram (CED) [Ishikawa 1993] is another modeling language used. It was introduced to improve the quality of production processes. One or more quality characteristics is set as effect(s) (e.g., [Harkness et al. 1996; Klesse et al. 2004]). The main factors that influence the quality characteristics are set as causes (e.g., materials, workers). Although the CED is concerned with process improvement, it focuses on causes and effects and does not include activities or process steps.

Table V. Applied Process Modeling Languages and Their Representational Characteristics

Article	Process modeling language (no. of figure in article)	Swim lane	Time axis	Sequence	Data quality element
[Balka et al. 2012]	PFC (4)	—	—	X	—
[Ofner et al. 2012]	PFC (6)	X	—	X	quality check
[Dejaeger et al. 2010]	PFC (2)	X	—	X	—
[Xie and Helfert 2010]	PFC (3)	—	—	(X)	—
[Gaynor and Shankaranarayanan 2008]	IP-MAP (1)	—	—	X	—
[Hakim 2008]	PFC (3)	—	—	X	—
[Laumann and Rosenkranz 2008]	PFC (1)	—	—	X	—
	DFD (2)	—	—	(X)	—
[Lee et al. 2007b]	context-embedded IP-MAP (2-7)	X	X	X	quality check
[Thi and Helfert 2007]	IP-MAP (2)	—	—	X	quality check
	PFC (3)	—	—	X	quality check
	DFD (7)	—	—	(X)	—
[Shankaranarayanan and Cai 2006]	IP-MAP (1)	—	—	X	—
[Keenan and Simmons 2005]	DFD (1)	—	—	(X)	—
[Mielke 2005]	PFC (1)	X	(X)	X	—
	process meta model (6)	X	—	—	quality dimensions
	DFD (7)	—	—	X	quality dimensions, quality metrics
	DFD (8)	X	—	X	—
	PFC (9)	—	—	X	quality dimensions, quality metrics
[Davidson et al. 2004]	DFD (1-3)	—	—	X	—
	DFD (4)	—	—	—	—
[Klesse et al. 2004]	PFC (3,6)	X	—	X	—
	CED (4-5)	—	—	X	generic quality types in life cycle
[Shankaranarayanan et al. 2003]	IP-MAP (1-3)	—	—	X	quality check
[Katz-Haas and Lee 2002]	DFD (6)	—	—	—	—
	DFD (7)	—	—	—	quality dimension (timeliness)
[Kovac and Weickert 2002]	PFC (1)	X	—	(X)	—
	PFC (2)	X	—	X	—
[Wang et al. 2002] (IP-MAPs cf. [Shankaranarayanan et al. 2003])	PFC (3)	—	—	X	—
[Helfert and von Maur 2001]	PFC (5)	X	—	X	quality check
[Kahn et al. 2001]	DFD (3)	—	—	—	—
[Millard and Lavoie 2000]	use case diagram (1)	—	—	X	quality check
	state diagram (2)	—	—	—	quality check

Table V. Continued

Article	Process modeling language (no. of figure in article)	Swim lane	Time axis	Sequence	Data quality element
[Ballou et al. 1998]	IMS (7)	—	—	X	quality check
[Kovac et al. 1997]	PFC (6)	X	—	X	quality check
[Harkness et al. 1996]	CED (3)	—	—	X	—
	PFC (4)	X	—	X	quality check
[Meyer and Zack 1996]	PFC (7)	—	—	(X)	—
	PFC (8)	—	—	(X)	—
[Zack 1996]	PFC (2)	—	—	(X)	—
	workflow diagram (3)	—	—	(X)	quality check
DFD = Data Flow Diagram; PFC = Process Flow Chart; CED = Cause-and-Effect Diagram					

Table VI. Representational Characteristics of Process Modeling Languages

Process modeling language	No. of models	Swim lane	Time axis	Sequence	Quality check	Quality dimension	Quality metrics
IMS, IP-MAP map	14	6	3	14	10	—	—
PFC	20	10	1	20	5	1	1
DFD	12	1	—	8	—	2	1
Total	46	17	4	42	15	3	2
IMS = Information Manufacturing System; PFC = Process Flow Chart; DFD = Data Flow Diagram							

In the following, we first give an overview of the process modeling languages applied for PDDQM in the particular articles. Table V extends Table I with a more detailed overview of the process modeling languages with respect to four characteristics (swim lane, time axis, sequence, and data-quality elements). IP-MAPs already provide a basis specifically for PDDQM, due to their focus on IP production (cf. Section 2.3). More recent research [Lee et al. 2007b] presents the context-embedded IP-MAP, enhancing the IP-MAP with swim lanes and a time axis. Therefore, we examine the use of swim lanes and the time-axis (also implying the sequence of process steps). We further examine the sequence since, as already pointed out, it is a major differentiation between PFCs and DFDs. Moreover, swim lanes impact the representation of the sequence, and the time-axis facilitates the visual representation of the quality dimension timeliness. To look beyond the IP-MAP to examine how data quality aspects are referred to in other models (PFC and DFD) or in customized IP-MAPs, we identify further elements for representing data quality. We discuss the characteristics in the subsequent sections, including an overview of the characteristics across the modeling languages (Table VI). In the following overview (Table V), we used parentheses if the characteristics are referred to but not applied in the model [Mielke 2005] or the characteristic is provided in a limited way [Keenan and Simmons 2005; Kovac and Weickert 2002; Meyer and Zack 1996; Xie and Helfert 2010; Zack 1996].

Application of Representational Characteristics. Swim lanes are applied throughout several process modeling languages (Table VI). In the context-embedded IP-MAP, they represent stakeholder groups involved into the IP process. The use of swim lanes in the PFCs and DFDs refers to departments [Kovac and Weickert 2002; Mielke 2005], external stakeholders [Harkness et al. 1996; Kovac et al. 1997; Kovac and Weickert 2002; Ofner et al. 2012], specific roles [Dejaeger et al. 2010; Klesse et al. 2004], systems and databases [Helfert and von Maur 2001; Kovac et al. 1997], and tasks [Harkness et al. 1996; Kovac and Weickert 2002]. Examples for possible variations within one model are provided [Dejaeger et al. 2010; Harkness et al. 1996; Kovac et al. 1997; Kovac and Weickert 2002]. The swim lanes include stakeholders, tasks, and products. In one of these cases, the swim lanes are applied in the rows and columns, constituting a matrix with stakeholders and tasks [Kovac and Weickert 2002]. Hence, the model includes additional information, but it is not possible to add a time axis. Further characteristics are the time axis and the logical sequence of steps. A time axis shows the time needed to conduct processes or process steps. The logical sequence shows the logical flow of the steps regarding the predecessor and successor relations. The time axis is usually represented by the X-axis and shows the flow of the process (in- or excluding data) from left to right [Lee et al. 2007b; Mielke 2005]. In this context, the sequence of the process steps is defined as well. However, most models do not include a time axis, but the sequence of the process steps.

All PFCs provide a sequence, although in some cases, the clarity of the sequence is inhibited by additional arcs or elements, which are not clearly separated from the control flow (e.g., [Meyer and Zack 1996; Xie and Helfert 2010]. An existing sequence fits the definition of PFCs. However, from most DFDs, the sequence of the processes can be derived as well (e.g., [Laumann and Rosenkranz 2008; Keenan and Simmons 2005]), although it is not inherent in this process modeling language. DFDs and PFCs seem to be combined in several cases (e.g., [Davidson et al. 2004; Keenan and Simmons 2005; Mielke 2005]).

Integrating Data Quality into Models. Although the presented models are applied within projects to assess or improve data quality, the integration of data quality directly into the models is rather rare. On the other hand, several possibilities exist to integrate data quality without using IP-MAPs. We structured the identified data-quality-specific elements into data-quality checks, quality dimensions, and quality metrics (cf. Tables V and VI). Focusing on the process model representation, we consider data-quality checks as modeling elements determining that some sort of data-quality check is carried out at one or more points in the process. It depends on the metadata if the data-quality check is merely an element with a label, some sort of task, or a more sophisticated modeling element specifically related to data-quality information. In the case of data-quality dimensions, specific quality dimensions are visibly incorporated into the representation of the process model. Data-quality metrics are quantifications of data quality.

In the presented models, the applied data-quality elements are mostly data-quality checks integrated into flow charts or diagrams as process steps [Kovac et al. 1997; Millard and Lavoie 2000; Thi and Helfert 2007; Zack 1996] or as specific elements that are attached to process steps [Helfert and von Maur 2001]. In one case, the data-quality checks are integrated as a swim lane, as the (intermediate) process deliverable is jointly agreed upon regularly between two parties [Harkness et al. 1996].

Within four models, data-quality checks are not dedicated model elements but tasks determining that data-quality checks take place [Harkness et al. 1996; Millard and Lavoie 2000; Zack 1996].

Data-quality checks within IMS or IP-MAPs are dedicated model elements which can be referred to the according meta data [Ballou et al. 1998; Shankaranarayanan et al. 2000, 2003]. Beside adapting the IMS or IP-MAP, further approaches exist, relating data-quality metadata to process models, not necessarily data quality checks [Helfert and von Maur 2001; Kovac et al. 1997; Mielke 2005; Ofner et al. 2012].

Further approaches exist into which the process models and included data-quality checks are embedded. Kovac et al. [1997] derive metrics for data-quality dimensions timeliness and accuracy, focusing on process hand-offs between stakeholders. Beside single tasks for checking data-quality, defined data quality measures between process hand-offs are indicated by dedicated arcs. Helfert and von Maur [2001] annotate modeling elements in a data delivery process. The numbered annotations refer to verbalized dataflow processes and are linked to data-quality dimensions, according data-quality indicators, and measuring points to the data delivery process elements. Ofner et al. [2012] provide a formalized meta model, building on BPMN, for assessing data quality related to process model tasks.

Two other approaches visibly integrate data quality into single process models without using data-quality checks. Katz-Haas and Lee [2002] focus on timeliness, since a process' cycle time led to delayed information provision, thus, causing high costs. To visualize why information does not arrive in a timely manner, they enhance a process model, assigning time stamps to process steps. Without using data-quality checks, Mielke [2005] provides quality dimensions and metrics to measure data quality within process models. Moreover, only Mielke [2005] integrates data quality across several models without applying IP-MAPs. At an abstract level, the model gives an overview of the most important data-quality dimensions for the main processes and departments. At a more detailed level, the subprocesses and their IP in- and outputs are provided. The data quality of each subprocess is determined, using weighted key performance indicators to measure data quality, based on the most important data-quality dimensions. The performance of the subprocesses is summed up to the process performance. Furthermore, the overall degree of data quality performance is calculated from weighted data quality across the processes.

5. REPRESENTATIONAL REQUIREMENTS FOR PDDQM

Based on the primary studies and the use of process modeling languages, we derive representational requirements for PDDQM. Since the requirements are derived from several different articles, they constitute an enhanced set of relevant elements for PDDQM, which needs to be examined in future research. Since the application of elements within process modeling languages is a subject to interpretation (cf. [Recker et al. 2010]), in a first step, we describe the identified elements with respect to their possible redundancy or complementary use with respect to the IP-MAP's basic modeling elements (cf. Figure 3 within bold frame). We build on the IP-MAP, since it is applied consistently across the primary studies, and its elements are provided within a defined process modeling language. Furthermore, the IP-MAP was specifically developed for the manufacturing process of an IP. However, this specificity may also inhibit the possibilities for process modeling. We include the requirements from Section 4.4 as enhancements to the IP-MAP (cf. Table V). Additionally, we derive further elements from the primary studies that we consider relevant for PDDQM and which will be addressed in more detail. Figure 3 summarizes the requirements based on the primary studies within their context-specific representation. It shows if they are complementary or redundant with respect to the IP-MAP elements. We omitted elements that are not specific for data representation (e.g., events, general textual annotations).

We already presented the application of swim lanes, time axis, and of a given sequence. Since swim lanes can be used in several ways, they can serve as organizational and IS boundaries in the IP-MAP. If both elements are given within one model or within one process modeling language, the possible application should be clearly defined to mitigate redundancy of swim lanes and the boundary elements and to allow for a complementary use.

Assuming a given sequence, a time axis can be included within the IP-MAP, where the sequence of process steps is provided by the data-flows. That is, the data-flow

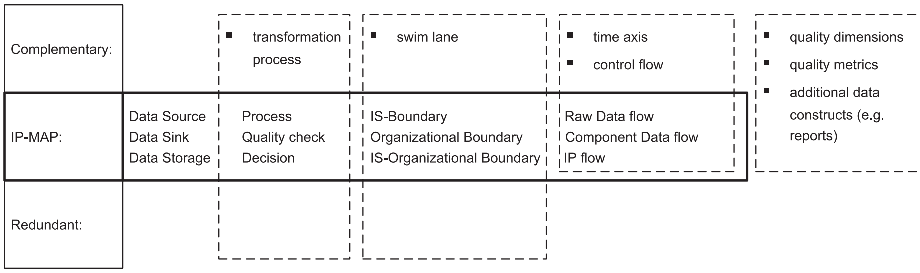

Fig. 3. Process modeling language requirements.

substitutes the control flow, since each process represents necessary processing steps to produce the IP. Therefore, the data output of a process step is the input of the following process step. However, control flows can be applied complementary. The combination of both types of flows is used in customized [Kovac and Weickert 2002; Xie and Helfert 2010] and standardized process models [Ofner et al. 2012]. Therefore, the control flow would allow for additional modeling possibilities enhancing the IP-MAP, for instance, representing process steps not directly involved in IP processing.

Regarding data-quality-specific elements, Table V already provides quality checks and quality dimensions, partially extended by quality metrics. Quality dimensions and quality metrics facilitate to focus on specific quality dimensions or to prioritize importance of quality dimensions or IPs. Within the IP-MAP, quality dimensions and metrics are not included visibly as elements (cf. Table VI). In contrast, quality checks are represented by specific elements within the IP-MAP. Within other process modeling languages, quality checks are represented through general-process step elements.

Further elements for representing data within process models might be necessary in addition to the IP-MAP elements. In one case, elements represent ‘core information’ (e.g., needed to start a subprocess) and ‘reports’ [Mielke 2005]. The several identified types of IPs (cf. Table III) support the potential need for additional data constructs. Although most IPs constitute reports, they have various purposes and therefore different characteristics. Consequently, a detailed definition and representation of IPs is necessary for a clear differentiation within organizations. Finally, Mielke [2005] provides an element for representing a data transformation process, which would be redundant within the IP-MAP since it focuses on data production and no differentiation between data and control flows is given. In contrast, in the given example, there is only one connection, a directed arrow, for linking data elements with processes and processes among each other. Hence, the process model provides a sequence with events and process steps, rather focusing on control flows additionally embedding data processing. In contrast to the IP-MAP, the process model does not seem to trace the production of a specific IP. Here again, the differentiation of data and process flow might be important for considering additional process steps not involved in data processing, allowing for broader application of the IP-MAP. In such a context, an additional element for data transformation processes could be used complementarily.

Another difference, specifically between DFDs and IP-MAPs, is the inclusion of data objects or even data elements within models. Instead of referring to dataflows with the predefined naming of the IP-MAP (raw data (RD), component data (CD), IP), data objects’ names can be included within the model. The inclusion of data elements as textual annotations has to be considered similarly to the data quality dimensions, whether if these information should be included within the model.

Table VII. Process Modeling Language and Process Type

	Process type		
Process modeling language	**P**	**S**	**IP**
IMS, IP-MAP	—	1	11
PFC	2	6	9
DFD	1	4	4
IMS = Information Manufacturing System; PFC = Process Flow Chart; DFD = Data Flow Diagram Deliverables of process: P = tangible product; S = service; IP = information product			

6. DISCUSSION

6.1. Data-Quality Management at the Process and Organizational Level

In order to provide the context for our RQs, we examined the types of processes and organizations at which the PDDQM efforts aim. Since the process type does not necessarily depend on the type of the organization and IP processes are present regardless of the organization's type (cf. Section 4.1 and Table II), we examined the processes in more detail. However, for PDDQM to be successful in the long term, we argue that it is necessary to manage data quality at both levels. The importance of establishing data-quality management at an organizational level is addressed in extant research (e.g., [Ana 2010; Otto 2011; Weber et al. 2009]).

We assumed that the process type an influence on the applied process modeling language (cf. Section 4.2). However, if we examine which process modeling languages are applied to model the process types, we do not get a clear picture (Table VII). No process modeling language is exclusively used for a specific process type. Even a process delivering a tangible product can be consequently modeled from an IP perspective, respectively, from an internal information consumer perspective [Thi and Helfert 2007]. Although IP-MAPs and related languages are only applied in one case to model a service process, this might be due to our literature selection, since most processes are IP processes. Regarding the IP processes, no clear pattern can be derive, because different languages are applied to model IP processes. In contrast to our assumption, we argue that the process modeling language influences how a process is perceived, especially due to separating (tangible) products from services (cf. Section 3.3). For example, the delivery of an IP could be customized by the consumer (cf. [Zack 1996]), constituting a service, or reports are not solely delivered but additionally bounded to a consulting-service (cf. [Kovac et al. 1997]).

6.2. Methodologies and Process Modeling Languages

Methodologies for PDDQM are highly customized, and even if particular methodologies are referred to, the phases and methods applied vary (cf. Section 4.3 and Table IV). The context dependency of data quality also leads to several conceptualizations and methodologies for assessing and improving data quality in specific contexts (e.g., [Knight 2011; Lin et al. 2007]). Besides using already contextspecific methodologies, another possibility for adjusting methodologies to a specific organizational context is to provide a customizable toolbox with several methods across, for instance, the TDQM phases (cf. CSDQ methodology [Keenan and Simmons 2005]). The process modeling languages are applied heterogeneously across methodologies' phases as well. A determination of the process modeling language based on the data-quality methodology cannot be derived. Regarding the rather low data-quality management maturity across sectors [Aiken et al. 2007; Glowalla and Sunyaev 2012b], further research is necessary to identify adequate data-quality management approaches for organizations.

Moreover, most data-quality management efforts are reactive, and the identification of root causes for data quality issues is prevalent (cf. [Balka et al. 2012; Eppler

2001; Harkness et al. 1996; Kahn et al. 2001; Katz-Haas and Lee 2002; Klesse et al. 2004; Sulong et al. 2012; Xie and Helfert 2010]). This focus is not surprising, since the causes for low data quality may not be simply linked to visible problems. But besides seeking to identify root causes of errors and to eliminate them, PDDQM also proposes sustaining the improvements. Sustaining or continuously improving data quality is important (e.g., [Dravis 2005; Keenan and Simmons 2005]). Therefore, successful reactive data-quality efforts should be used as a basis for the development of proactive methodologies. An example is provided by Harkness et al. [1996], where the success of reactive methodologies led to the development of a proactive methodology to assure data quality for newly developed processes. In such cases, it would be necessary not only to keep up improvement on specific processes, but to standardize data-quality approaches. Another example for the need of a proactive methodology is the foundation of a new organization [Kovac and Weickert 2002]. A major step towards proactive PDDQM is the integration of the data-quality perspective into broadly-used process modeling languages (e.g., [Ofner et al. 2012]). Such enhanced process modeling languages may build awareness and provide a basis for addressing data quality during process design. Although not considered in detail in this article, CEDs might be helpful for identifying root causes for specific data-quality issues. Specific process models for PDDQM would support identification of reasons for poor data quality and to provide additional context, for instance, about involved shareholders and systems. Finally, data-quality research identifies generic root causes for data-quality issues [Eppler 2001; Lee 2006; Liu and Chi 2002;]. The presented generic principles for improving data-quality constitute another step towards a proactive methodology for data-quality management. Using general principles instead of the numerous data-quality criteria could decrease complexity of data-quality management.

6.3. IP-MAP and Its Benefit

With the results showing the possibilities of integrating data quality into existing models, the benefit of IP-MAPs could be questioned. Whether data quality is integrated marginally or in a holistic way, both seem possible with the enhancement of existing models or process modeling languages (e.g., cf. [Klesse et al. 2004; Mielke 2005]). However, the perspective of IP-MAP brings important advantages. IP-MAP focuses on the delivery of a specific IP and, similar to a PFC, on the necessary sequential steps to manufacture such an IP. Additionally, it presents the necessary (raw, component) data and its sources. 'Necessary' means that the presented dataflow is limited to the purpose of producing the IP. The focus on a specific IP and its manufacturing process makes it possible to add further elements (databases, stakeholders, pre- and post-conditions of data- or information products etc.) into the model and reduce abstraction while keeping an understandable model. Furthermore, other IP-specific information can be linked more easily, whether within the model or referenced in additional documentation, for example, the IMS analysis matrix [Ballou et al. 1998] or metadata [Shankaranarayanan et al. 2000].

6.4. Integration of Different Process Models

Process modeling languages are applied heterogeneously. They are enhanced (e.g., [Katz-Haas and Lee 2002; Lee et al. 2007b]) or combined (e.g., [Davidson et al. 2004; Mielke 2005]). This heterogeneity poses a problem in terms of tool support (e.g., [Fahland et al. 2011; Koschmider et al. 2011]) and understanding between stakeholders with customized process modeling languages. However, using mature and known models and adjusting them to emerging needs might be useful in terms of model acceptance and understanding within an organization. Furthermore, a model can provide several perspectives on a single process (e.g., cf. [Katz-Haas and Lee 2002; Klesse et al. 2004; Mielke 2005]). Thus, we consider the integration of different models an important topic, especially since data quality can be integrated into existing process modeling languages (cf. Table VI) and IMS and IP-MAPs seem to be used rather seldomly, even when applying TDQM (Table IV). In this context, integration does not necessarily refer to the use of one model for all purposes, but to keep up comprehensibility across different and possibly customized models. Comprehensibility is essential, since process models are widely used for documentation and communicating processes [Bandara et al. 2005; Davios et al. 2006]. Process stakeholders must have sufficient knowledge about the context, that is, to know 'what' data are collected 'how' and 'why' in order to solve data-quality problems [Lee 2003; Lee and Strong 2003]. The primary studies show some examples of limited model integration [Davidson et al. 2004; Kovac and Weickert 2002; Klesse et al. 2004]. The integration ranges from referring to process steps throughout models by numbering them [Klesse et al. 2004] through presenting the same objects from different models within one integrating model with a different perspective [Davidson et al. 2004] to specifying processes from a more abstract view [Kovac and Weickert 2002]. In the last case, although swim lanes are applied, the stakeholder naming differs, and the presented specific process is not visible in the overview model. More consequent integration is provided as well [Helfert and von Maur 2001; Mielke 2005], applying overview models to integrate models into a broader context and to derive and use swim lanes, processes, and data-quality elements consistently.

6.5. Integration of Data Quality into Process Models and Languages

We identify several requirements from the customized process modeling languages that can extend the IP-MAP (cf. Figure 3). These requirements either allow for a broader application of the IP-MAP, for instance, including control flows, or for a more specific application, for instance, including elements to refine the representation of data. However, as pointed out previously, the customized models should be integrated. Providing integration may be a reason why organizations stick to existing process modeling languages for PDDQM. This leads to the question of whether the benefits of the IP-MAP could be realized, respectively integrated, within an already broadly accepted process modeling language. Despite upcoming research (e.g., [Cappiello et al. 2013; Ofner et al. 2012]), the existing lack of application of current process modeling languages for PDDQM in which data and control flows can be modeled, is a research gap. Other modeling elements, such as the several kinds of gateways, could have an impact on process modeling from a data-quality perspective. Adequate definition and representation of information is necessary for efficient and, for instance, automated processes [Glowalla and Sunyaev 2012a; Zhao et al. 2012]. New requirements arise for process modeling from an information perspective. Representational analysis using the Bunge-Wand-Weber representation model would allow for an overlap analysis [Green et al. 2007; Weber 1997]. That is, comparing IP-MAPs with other process modeling languages with respect to a minimal redundancy of elements and a maximum completeness for the representation of real-world phenomena. Furthermore, such an analysis should take into account the additional requirements and elements identified in our study (cf. Figure 3). These requirements represent the organization-specific level of representational analysis, that is, how organizations apply a process modeling language (cf. [Rosemann et al. 2009]). The adequate representation is important in order to cover practitioners' needs and to keep the process modeling language understandable to facilitate adoption.

Given the identified primary studies, we focus on the integration of data quality into process model instantiations. A visible integration of data quality into process models facilitates awareness and understanding of data-quality issues within processes across stakeholders, including non-modeling experts. The IP-MAP allows visible integration, facilitating communication across stakeholders (modeling and non-modeling experts) while providing a data-quality metamodel for a sophisticated definition of data-quality metrics [Shankaranarayanan et al. 2003]. If integrated into a specific process modeling language, the metamodel needs to define the possible application of the data-quality check within the process model instantiations. Moreover, allowing for process automation, continuous control, and tool support, a structured approach with a clearly defined meta-model is necessary. Ofner et al. [2012] provide such a sophisticated enhancement of BPMN for the integration of data-quality information. A clear visualization of data-quality aspects within the process model is neglected, but the extension of the BPMN metamodel is a complementary approach to the visible integration of data quality. Moreover, BPMN may leverage familiarity due to its broad application. Cappiello et al. [2013] build on BPMN as well and develop a data-quality-aware process design methodology. Their process-driven methodology supports linking data-quality issues to root causes and defining improvement activities. The integration of data quality into broadly-applied process modeling languages is at an initial but promising stage and would allow for a proactive approach to PDDQM right from the process design.

Both approaches, integrating data quality visibly into process model instantiation and building on clearly defined process modeling languages supporting PDDQM, are complementary and should be applied depending on the modeling purpose. The visible integration facilitates communication across stakeholders, especially if non-modeling experts are involved. For instance, a conceptual model with a broad set of data-quality dimensions can be used for an intuitive process-driven approach, allowing an in-depth exploration of the production and control of critical IPs [Glowalla et al. 2014]. Such a pragmatic approach is adequate for exploring unstructured or knowledge-intensive processes and might support process design in early stages. The final design, implementation, and ongoing control of processes should be supported by clearly defined process modeling languages in order to assure correct process models, process execution, and high data quality with a long-term perspective.

6.6. Limitations

Our literature review aims at a detailed analysis of process modeling languages and their application. Therefore, we considered articles describing such an application of process-driven data quality. This rather specific selection led to a small number of articles. Hence, a generalization beyond the identified articles regarding the application of process modeling languages and methodologies is not possible.

The research approaches in the articles are heterogeneous, that is, rather explorative approaches, deriving (enhanced) methodologies for data-quality management in contrast to the use of existing methodologies. Since each article focuses on specific aspects of a case, it remains open for further research how far the described methodologies and process modeling languages reflect real-world cases, especially since in some articles, the presented methodologies or phases might be used only to structure the article. Being aware of the limited possibility for making general statements, we focused on the heterogeneity and the many possibilities that arise from their detailed exploration.

Finally, we explicitly excluded processes that are inherent to IT-systems (e.g., the optimization of data warehouse internal processes). It should be mentioned that, for instance, TDQM is applied in this context as well (e.g., [Nadkarni 2006]). This corroborates the view that the methodologies can be used as generic concepts and that it is necessary to provide methods for specific applications.

7. CONCLUSION

7.1. Implications for Practice and Research

Regarding RQ1, we provide a detailed synthesis of how organizations apply process modeling languages within PDDQM. The organization or process types do not determine the process modeling language. Regarding PDDQM methodologies, TDQM is most often referred to. However, the phases and methods (including process modeling languages) that are defined by the methodology are seldom applied. TDQM is used as a general concept to improve data quality, to structure data-quality programs at an abstract level, or it is applied as a data-quality methodology to improve specific processes. Although IMS is part of TDQM, we cannot derive a dependency of the use of TDQM and of IMS or its extension, the IP-MAP. Organizations customize existing process modeling languages and do not prefer a specific process modeling language for PDDQM.

Regarding RQ2, we identified process modeling requirements for PDDQM, examining instantiations of process modeling languages. Process modeling with an emphasis on data quality can be realized in several ways. The IP-MAP focuses specifically on the production of one IP. However, our results show possibilities of either broadening or refining the application of IP-MAPs. Moreover, our review presents process modeling languages that already provide useful characteristics for PDDQM and are used in several organizations. Either way, it is necessary to consider the integration of different process modeling languages and their organization-specific application to fulfill practitioners' needs.

Using Existing Process Models for an Initial Engagement in Data Quality. In current practice, the data-quality perspective can be integrated within and across PFCs and DFDs. The models' differentiation is vague, since the process models are enhanced or combined. How an existing process model can be enhanced depends on the existing representational characteristics. In order to support adoption, the models should be kept familiar for stakeholders. Then, enhancing existing process models creates awareness for data quality without radical changes in the existing model landscape and counteracts the problem of missing data-quality management.

Defining Critical IPs to Apply IP-Centric Process Models. An organization which seeks to improve its processes with respect to data quality should be aware of its main internal and external deliverables and clearly classify them, especially if the deliverable is an IP that can be delivered in several ways or is attached to a service. The application of IP-MAPs, or at least an IP production perspective, depends on the importance of the IP and the amount of information that should be represented. We argue that enhancing mature models with additional data-quality information is a first step towards this IP definition, especially if data-quality problems occur that cannot be attributed to a single IP. IP-MAPs can yield further advantages if data quality efforts are concentrated on specific (critical) IPs. Then the product perspective is helpful for describing the IP and its flow within a broad scope at a detailed level [Glowalla et al. 2014]. This application implies an iterative approach to data-quality management and allows for leveraging the organization's proven process modeling languages.

Bridging the Organizational and Process Level with a PDDQM Methodology. For sustainable PDDQM, methodologies, for instance, TDQM and CSDQ, bridge the gap between the organizational and process level, providing general guidelines and techniques for data-quality improvement to align process improvements. Moreover, continuous application and improvement of a PDDQM methodology facilitate transition from initial, reactive PDDQM to proactive approaches (Section 6.2).

The encountered heterogeneity of the application and customization of methodologies (cf. Section 4.3) inhibits general recommendations about which methodology should be used in what way. Overall, organizations tailor methodologies and included methods to their specific needs. Existing methods and process modeling languages might be incorporated to lever familiarity instead of switching to new and unknown ones. Further research is needed to examine how to use existing methodologies to meet an organizations' requirements for PDDQM, including customizable methodologies that already suggest several methods within, for instance, a toolbox.

Identifying or Developing an Adequate Process Modeling Language for PDDQM. The integration of different process models is an important and apparently underestimated issue. Every production of tangible products and services carries information that is somewhere created, updated, and maintained. Therefore, organizations need to consider several different processes and their models for IP production. Since IP-MAPs are still evolving and other process modeling languages are used for PDDQM, the integration of IP-MAPs and other broadly used process modeling languages should be addressed (e.g., [Cappiello et al. 2013; Lee et al. 2007a; Ofner et al. 2012]). With respect to upcoming application of current process modeling languages with a data-quality focus, for instance, BPMN [Cappiello et al. 2013; Ofner et al. 2012], we see a research gap from the data-quality perspective in process modeling. The Bunge-Wand-Weber representation model allows comparing IP-MAPs with other process modeling languages [Rosemann et al. 2009] to provide a meaningful combination of languages or even a substitution for IP-MAPs. An example for complementing the IP-MAP is the specific modeling element for reports applied in Mielke [2005], especially since most IPs can be categorized as reports (cf. Table III). To cover practitioners' needs and further support the adoption of a new or enhanced process modeling language, the identified requirements (Figure 3) are a basis for enhanced use and further research. However, for improving automation, control, and tool support of process modeling languages (e.g., [Fahland et al. 2011; Koschmider et al. 2011; Meda et al. 2010]) a substitution might be desirable.

The integration of elements into process model instantiations and the development of process modeling languages supporting PDDQM are both feasible approaches and should be applied according to the modeling purpose. BPMN might be an adequate platform to bridge the gap between a visible and pragmatic approach for process exploration and communication as well as documenting final processes.

The potential negative impact on ontological clarity needs to be examined if a process modeling language, such as BPMN, provides high coverage and thus potential use of redundant modeling elements (cf. [Recker et al. 2010; Wand and Weber 1993, 1995]). Additionally, understandability and complexity need to be ensured in process modeling [Reijers and Mendling 2011], especially if enhancing process models, for instance, with data-quality aspects [Glowalla and Sunyaev 2013b]. Besides the process modeling language itself [Recker 2010b], the familiarity with existing process modeling languages affects its use [Recker 2010a]. Research on model understandability or complexity, respectively, needs to assess the right trade-off between familiarity and usefulness of a new or enhanced process modeling language and its instantiations. Several factors have to be considered, which could be categorized as contextual factors [Rosemann et al. 2008], personal factors related to the model reader, and factors related to the model itself [Reijers and Mendling 2011]. Quality frameworks, pragmatic guidelines, and process model quality metrics [Mendling et al. 2010] could be applied to keep models understandable after integration of data-quality elements [Glowalla and Sunyaev 2013b]. Both approaches—enhancing an existing or developing a new process modeling language—ask for further, differentiated research. Practitioners and researchers should be aware of both approaches and according trade-offs to allow for a holistic PDDQM methodology in the long term.

7.2. Contribution and Impact

The article represents a thorough literature review and analysis of PDDQM which, to our knowledge, have not been done before.

For scholars, this critical literature review is a sound basis of the body of knowledge in PDDQM process modeling. According to our results, there is much potential for further research in order to develop a methodology, with tools and techniques for holistic and sustainable PDDQM. The results provide researchers with a detailed view on the application of methodologies and process modeling languages at the organizational and process level. The results show current gaps, further requirements, and trade-offs, questioning the adequacy and operationalization of extant methodologies and process modeling languages. Moreover, by presenting potential future research for developing and validating process modeling languages, this article guides scholars who aim to improve PDDQM.

The results show practitioners which and how methodologies and process modeling languages are applied for PDDQM. While supporting choosing solutions for their specific organizational setting, it also sensitizes practitioners to the fact that currently no standard solutions are available. A label to a methodology like TDQM might refer to a plethora of potential application choices and customizations. We enable practitioners to scrutinize the application of methodologies and process modeling languages for their PDDQM efforts. The presented results increase awareness of possibilities and of potential effects when tailoring PDDQM methodologies and process modeling languages to their context-specific needs.

APPENDIX

These are the 74 Journals used in the literature review. Academy of Management Journal; Academy of Management Review; ACM Computing Surveys; ACM Journal of Data and Information Quality; ACM Transactions on Computation Theory; ACM Transactions on Computer Systems; ACM Transactions on Computer-Human Interaction; ACM Transactions on Database Systems; ACM Transactions on Information and System Security; ACM Transactions on Information Systems; ACM Transactions on Intelligent Systems and Technology; ACM Transactions on Internet Technology; ACM Transactions on Knowledge Discovery from Data; ACM Transactions on Modeling and Computer Simulation; ACM Transactions on Multimedia Computing Communications and Applications; ACM Transactions on Reconfigurable Technology and Systems; ACM Transactions on Software Engineering and Methodology; Administrative Science Quarterly; AI Magazine; Artificial Intelligence; Business Process Management Journal; California Management Review; Communications of the ACM; Communications of the AIS; Computers and Operations Research; Data & Knowledge Engineering; Decision Sciences; European Journal of Information Systems; Harvard Business Review; Human-Computer Interaction; IEEE Computer; IEEE Software; IEEE Transactions on Communications; IEEE Transactions on Computers; IEEE Transactions on Engineering Management; IEEE Transactions on Evolutionary Computation; IEEE Transactions on Industrial Informatics; IEEE Transactions on Knowledge and Data Engineering; IEEE Transactions on Reliability; IEEE Transactions on Services Computing; IEEE Transactions on Software Engineering; IEEE Transactions on Systems, Man, and Cybernetics; IEEE Transactions on Systems, Man, and Cybernetics, Part A: Systems and Humans; IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics; IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews; Information & Management; Information

Processing & Management; Information Systems; Information Systems Frontiers; Information Systems Journal; Information Systems Research; Information Technology and People; Informing Science; International Journal of Electronic Commerce; International Journal of Information Quality (searched manually); Journal of Computer and System Sciences; Journal of Database Management; Journal of Global Information Management; Journal of Global Information Technology Management; Journal of Information Technology; Journal of Management Information Systems; Journal of Strategic Information Systems; Journal of the ACM; Journal of the AIS; Journal on Computing; Management Science; MIS Quarterly; Operations Research; Organization Science; Sloan Management Review; The DATABASE for Advances in Information Systems.

The following journals were not available and therefore not included: Database Programming and Design; Journal of Management Systems; Journal of Information Management; MISQ Discovery.

REFERENCES

Aiken, P., Allen, M., Parker, B., and Mattia, A. 2007. Measuring data management practice maturity: A community's self-assessment. *Computer 40*, 4, 42–50.

Alavi, M. and Leidner, D. E. 2001. Review: Knowledge management and knowledge management systems: Conceptual foundations and research issues. *MIS Quart. 25*, 1, 107–136.

Ana, L. 2010. Corporate data quality management in context. In *Proceedings of the International Conference on Information Quality*, 1–19.

Balka, E., Whitehouse, S., Coates, S. T., and Andrusiek, D. 2012. Ski hill injuries and ghost charts: Sociotechnical issues in achieving e-Health interoperability across jurisdictions. *Inf. Syst. Frontiers 14*, 1, 19–42.

Ballou, D., Wang, R., Pazer, H., and Tayi, G. K. 1998. Modeling information manufacturing systems to determine information product quality. *Manag. Sci. 44*, 4, 462–484.

Bandara, W., Gable, G. G., and Rosemann, M. 2005. Factors and measures of business process modelling: Model building through a multiple case study. *Euro. J. Inf. Syst. 14*, 4, 347–360.

Batini, C., Cappiello, C., Francalanci, C., and Maurino, A. 2009. Methodologies for data quality assessment and improvement. *ACM Comput. Surv. 41*, 3, 1–52.

Boisot, M. and Canals, A. 2004. Data, information and knowledge: Have we got it right? *J. Evol. Econ. 14*, 1, 43–67.

Buhl, H. U., Röglinger, M., Stöckl, S., and Braunwarth, K. S. 2011. Value orientation in process management. *Bus. Inf. Syst. Eng. 3*, 3, 163–172.

Cappiello, C., Caro, A., Rodríguez, A., and Caballero, I. 2013. An approach to design business processes addressing data quality issues. In *Proceedings of the European Conference on Information Systems*.

Davidson, B., Lee, Y. W., and Wang, R. 2004. Developing data production maps: Meeting patient discharge data submission requirements. *Int. J. Healthcare Technol. Manag. 6*, 2, 223–240.

Davies, I., Green, P., Rosemann, M., Indulska, M., and Gallo, S. 2006. How do practitioners use conceptual modeling in practice? *Data Knowl. Eng. 58*, 3, 358–380.

Dehnert, J. and van der Aalst, W. M. 2004. Bridging the gap between business models and workflow specifications. *Intern. J. Coop. Inform. Syst. 13*, 3, 289–332.

Dejaeger, K., Hamers, B., Poelmans, J., and Baesens, B. 2010. A novel approach to the evaluation and improvement of data quality in the financial sector. In *Proceedings of the International Conference on Information Quality*.

Deming, W. E. 1989. *Out of the Crisis*. MIT Press, Cambridge, MA.

Dravis, F. 2005. The IQ solution cycle. In *Proceedings of the International Conference on Information Quality*. 1–12.

English, L. P. 1999. *Improving Data Warehouse and Business Information Quality: Methods for Reducing Costs and Increasing Profits*. Wiley, New York.

Eppler, M. J. 2001. A generic framework for information quality in knowledge intensive industries. In *Proceedings of the Conference on Information Quality*. 329–346.

Even, A. and Shankaranarayanan, G. 2007. Utility-driven assessment of data quality. *Data. Adv. Inf. Syst. 38*, 2, 75–93.

Fahland, D., Favre, C., Koehler, J., Lohmann, N., Völzer, H., and Wolf, K. 2011. Analysis on demand: Instantaneous soundness checking of industrial business process models. *Data Knowl. Eng. 70*, 5, 448–466.

Gaynor, M. and Shankaranarayanan, G. 2008. Implications of sensors and sensor-networks for data quality management. *Int. J. Inform. Qual. 2*, 1, 75–93.

Glowalla, P. and Sunyaev, A. 2012a. A process management perspective on future ERP system development in the financial service sector. *AIS T. Enterprise Syst. 3*, 1, 18–27.

Glowalla, P. and Sunyaev, A. 2012b. Process-driven data and information quality management in the financial service sector. In *Proceedings of the Americas Conference on Information Systems.*

Glowalla, P. and Sunyaev, A. 2013a. Managing data quality with ERP systems - Insights from the insurance sector. In *Proceedings of the European Conference on Information Systems.*

Glowalla, P. and Sunyaev, A. 2013b. Process-driven data quality management – Integration of data quality into existing process models. *Bus. Inf. Syst. Eng. 5*, 6, 433–448.

Glowalla, P., Balazy, P., Basten, D., and Sunyaev, A. 2014. Process-driven data quality management – An application of the combined conceptual life cycle model. In *Proceedings of the 47th Hawaii International Conference on System Sciences*. IEEE Computer Society.

Gorla, N., Somers, T. M., and Wong, B. 2010. Organizational impact of system quality, information quality, and service quality. *J. Strategic Inf. Syst. 19*, 3, 207–228.

Green, P., Rosemann, M., Indulka, M., and Manning, C. 2007. Candidate interoperability standards: An ontological overlap analysis. *Data Knowl. Eng. 62*, 2, 274–291.

Hakim, L. A. 2008. Modelling information flow for surgery management process. *Int. J. Inform. Qual. 2*, 1, 60–74.

Harkness, W. L., Segars, A. H., and Kettinger, W. J. 1996. Sustaining process improvement and innovation in the information services function: Lessons learned at the Bose Corporation. *MIS Quart. 20*, 3, 349–368.

Haug, A., Arlbjørn, J. S., and Pedersen, A. 2009. A classification model of ERP system data quality. *Ind. Manag. Data Syst. 109*, 8, 1053–1068.

Haug, A., Zachariassen, F., and van Liempd, D. 2011. The costs of poor data quality. *J. Ind. Eng. Manag. 4*, 2, 168–193.

Helfert, M. and von Maur, E. 2001. A strategy for managing data quality in data warehouse systems. In *Proceedings of the Conference on Information Quality*, Cambridge. 62–76.

Ishikawa, K. 1993. *Guide to Quality Control*. Asian Productivity Organization, Tokyo.

Kahn, B. K., Katz-Haas, R., and Strong, D. M. 2001. Organizational realism meets information quality idealism: The challenges of keeping an information quality initiative going. In *Proceedings of the Conference on Information Quality.* 20–32.

Kahn, B. K., Strong, D. M., and Wang, R. Y. 2002. Information quality benchmarks: Product and service performance. *Commun. ACM 45*, 4, 184–192.

Katz-Haas, R. and Lee, Y. W. 2002. Understanding hidden interdependencies between information and organizational processes in practice. In *Proceedings of the International Conference on Information Quality*. 18–30.

Keenan, S. L., and Simmons, T. 2005. CSDQ: A user-centered approach to improving the quality of customer support data. In *Proceedings of the International Conference on Information Quality*.

Kim, Y. J., Kishore, R., and Sanders, G. L. 2005. From DQ to EQ: Understanding data quality in the context of e-business systems. *Commun. ACM 48*, 10, 75–81.

Klesse, M., Herrmann, C., Maier, D., Mügeli, T., and Brändli, P. 2004. Customer investigation process at Credit Suisse: Meeting the rising demand of regulators. In *Proceedings of the International Conference on Information Quality*. 418–432.

Knight, S.-A. 2011. The combined conceptual life-cycle model of information quality: Part 1, an investigative framework. *Int. J. Inform. Qual. 2*, 3, 205–230.

Ko, R. K., Lee, S. S., and Lee, E. W. 2009. Business process management (BPM) standards: A survey. *Bus. Process Manag. J. 15*, 5, 744–791.

Koschmider, A., Hornung, T., and Oberweis, A. 2011. Recommendation-based editor for business process modeling. *Data Knowl. Eng. 70*, 6, 483–503.

Kovac, R., Lee, Y. W., and Pipino, L. 1997. Total data quality management: The case of IRI. In *Proceedings of the Conference on Information Quality.* 63–79.

Kovac, R. and Weickert, C. 2002. Starting with quality: Using TDQM in a start-up organization. In *Proceedings of the International Conference on Information Quality.* 69–78.

Laumann, M. and Rosenkranz, C. 2008. Analysing information flows for controlling activities within supply chains: An Arvato (Bertelsmann) business case. In *Proceedings of the European Conference on Information Systems*.

Lee, G., Eastman, C. M., and Sacks, R. 2007a. Eliciting information for product modeling using process modeling. *Data Knowl. Eng. 62*, 2, 292–307.

Lee, Y., Chase, S., Fisher, J., Leinung, A., McDowell, D., Paradiso, M., Simons, J., and Yarsawich, C. 2007b. CEIP Maps: Context-embedded information product maps. In *Proceeding of Americas Conference on Information Systems*.

Lee, Y. W. 2003. Crafting rules: Context-reflective data quality problem solving. *J. Manag. Inf. Syst. 20*, 3, 93–119.

Lee, Y. W. 2006. *Journey to Data Quality*. MIT Press, Cambridge, MA.

Lee, Y. W., Allen, T., and Wang, R. Y. 2001. Information products for remanufacturing: Tracing the repair of an aircrfat fuel-pump. In *Proceedings of the Conference on Information Quality*. 77–82.

Lee, Y. W., Pipino, L., Strong, D. M., and Wang, R. Y. 2004. Process-embedded data integrity. *J. Datab. Manag. 15*, 1, 87–103.

Lee, Y. W. and Strong, D. M. 2003. Knowing-why about data processes and data quality. *J. Manag. Inf. Syst. 20*, 3 (Apr.), 13–39.

Lee, Y. W., Strong, D. M., Kahn, B. K., and Wang, R. Y. 2002. AIMQ: A methodology for information quality assessment. *Inform. Manag. 40*, 2, 133–146.

Lindland, O. I., Sindre, G., and Solvberg, A. 1994. Understanding quality in conceptual modeling. *IEEE Softw. 11*, 2, 42–49.

Lin, S., Gao, J., Koronios, A., and Chanana, V. 2007. Developing a data quality framework for asset management in engineering organisations. *Int. J. Inform. Qual. 1*, 1, 100–126.

Liu, L. and Chi, L. 2002. Evolutional data quality: A theory-specific view. In *Proceedings of the International Conference on Information Quality*. 292–304.

Loshin, D. 2001. *Enterprise Knowledge Management: The Data Quality Approach*. Morgan Kaufmann, San Diego.

Madnick, S. E., Wang, R. Y., Lee, Y. W., and Zhu, H. 2009. Overview and framework for data and information quality research. *ACM J. Data Inform. Qual. 1*, 1, 1–22.

Meda, H. S., Sen, A. K., and Bagchi, A. 2010. On detecting data flow errors in workflows. *ACM J. Data Inform. Qual.* 2, 1, 1–31.

Mendling, J., Reijers, H. A., and van der Aalst, W. M. P. 2010. Seven process modeling guidelines (7PMG). *Inform. and Softw. Technol. 52*, 2, 127–136.

Meyer, M. H. and Zack, M. H. 1996. The design and development of information products. *Sloan Manag. Rev. 37*, 3, 43–59.

Mielke, M. 2005. IQ principles in software development. In *Proceedings of the International Conference on Information Quality*.

Millard, F. H., and Lavoie, M. 2000. Developing data product maps for total data quality management: The case of Georgia Vital Records. In *Proceedings of the Conference on Information Quality*. 17–28.

Moody, D. 2009. The "Physics" of notations: Toward a scientific basis for constructing visual notations in software engineering. *IEEE Trans. Softw. Eng. 35*, 6, 756–779.

Nadkarni, P. 2006. Delivering data on time: The assurant health case. In *Proceedings of the International Conference on Information Quality*. 341–355.

Ofner, M. H., Otto, B., and Österle, H. 2012. Integrating a data quality perspective into business process management. *Bus. Process Manag. J. 18*, 6, 1036–1067.

Otto, B. 2011. Data governance. *Bus. Inf. Syst. Eng. 3*, 4, 241–244.

Otto, B., Hüner, K. M., and Österle, H. 2009. Identification of business oriented data quality metrics. In *Proceedings of the International Conference on Information Quality*. 122–134.

Pitt, L. F., Watson, R. T., and Kavan, C. B. 1995. Service quality: A measure of information systems effectiveness. *MIS Quart. 19*, 2, 173–187.

Recker, J. 2010a. Continued use of process modeling grammars: The impact of individual difference factors. *Euro. J. Inf. Syst. 19*, 1, 76–92.

Recker, J. 2010b. Explaining usage of process modeling grammars: Comparing three theoretical models in the study of two grammars. *Inform. Manag. 47*, 5-6, 316–324.

Recker, J. C., Rosemann, M., Indulska, M., and Green, P. 2009. Business process modeling - A comparative analysis. *J. Assoc. Inf. Syst. 10*, 4, 333–363.

Recker, J., Indulska, M., Rosemann, M., and Green, P. 2010. The ontological deficiencies of process modeling in practice. *Euro. J. Inf. Syst. 19*, 5, 501–525.

Recker, J. and Rosemann, M. 2010. The measurement of perceived ontological deficiencies of conceptual modeling grammars. *Data Knowl. Eng. 69*, 5, 516–532.

Redman, T. C. 1996. *Data Quality for the Information Age*. Artech House, Boston.

Redman, T. C. 1998. The impact of poor data quality on the typical enterprise. *Commun. ACM 41*, 2, 79–82.

Redman, T. C. 2004. Data: An unfolding quality disaster. *DM Rev. 14*, 8, 21–23.

Reijers, H. A. and Mendling, J. 2011. A study into the factors that influence the understandability of business process models. *IEEE Trans. Syst., Man Cybernetics, Part A 41*, 3, 449–462.

Rosemann, M. 2006. Potential pitfalls of process modeling: Part A. *Bus. Process Manag. J. 12*, 2, 249–254.

Rosemann, M., Green, P., Indulska, M., and Recker, J. C. 2009. Using ontology for the representational analysis of process modelling techniques. *Intern. J. Bus. Process Integrat. Manag. 4*, 4, 251–265.

Rosemann, M., Recker, J. C., and Flender, C. 2008. Contextualisation of business processes. *Int. J. Bus. Process Integrat. Manag. 3*, 1, 47–60.

Shankaranarayanan, G. and Cai, Y. 2006. Supporting data quality management in decision-making. *Decis. Support Syst. 42*, 1, 302–317.

Shankaranarayanan, G. and Wang, R. 2007. IPMAP research status and direction. In *Proceedings of the International Conference on Information Quality*, 500–517.

Shankaranarayanan, G., Wang, R. Y., and Ziad, M. 2000. IP-MAP: Representing the manufacture of an information product. In *Proceedings of the Conference on Information Quality*. 1–16.

Shankaranarayanan, G., Zhu, B., and Cai, U. 2008. Decision support with data quality metadata. In *Proceedings of the International Conference on Information Quality.* 281–295.

Shankaranarayanan, G., Ziad, M., and Wang, R. Y. 2003. Managing data quality in dynamic decision environments: An information product approach. *J. Database Manag. 14*, 4, 14–32.

Škerlavaj, M., Dimovski, V., and Desouza, K. C. 2010. Patterns and structures of intra-organizational learning networks within a knowledge-intensive organization. *J. Inform. Technol. 25*, 2, 189–204.

Spiegler, I. 2003. Technology and knowledge: Bridging a “generating” gap. *Inform. Manag. 40*, 6, 533–539.

Sulong, M. S., Koronios, A., Gao, J., and Abdul-Aziz, A. 2012. Service-oriented architecture intiatives with information quality: Exploring the issues. In *Proceedings of the European Conference on Information Systems*.

Sunyaev, A. and Chornyi, D. 2012. Supporting chronic disease care quality: Design and implementation of a health service and its integration with electronic health records. *ACM J. Data Inform. Qual.* 3, 2, 1–21.

Tallon, P. P. 2010. Understanding the dynamics of information management costs. *Commun. ACM 53*, 5, 121–125.

Tee, S. W., Bowen, P. L., Doyle, P., and Rohde, F. H. 2007. Data quality initiatives: Striving for continuous improvements. *Int. J. Inform. Qual. 1*, 4, 347–367.

Thi, T. T. P. and Helfert, M. 2007. Modelling information manufacturing systems. *Int. J. Inform. Qual. 1*, 1, 5–21.

Wamba, S. F. 2012. Achieving supply chain integration using RFID technology: The case of emerging intelligent B-to-B e-commerce processes in a living laboratory. *Bus. Process Manag. J. 18*, 1, 58–81.

Wand, Y., and Weber, R. 1993. On the ontological expressiveness of information systems analysis and design grammars. *Inform. Syst. J. 3*, 4, 217–237.

Wand, Y. and Weber, R. 1995. On the deep structure of information systems. *Inform. Syst. J. 5*, 3, 203–223.

Wang, R. W., and Strong, D. M. 1996. Beyond accuracy: What data quality means to data consumers. *J. Manag. Inf. Syst. 12*, 4, 5–33.

Wang, R. Y. 1998. A product perspective on total data quality management. *Commun. ACM 41*, 2, 58–65.

Wang, R. Y., Allen, T. J., Harris, W., and Madnick, S. 2002. *An Information Product Approach for Total Information Awareness*. MIT Sloan Working Paper No. 4407-02; CISL No. 2002-15.
http://papers.ssrn.com/sol3/papers.cfm?abstract_id=376820. Accessed February 2014.

Weber, K., Otto, B., and Österle, H. 2009. One size does not fit all - A contingency approach to data governance. *ACM J. Data Inform. Qual. 1*, 1, 4:1–4:27.

Weber, R. 1997. *Ontological Foundations of Information Systems*. Coopers & Lybrand and the Accounting Association of Australia and New Zealand.

Webster, J. and Watson, R. T. 2002. Analyzing the past to prepare for the future: Writing a literature review. *MIS Quart. 26*, 2, xiii–xxiii.

Wijnhoven, F., Boelens, R., Middel, R., and Louissen, K. 2007. Total data quality management: A study of bridging rigor and relevance. In *Proceedings of the European Conference on Information Systems*. 925–937.

Xie, S. and Helfert, M. 2010. Assessing information quality deficiencies in emergency medical service performance. In *Proceedings of the International Conference on Information Quality*.

Zack, M. H. 1996. Electronic publishing: A product architecture perspective. *Inform. Manag. 31*, 2, 75–86.

Zack, M. H. 1999. Managing codified knowledge. *Sloan Manag. Rev. 40*, 4, 45–58.

Zhao, X., Liu, C., and Lin, T. 2012. Incorporating business logics into RFID-enabled applications. *Inform. Process. Manag. 48*, 1, 47–62.

Received July 2013; revised February 2014; accepted May 2014