Process-driven data and information quality management in the financial service sector

Paul Glowalla
Department of Information Systems, University of Cologne, Cologne, Germany., glowalla@wiso.uni-koeln.de

Ali Sunyaev
Department of Information Systems, University of Cologne, Cologne, Germany., sunyaev@wiso.uni-koeln.de

Follow this and additional works at: http://aisel.aisnet.org/amcis2012

Recommended Citation
http://aisel.aisnet.org/amcis2012/proceedings/DataInfoQuality/2

This material is brought to you by the Americas Conference on Information Systems (AMCIS) at AIS Electronic Library (AISeL). It has been accepted for inclusion in AMCIS 2012 Proceedings by an authorized administrator of AIS Electronic Library (AISeL). For more information, please contact elibrary@aisnet.org.
Process-driven data and information quality management in the financial service sector

Paul Glowalla
University of Cologne
glowalla@wiso.uni-koeln.de

Ali Sunyaev
University of Cologne
sunyaev@wiso.uni-koeln.de

ABSTRACT

Highly regulated sectors face challenges on data and information quality management (DIQM) to conform to increasing regulations. With the financial service sector, as the most highly regulated industry, we are interested in current and future DIQM challenges. For a sustaining improvement, data quality should be managed process-driven. Process-driven data quality management (PDDQM) provides continuous improvement of data quality by redesigning processes that create or modify data. Therefore, business process management (BPM) is a basis for PDDQM. In an information systems’ context, enterprise resource planning (ERP) systems offer a platform for integrating processes and data. We examine market developments and IT trends by conducting semi-structured expert interviews with participants in IT-strategic decision making. We present current trends in the insurance sector and identify three main DIQM challenges: The IT-independent management of data, an increasing need to engage in PDDQM, and guiding existing and future measures by a data governance framework.

Keywords

Data quality management, information quality management, business process management, ERP, IT trends, financial services, insurance

INTRODUCTION

Importance of data quality is well acknowledged and data quality is critical to organizational success as vast quantities of data need to be processed by organizations (Madnick, Wang, Lee and Zhu, 2009; Shankaranarayanan and Wang, 2007). Poor data quality leads to high cost; an additional problem of lacking data quality management is that unknown data quality issues may exist, negatively affecting decision making and strategic planning (Redman, 2004, Redman, 1998). Data and information quality management (DIQM) research includes improving data and processes in order to improve data quality. Process-driven data quality management (PDDQM) provides continuous improvement of data quality by redesigning processes that create or modify data (Batini, Cappiello, Francalanci and Maurino, 2009). PDDQM outperforms data-driven data quality management – the direct modification of data values – in the long-term perspective as it seeks to identify and eliminate root causes of errors and sustain the improvements (English, 1999). Highly regulated sectors face the challenge to provide transparency through reporting (Khatri and Brown, 2010; Otto, 2011), that in turn posing challenges on DIQM. The financial service sector, as the most highly regulated industry (Syed Abdullah, Sadiq and Indulska, 2010), has to conform to several directives. Due to the continuing trend towards increased regulations (Becker, Bergener, Delfmann and Weiβ, 2011), ongoing DIQM is necessary. Therefore, we are interested in the sector’s future DIQM challenges.

Research question (RQ): What are current and future DIQM challenges in the financial service sector?

In conjunction with PDDQM, business process management (BPM) provides a basis for process improvement. Researchers’ and practitioners’ interest in BPM is increasing for decades, resulting in several standards (Ko, Lee and Lee, 2009) and process management maturity models (Röglinger, Pöppelbuß and Becker, forthcoming). From an information systems (IS) perspective, enterprise resource planning (ERP) systems provide a platform for integrating processes and data (Davenport, 1998). Most organizations, including the service sector, apply ERP systems (Botta-Genoulaz and Millet, 2006). However, an issue of ERP system implementation is the choice to adapt to the system’s generic processes or to customize the system (Hoermann, Kienegger, Langermeier, Mayer and Krcmar, 2010). The sole integration of data in an ERP system may not guarantee the quality of data and information as data and information should fit the requirements of the stakeholders (cf. Madnick et al. 2009). Hence, DIQM is necessary to manage data, focusing on stakeholders’ needs.

We conducted semi-structured expert interviews (Flick, von Kardorff and Steinke, 2009) with participants in IT-strategic decision making. The main contribution of our study is twofold. First, as a basis for answering our RQ, we present the current
application of ERP systems and the maturity of BPM and DIQM in the examined organizations. Second, to answer our RQ, we provide current strategic developments and their impact on DIQM, identifying current and future management issues.

Section 2 provides the applied definitions of ERP systems, BPM, DIQM and presents the characteristics of the financial services sector. In section 3, we present our research method and the study’s context. Section 4 provides the organizations’ current state and interdependencies regarding ERP systems, BPM, and DIQM. In section 5, we provide our findings, the main strategic development trends in the examined organizations and the resulting challenges for DIQM. We provide a summary and outlook in section 6.

THEORETICAL BACKGROUND

Table 1 forecloses the applied definitions of ERP systems, BPM, and DIQM. We applied the definitions within the study to provide a basic common understanding. However, we kept the definitions rather broad to avoid limitation on specific aspects in the interviews. We asked all participants for questions or comments for clarification regarding all definitions. Considering how BPM and DIQM are conducted within the examined organizations, we addressed the lifecycle of analyzing, modeling, implementing, and controlling processes, respectively defining, measuring, and optimizing data quality.

| BPM | Business process management is concerned with the iterative and incremental optimization of business processes. The optimization of business processes is represented in a continuous lifecycle. The lifecycle encompasses the process analysis or respectively diagnosis, process enactment, and continuous control. |
| DIQM | Data and information quality describes the information’s fitness for their intended use or task by users or user groups. Users can be internal as well as external information consumers. Data quality can be measured by several quality dimensions. The quality is measured with regard to the user requirements and the intended use. |
| ERP system | An ERP system is an integrated software for supporting main processes and important administrative functions in an organization. |

Table 1. BPM/DIQM/ERP system definitions

Derivation of applied definitions

BPM is concerned with the efficient management of business processes and their continuous improvement (van der Aalst, ter Hofstede and Weske, 2003). For managing processes, BPM provides a set of structured methods and technologies (Bandara, Chand, Chircu, Hintringer, Karagiannis, Recker, van Rensburg, Usoff and Welke, 2010). Current research provides an overview of the several BPM standards across the BPM lifecycle (Ko et al. 2009). Several BPM maturity models exist to assess organizations maturity and provide guidance for its improvement (Röglinger et al. forthcoming). Considering the definition, we focus on the continuous improvement of processes across the BPM lifecycle.

Data quality can be defined as the data’s fitness for use (Madnick et al. 2009). Information is derived from data by putting data into context; that is, giving data a meaning (English, 1999). An important milestone to data quality research was 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). There is no general agreement which set of dimensions defines data quality or the exact meaning of each dimension (Batini et al. 2009; Haug, Arlbjørn and Pedersen, 2009). Further, with new types of IS, it will be necessary to match data quality dimensions to new technological contexts (Batini et al. 2009; Madnick et al. 2009). The definition focuses on data consumers’ needs, not on specific and potentially misleading quality dimensions. As we consider data quality with respect to the intended use, we apply the terms data and information synonymously.

ERP systems are commercial software packages enabling data integration across the organization (Davenport, 1998; Haug et al.2009). Further, ERP systems are defined as “integrated standard software systems supporting core business processes across several functions” (Hoermann et al. 2010, p. 1). However, implementing a standardized ERP system imposes generic processes on the enterprise. As standard ERP systems lack specific functions for the service sector (Botta-Genoulaz and Millet, 2006), we broadened the definition to non-standardized systems.

Financial services sector

The financial services sector is the most highly regulated industry (Syed Abdullah et al. 2010) and there is a continuing trend towards increased regulations (Becker et al. 2011). Regarding insurance organizations, for instance, the current directive “Solvency II” demands increased transparency to ensure insurance organizations with regard to several economic risks (European Commission, 2011). Therefore, insurance providers face new requirements to deliver these reports.
Considering data quality, there is a gap between the ascribed importance to data quality and the already applied solution in the financial services sector. Even organizations with a high propensity for data quality investments show rather low objective investments in data quality projects (Becker, Poeppelbuss, Gloerfeld and Bruhns, 2009). Regarding BPM, different sectors are on different maturity levels (Becker, Weiß and Winkelmann, 2010). Competitive pressure forces financial services organizations to permanently improve their business processes (Heckl, Moormann and Rosemann, 2010). In this context, insufficient data quality and quantity is a main problem for process improvement.

**RESEARCH APPROACH**

We conducted 15 semi-structured expert interviews (Flick et al. 2009) between September 2011 and January 2012 to understand the IT-strategic challenges within the examined organizations. That is, besides participants directly addressing IT trends, possibly with deviating perceptions of each trend, we are interested in the underlying and currently unknown issues. Therefore, we applied semi-structured interviews for posing open questions and following up new aspects (Kvale, 2007). Due to the particularity of financial services as the most highly regulated sector (Syed Abdullah et al. 2010), we narrow our focus to the insurance sector.

We designed an interview guideline to ensure comparability between the interviews. The guideline was structured into four main sections regarding (1) the developments and IT trends in the organizations and in the service sector in general, (2) the application of ERP systems, and the maturity and application of (3) BPM and (4) DIQM within the organizations. The guideline was checked, including two test interviews, and approved by researchers and practitioners.

We chose the participants with respect to their influence on potential IT trends in the respective organizations. Whenever possible, we conducted the interview with CIOs / Heads of IT, as they are the main drivers for IT innovations (Capgemini, 2011). We conducted the 15 interviews across twelve insurance providers. One insurance provider is represented by two participants in different countries. Two participants work for other insurance-related organizations. For the analysis with regard to the organizations’ ERP systems, BPM, and DIQM, we focus on the twelve insurance providers. The insurance providers are presented in Table 2 ordered by their premium income and number of employees.

Most participants are Heads of IT departments or belong to the middle or executive management. 13 of 15 participants are involved in IT-strategy decision making as decision makers or direct advisors (Table 2). All participants are experienced, with a minimum job experience in the financial services sector of 11 years and a mean of 19.8 years. We recorded and transcribed the interviews and sent them back to the participants for communicative validation (Flick et al. 2009) resulting in minor wording adjustments. We analyzed the interviews with regard to our RQ by iterative descriptive and interpretive coding (Myers, 2011).

<table>
<thead>
<tr>
<th>Organizations’ premium income in million Euro</th>
<th>Organizations’ no. of employees</th>
<th>Participants’ position regarding IT-strategic decision making</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt; 1000</td>
<td>4</td>
<td>CIO / Head of IT (national)</td>
</tr>
<tr>
<td>&lt; 5000, &gt; 1000</td>
<td>4</td>
<td>Directly reporting to CIO / Head of IT (international)</td>
</tr>
<tr>
<td>&lt; 10000, &gt; 5000</td>
<td>2</td>
<td>Directly reporting to CIO / Head of IT (national)</td>
</tr>
<tr>
<td>&gt; 10000</td>
<td>2</td>
<td>Other executive board members</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Other participants from insurance sector</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>12</strong></td>
<td><strong>15</strong></td>
</tr>
</tbody>
</table>

Table 2. Examined organizations and participants

**CURRENT STATE OF ERP SYSTEMS, BPM, AND DIQM IN THE EXAMINED ORGANIZATIONS**

**Application of ERP systems and BPM, DIQM maturity**

In this section, we provide an overview of the application of the ERP systems within the examined organizations. Further, we assess the organizations’ DIQM and BPM maturity to build the basis for understanding our further results within the given organizational context. We consider the BPM maturity as DIQM depends on BPM to a certain degree since DIQM processes
have to be defined. Especially when managing data quality from a long-term perspective, that is, improving business processes with regard to data quality, a certain maturity of BPM is required. We chose the DIQM and BPM maturity levels with regard to their comparability. We applied the BPM Maturity Model (BPMMM) (Rosemann, de Bruin and Power, 2008) as it focuses on BPM as a holistic management practice (Röglinger et al. forthcoming). The stages are similarly applied in DIQM (Aiken, Allen, Parker and Mattia, 2007; Loshin, 2011).

**ERP system application**

Ten of the twelve insurance providers rely on standard ERP software. To achieve integrated support of general (e.g., accounting) and insurance-specific functions (e.g., product development, claims processing), the standard systems had to be integrated with other existing systems or had to be highly customized, entailing increased maintenance. The support of general functions can be achieved by standard and non-standard ERP systems. The main reasons for the implementation of standard ERP systems in the examined organizations are the standardization of systems across different sites and standardization of administrative functions and processes.

**DIQM Maturity**

As the maturity encompasses several capability areas (cf. Loshin, 2011; Rosemann et al. 2008), it is possible that organizations show different developments across capabilities. Therefore, organizations are assigned between the maturity levels if the maturity of capability areas differs clearly.

DIQM is in most cases at its very beginning (Figure 1). Organizations on the first maturity level manage data issues ad-hoc if reports deviate from expected results or data quality issues appear in data migration projects. Surprisingly, extending the approaches on DIQM in these organizations is rather not planned, although data quality itself is a main topic. This may be explained with the reliance of organizations on data checks that are built into systems. Organizations on maturity level 2 provide rudimental, but structured approaches to DIQM. Approaches to DIQM on higher maturity levels encompass automation of DIQM processes and the transfer of proven migration practices to continuous operational business.

![Figure 1. BPM and DIQM Maturity of insurance providers](image)

**BPM Maturity**

All organizations are beyond documentation of first processes. In the organization on level 1-2 (Figure 1), processes are partially documented, but process adjustments are caused by e.g., IT projects, and not conducted process-driven. Organizations on level 3, besides modeling processes, rely on process improvements based on employees’ suggestions or utilize centralized know-how for process improvement. However, the initiatives are rather sporadic and the organizations lack controlling methods to measure improvements. Organizations on higher maturity levels derive measures and control process improvements. The organization on level 5 continuously controls the process lifecycle applying process release management.
ERP systems and DIQM

The examined organizations strongly limit their DIQM to quality checks within applications. Participants emphasize data quality as an important topic and do not complain about low data quality. Therefore, standard ERP systems may provide adequate support for necessary reports. However, participants tend to adopt an IT-oriented perspective and implicitly refer to data relevant in the systems’ context. Another reason might be that data quality, although managed reactively and ad-hoc, is often improved data-driven due to mergers and acquisitions (M&A) and changes in the IT landscape, both leading to system migrations. ‘Regular migrations’ or standard ERP systems may provide adequate support to maintain a certain level of data quality within the applications. However, relying on the ERP system bears the risk to overlook data in other, not integrated, systems and individual organizational requirements.

FINDINGS

Market developments

Regulations

The already high and still increasing grade of regulation is a major concern in the insurance sector. 14 of our 15 participants take up this topic. The main concern is the current directive Solvency II. Its implementation affects the IT landscape and processes. Eventually, the implementation efforts aim at establishing the necessary data quality to fulfill the reporting specifications.

Because of Solvency II, we face increasing requirements on the core system, to assure data quality, disclose data lineage, their definition [...] I have to collect additional data to fulfill Solvency II specifications. (enterprise architect)

Due to the high priority of the specification implementation, regulations are a major driver for organizations to engage in data quality efforts (e.g., updating data models, collect additional data) and identify potential deficiencies. Because of the lack of knowledge regarding future regulations and their requirements, there is uncertainty regarding the sustainability of the existing systems, even if the current efforts aim at fulfilling the current requirements.

Regulations’ number and severity [...] are heavily increasing during the last years [...] and have gigantic impacts into our IT systems. [...] Currently there is no terminal legislative clarity. [...] That could mean, that parts of our current [IT] landscape are not adequate for conducting such reconstructing processes, [and] that we have to replace them in the future. (other executive board member)

We see the risk that organizations adopt the data quality perspective of the respective legislative body when focusing on regulation implementation. Although Solvency II is an impetus for DIQM, the data quality requirements do not match the organization-specific requirements. Therefore, data that are not considered within Solvency II are neglected.

Mergers and Acquisitions

Seven participants address M&A while five of the examined organizations currently deal with system integration due to M&A. M&A lead to fundamental changes within the current IT landscape as both data and processes need to be integrated. Standard ERP systems and their consolidating nature are applied to integrate data and processes within organization groups. As the need for M&A will persist or increase, the need for data, process, and IS integration will persist as well.

What we always see are Merger and Acquisitions. We have to adjust ourselves to buying another organization and naturally this has to be integrated. Mostly, such things have topmost priority. (CIO)

IT trends

IT trends overview

We identified five IT trends and their interdependencies within our study’s context. The IT trends are provided in Figure 2 including the number of participants that addressed the IT trends. Further IT trends such as cloud computing are not considered relevant in the context of our study. Cloud computing is addressed by five participants, but three explicitly demarcate cloud computing from relevant organizational topics. The other two participants emphasize the need for private clouds, however questioning if the addressed solutions really are cloud computing.

In this section, we focus on process automation and data analysis as these trends are closely interrelated with data quality and its management.
Figure 2: IT trends

Process automation

Participants consider the automation of processes with regard to the improvement of core business processes, that is processes that do not follow standards and therefore cannot be mapped on standard software. In this context, process automation is closely interrelated with BPM as it is necessary to (re)design processes for automation to gain the sought advantages.

[Other organizations] automate processes as they currently exist […] and that is very laborious and they only get suboptimal results. (enterprise architect)

In addition, DIQM is mutually interrelated with process automation. On the one hand, data and their quality have to be controlled for continuously as it is essential for correct process execution. On the other hand, automated processes have a positive impact on data quality.

When it [the machine] has the right information it processes them. Whether they are correct or not. (Head of IT department)

The more [manual] interventions I have […] the more problems I have with data quality in the end. (CIO)

Process automation is closely interrelated to BPM and DIQM. Regarding core business processes, a sophisticated process management approach is necessary to define, model, and map processes for automated execution. Further, the processes have to be monitored continuously. Regarding the data, their quality has to be assured as well to avoid manual intervention. Further, data quality should be improved due to reduced manual intervention. Therefore, data quality is a prerequisite for and a benefit of process automation. This mutual relationship between data quality and process automation requires an iterative process for continuous improvement.

Data analysis

Data analysis is referred to several purposes, for example, the analysis of customer data, data analysis for product development, and strategic planning. Considering strategic planning, regulations are also a driver for data analysis. The analysis of customer data and the product development are interrelated as the products should fit the customers’ requirements. Further, the effects of product changes have to be analyzed. The longevity of insurance products and the associated data implies an above-average planning horizon.

As an insurance provider has a perspective of at least ten to fifteen years, it is interesting to know, on which basis respective rate and product calculations etc. can be made. […] If we launch a new life insurance product, we would perceive the consequences in perhaps sixty, seventy years. (CIO)

Data analysis is a cross-sectional issue as it relates to all kinds of data across organizational departments, processes, and activities. Data from different sources have to be considered, for example, different organizational IS but also external sources such as statistics and social networks. Therefore, it is necessary to consider data analysis independent from existing data, data structures, and IS and to align data analysis to organizational needs. Because of the data longevity, data analysis within the insurance sector poses the need for and a special challenge on DIQM. When collecting, organizing, and
interpreting data from different sources, it is necessary to track the data’s quality to assess the quality of the analysis. The data quality information has to be transparent to data consumers, as they have to understand the data for a correct interpretation that in turn is essential for strategic planning.

*That has to be understood. That one person being used to say ‘I have fairly robust numbers. After extrapolating these over the next five years, than the result is right and the deviation is this’ now is forced to work with numbers, where another person says ‘the inaccuracy is much greater and you cannot presume, that it will be that way, but there are other scenarios that might occur.’(CIO)*

**DIQM challenges**

With respect to our RQ, we see three main DIQM challenges. The IT-independent management of data, an increasing need to engage in PDDQM, and guiding existing and future measures by a data governance framework.

Data quality in the financial services sector is seen as a priority topic by the participants and in literature (e.g., Becker et al. 2009). Despite the low DIQM maturity, participants consider their data quality high and link data quality to users’ needs. However, when referring to the data quality in the organization, participants apply an IT-oriented view. Herein lays a discrepancy between the understanding of DIQM and the perspective on the actual data quality. When considering data provided by specific IS, data quality seems to be assessed in relation to necessary or provided reports. It is a key issue to consider data independent of the applied IS. The implementation of regulations and the reliance on standard ERP systems increases this risk. Data quality has to be considered from a management perspective aligned to individual information needs (Kettinger and Marchand, 2011). These information needs arise from regulations, process automation, and data analysis. Especially data analysis imposes specific requirements on data quality across applications and even across organizational borders.

The increasing need to engage in PDDQM and managing processes with a focus on data quality is driven by process automation and the need to keep up data quality after system migrations. Process automation, especially of core business processes, is closely linked to BPM and DIQM. In this context, the automation of core business processes promises more efficient processes while maintaining organization specific processes and entailed competitive advantage. Migrations, according to the participants, lead to an identification of data quality issues and provide the possibility to improve data. However, to keep up these improvements, the processes creating and modifying data have to be adjusted as well. In this context, measures from migrations should be transferred to the operational business.

The data quality requirements might differ with the data’s purpose, but the data should reflect a single version of the truth (Khatri and Brown, 2010). A data governance framework should be established to guide existing and the proposed measures of DIQM and maximize their utilization. To avoid overhead as far as possible, the framework should provide a possibility to embed and consolidate already existing measures and identify synergies. For instance, measures that are necessary to implement regulations requirements, such as collecting new data to fulfill requirements or updating data models, should be conducted with regard to the organizations data requirements as well. In this context, DIQM is necessary to prioritize possible measures to reduce overhead. Further, the governance framework has to support communication of DIQM. To solve data quality problems, stakeholders need to understand what data quality is (Helfert and Hossain, 2010) and the data’s usage context, that is, to know ‘what’ data are collected ‘how’ and ‘why’ (Lee and Strong, 2003–4). A practitioner-oriented framework (Khatri and Brown, 2010) provides the possibility to embed our presented management issues, that is, to understand data as a corporate asset and therefore distinguish its management from IT management. Further, the data lifecycle has to be managed by organizations, to be aware of the changes that data undergoes throughout business processes. On a strategic level, a governance framework supports a continuous documentation, assessment, and improvement of DIQM, avoiding decisions on a day-to-day basis (Khatri and Brown, 2010), which is especially important in the insurance sector with its above-average planning horizon.

**SUMMARY AND OUTLOOK**

We identified market developments and IT trends that are drivers for DIQM. M&A provide the need for a holistic integration of heterogeneous landscapes. Regulation implementation is a reason for DIQM, but organizations should avoid limiting their DIQM on those requirements. DIQM will gain importance for increasing and more sophisticated process automation. In this context, with the automation of organizations’ core business processes BPM is important as well to analyze and optimize processes prior to automation to prevent automation of bad or flawed processes. DIQM is a prerequisite for data analysis, including contextual aspects such as providing the basis for a correct interpretation. Based on the developments and trend, we derived three DIQM challenges. The IT-independent management of data, an increasing need to engage in PDDQM, and guiding existing and future measures by a data governance framework.
Our results are relevant for most organizations in the insurance sector, as most organizations have ERP systems implemented. Nevertheless, our study has the following limitations. Due to the number and selection of organizations and participants, the generalization of our results is limited. The regulations of the financial services sector limit the generalizability to other sectors. Further, when aiming for a detailed maturity assessment, the criteria or capabilities may have to be adjusted for small and medium organizations, where operative requirements can be addressed on the executive board level without the necessity for a dedicated BPM team. There are several ways to arrive at a high process management maturity level (Röglinger et al. forthcoming).

Although we assumed sophisticated DIQM within the financial service sector, data quality is managed only partially, primarily driven by the necessary implementation of regulations. Further research should examine DIQM in highly regulated industries in more detail, to provide sector-specific guidelines and data governance approaches. For practice, our results provide challenges for DIQM and possible next steps to iteratively engage in a more holistic and continuous management approach.

ACKNOWLEDGMENTS

This study has been supported by ‘ConVista Consulting AG’.

REFERENCES


