The Influence of Regulation on Data Warehouse Engineering – Investigating an IT Consulting Case in the Financial Service Industry

Completed Research Paper

Jörg Becker
University of Münster - ERCIS
joerg.becker@ercis.uni-muenster.de

Mathias Eggert
University of Münster - ERCIS
mathias.eggert@ercis.uni-muenster.de

Jan Saat
zeb/information.technology
jsaat@zeb.de

Philipp Dirding
zeb/information.technology
pdirding@zeb.de

ABSTRACT

Financial service providers cope with an increasing amount of regulatory requirements causing a need to change IS systems, especially data warehouses and reporting applications. As there is no clear insight into the influence of regulation on the data warehouse (DWH) engineering process, a case study from the financial industry are conducted. Therefore, our research contributions are threefold. First, we provide an insight into regulatory-driven DWH development processes. Second, we can show that all DWH engineering phases differ if the project is triggered by regulatory requirements. Third, based on our findings, we derive a research agenda in order to address current shortcomings of research in this field.

Keywords

Regulation, Data Warehouse, Requirements Engineering, Compliance.

INTRODUCTION

Companies are exposed to a growing intensity of regulation – this is especially true for financial service providers after the peak of the financial crisis. As regulation affects all aspects of the value chain, banks are obliged to prepare their IS landscape accordingly. As core instance of financial data, data warehouses and respective reporting systems are main subjects to change in order to fulfill the requirements of the governing bodies, such as IAS39, Basel II & III or IFRS.

As DWH research has been conducted intensely over the last decades, various approaches for DWH engineering exist. Nevertheless, the influence of financial requirements on DWH projects is rarely addressed yet. In order to gain structure and exploit the potential of this matter, the goal of this paper is twofold: First, we aim to identify and clarify the effects that regulation has on the DWH engineering process. Second, based on these findings, we aim to derive research potential for an improved regulatory-driven DWH design process. In order to investigate in a real-life environment, four case studies with project leads of an IT consulting company are conducted and analyzed.

The remainder of this paper is structured as follows: related work on banking supervision and IS as well as DWH engineering is presented in the subsequent section. The research design is sketched in the third section before we present the results in section four. In the next section, we discuss the results and limitations of our analysis and conclude the paper by pointing out a research agenda to gain further insights into regulation in the context of DWH engineering in the final section.

RELATED WORK

Bank supervision and IS

The goal of bank supervision is to constrain market mechanisms by implementing certain instruments for assessing banks conditions and to conduct disciplinary actions, if required (Flannery 1998). Barth et al. (2004) provides three main reasons for bank supervision:
The monitoring of banks is costly and sophisticated, which leverages neglecting it. This in turn implies sub-optimal performance and stability. The correction of this market failure is done by official supervision.

Bank supervision prevents socially costly bank runs.

It enables the implementation of deposit insurance schemes, which motivates an restrictive risk behavior and reduces the need to monitor banks by depositors.

Institutions, such as the Federal Reserve or the European Central Bank, are responsible for bank supervision. They require the banks to submit comprehensive information regarding their activities and risk management procedures. Particularly, the reporting of adequate capital according to regulations like Basel or its transformation into European law (EU Capital Requirements Directive, CRD) (Bongaerts and Charlier 2009) becomes increasingly important (Barron and Staten 2003; Cowan and de Gregorio 2003; Tsai et al. 2011).

Craig and Diga (1996) list three types of regulations: Financial reporting-related legislation, which mainly comprises companies’ laws, security laws and tax statutes. Official directives and guidelines issued by government agencies, which comprise regulations, such as those executed by companies’ law administrators, securities market regulators and tax authorities. Rules and guidelines issued by private sector organizations keep all regulations that are posted by professional accountancy bodies and stock exchanges. In this paper, we focus on all three types of regulations, since we investigate the general relevance of regulations for the design of data warehouse and reporting systems. Further, we use the term regulation for laws, directives, acts, rules, recommendations, etc., and other principles, which are relevant for financial DWH design.

Data warehouse development process

DWHs contain huge data sets, which are used for statistical analysis, predictions and reports. They might even contain historical data, while the data sets of operational databases are smaller, serving a specific aspect of operations, and are stored as required for operations (Bischoff and Alexander 1997).

For these reasons, the traditional methods (cp. Vossen 1991) for database development cannot be applied directly to DWH engineering. End-users have different functionality and information expectations of data warehouses compared to expectations of operational databases. They want to access integrated, historical data that is relevant to their analysis and the basis for decision-making, rather than just store and retrieve operational data. Furthermore, the design aspects of operational databases and data warehouses differ.

In order to reduce data redundancies, operational (relational) databases are designed according to normalization principles, which results in complex data structures containing many relations between tables. As data warehouses are not updated by user processes, but by extraction, transformation, and load processes (ETL), which are usually performed less frequently, data redundancy is not a primary concern in data warehouses. Therefore, data warehouses are based on conceptual multidimensional models which focus on readability for users (Moody and Kortink 2000).

Figure 1: Major DWH development phases

In literature several DWH design processes have been proposed (e. g. Inmon 1996; Kimball and Ross 2002; Prat et al. 2006). Figure 1 sketches a summary of the phases of the identified development processes. The first phase is the requirements engineering, which is present in all proposals, but the one by Inmon (1996) comprises gathering of data and (non-)functional requirements. The conceptual design phase aims to develop data models (e. g. Dimensional Fact Models (DFM) or Entity Relationship Diagrams (ERD)) based on the data collected in the previous phase(s) (Golfarelli et al. 1998; Inmon 1996; Kimball and Ross 2002). The subsequent phase is the logical design. In this phase, DWH engineers decide which target logical model – being a relational or a multidimensional one – should be selected (Golfarelli and Rizzi 1998). The next phase is the implementation or physical design. An essential part of this step is the selection of a database management system (Golfarelli and Rizzi 1998) and the physical distribution of the logical structures (Luján-Mora and Trujillo 2006). The only development process that suggests testing is the one of Luján-Mora and Trujillo (2006). They explicitly call it a separate phase and define its goal as “to verify that the implementation works as desired” (Luján-Mora and Trujillo 2005, p. 20).

All identified DWH development approaches are generic and do not consider special requirements for regulatory-driven data warehouses and their development processes. In order to provide insights into DWH development projects influenced or
initiated by regulations, we interview data warehouse project managers of an IT consulting company for financial services. The following section briefly describes our research design before we present and discuss the interview results.

**RESEARCH DESIGN**

**Case study and case setting**

In order to investigate the influence of regulation on DWH development processes, we apply case study research (Yin 2009) as primary research method. Questionnaire-based expert interviews with project managers of DWH projects in the financial service sector were conducted. The questionnaire contains questions regarding activities and methods within the main DWH development phases (cf. Figure 1) focusing on the differences between regulatory-driven DWH projects. For a successful investigation of the regulations’ influence on DWH development processes, it is necessary that all respondents are experienced in both non-regulatory and regulatory-driven DWH projects. Respondents of this study are consulting and DWH experts of a medium-sized consultancy, specialized in the financial services industry. With more than 800 employees, it is one of the largest IT consulting companies for financial industries in Germany. All respondents conducted several DWH projects. In the following, we refer to the organization by using the name CONSULTANCY.

The main sources of data in this study are focused, individual, semi-structured interviews (Merton et al. 1990). This particular type of interview has an interview guide with questions that ensures that all relevant topics are covered, but leaves much freedom to the interviewees’ answers. The order in which questions are asked does not have to follow the order of the guideline. In addition, questions that are not in the guide at all may be asked to receive the desired depth of information. However, the usage of an interview guide ensures to some extent that the answers given by the interviewees have similar wordings (Bryman and Bell 2007). The interview guideline and procedure were verified in a pre-test with a consultant of CONSULTANCY. Its results are not part of the analysis as required by (Merton et al. 1990). The pretest confirmed that the structure and wording of the interview guide and its questions are adequate for this purpose.

**Data collection and analysis**

Interviews were conducted with four different DWH experts of the IT consulting unit of CONSULTANCY. In total, four interviews between 30 and 60 minutes of interview time were conducted. The selection of the interviewees was based on a respondent driven sampling (Salganik and Heckathorn 2007). CONSULTANCY provided a first contact to a partner responsible for DWH projects at CONSULTANCY. He recommended three other interview partners. This process ensures the eligibility of the candidates since the interview partner is a well connected expert in his field. Table 1 provides an overview of the interviewees respective experience level.

Each interview session was audio recorded and then transcribed and anonymized. The setting of the first interview differs to some extent from the others. The first interview was conducted in a closed office environment and without any obvious time constraints or distractions. The other three interviews took place during an internal conference at CONSULTANCY, where the interviewees had subsequent appointments. Nevertheless, the open environment did not lead to any interview interruptions.

<table>
<thead>
<tr>
<th>Interviewee no</th>
<th>Employment at CONSULTANCY [in years]</th>
<th>Data warehouse experience [in years]</th>
<th>Current job position</th>
</tr>
</thead>
<tbody>
<tr>
<td>#1</td>
<td>16</td>
<td>10</td>
<td>Partner</td>
</tr>
<tr>
<td>#2</td>
<td>10</td>
<td>8</td>
<td>Senior Manager</td>
</tr>
<tr>
<td>#3</td>
<td>9</td>
<td>9</td>
<td>Senior Manager</td>
</tr>
<tr>
<td>#4</td>
<td>9</td>
<td>9</td>
<td>Senior Manager</td>
</tr>
</tbody>
</table>

**Table 1: Interviewee Details**

During the interviews, we followed the analytical manipulation of putting information into different arrays as suggested by (Miles and Huberman 1994). The interviews were coded and categorized according to the literature-guided DWH development process. Afterwards the categorized data was interpreted regarding the influence of regulation.
RESULTS

Requirements engineering

“Regulatory requirements are at first functional requirements”, Interviewee1 said. They additionally require a set of system properties that are categorized as non-functional requirements. The functional requirements first have to be derived from the regulations. These regulations are part of “[…] German legislation and it may be difficult to understand its logic at the first glance”, Interviewee4 stated. The requirements are based on reporting obligations formulated in the legal texts and result in long lists of reporting requirements (“We have extensive lists of regulatory requirements”, Interviewee1). Using these lists, a pre-study is conducted to identify those points that are relevant in the project’s context. “… [We] do a pre-study at the bank, thus check the catalogue and look what is relevant for them”, Interviewee2 said. Most parts/ The largest part of regulatory requirements are already gathered at that point and serve as foundation for the detailed requirements definition. “This provides a good basis because you do not need to discuss these mandatory topics […]”, Interviewee2 said. Up to this point, the requirements are handled just as functional requirements stemming from non-regulatory contexts, but they have to be defined more precisely. Interviewee1 said “[…] you have to be much more accurate in the definition of the requirements”. When assessing all requirements according to critical business relevance and urgency, the regulatory requirements “[…] get always the highest priority”, Interviewee1 said. This is due to the mandatory nature of most regulations. “There is a supervisory act and this act has to be fulfilled until a certain date”, Interviewee2 said.

The non-functional parts of regulatory requirements are attributes, such as “data security, historization, […], auditability, documentation, process reliability, that processes run fully automated without manual intervention, timeliness, meaning that you have certain demands describing what should be reported […]”, Interviewee1 said. CONSULTANCY structures the requirements into the categories: stability, quality, availability, and error-proneness. For each non-functional category, the impact of regulation has to be assessed. “I have to consider every single non-functional requirement: Are there impacts from regulation that I have to consider, such as historization”, Interviewee2 said. As an example, Interviewee3 mentions that for “a normal business analysis, […] an availability of 90% during operations” is satisfying while for reporting software, “99.99% is required to be able to report within the deadline” (Interviewee3).

Overall the requirements engineering phase is perceived as easier. The previously gathered knowledge about possible regulatory requirements and the general applicability of those requirements to most of the institutions demands much less coordination between the customer and the consulting company (“So if you, like ourselves, have a relatively concrete picture of what the relevant regulatory requirements are, the requirements engineering is easier because you can say: This is the list, what else do you have?”), Interviewee1). Of course, the work of regulatory requirements analysis has to be taken into account which is presumably a lot. But the fact that it is an effort that only has to be expended once per novel regulation act reduces the marginal effort.

Conceptual and logical design

Generally, the design phase of the DWH process is not carried out differently due to regulatory influence. “Specifics in the sense of: I do another conceptual phase or another design phase, I would say rather not”, Interviewee3 said. One difference is the relevance of multidimensional organization of data for regulatory purposes. Multidimensional data structures are much less relevant compared to classical data warehousing. Interviewee4 put it this way: “[…] multidimensional reporting and classic data warehousing which, from my point of view, plays a rather inferior role, in particular for regulations, because you do not work so much with cubes. Rather it is an entity-related reporting”. Regulatory financial data is not strictly organized according to a multidimensional model. For some regulatory requirements, “[…] you simply extract data from the warehouse box, drag out to SAS or another statistical tool and then just draw reports from there. […] That is obviously highly non-multidimensional […]” (Interviewee4).

Beyond that, regulations do not require a certain physical design, hardware or software solution per se when a data warehouse already exists. But, as Interviewee1 states, “[…] through auditability and historization and so on, Excel can be ruled out for such a professional solution”. “It rather should be a market leading database management system”, he said. There is no preference for certain technology features. Therefore, functional regulatory requirements have no direct influence on the physical design. Rather, non-functional regulatory requirements have to be regarded though. The categories mentioned before, which are stability, availability, historization, auditability, etc. strongly influence the architecture of the DWH environment, e.g. the historization defining how long data has to be available and whether data needs to be archived.
Implementation (ETL)

Regulatory as well as non-regulatory data has to be extracted from the operational data sources, transformed through calculations and derivations, and loaded into the DWH. Interviewee3 said, “obviously I have to conceptualize [calculations] in order to comply with corresponding requirements […]”. The calculations that have to be computed to meet regulatory requirements naturally differ from those calculations that work on controlling or other data. However, the required ETL process is not handled differently. “The calculations are different, but they are not handled differently because it is regulation”, Interviewee1 said. The major differences between regulatory-driven and non-regulatory driven ETL processes are again the non-functional requirements. Auditable and completeness are named by Interviewee1 ([…] the great challenge is to ensure the completeness and auditability of the calculation”). The specification of these non-functional requirements “are considerably higher than for a non-regulatory scenario”, Interviewee1. A thorough documentation of the process models and the process implementation are the instruments to ensure the audibility of the ETL processes. This documentation “[…] has to be reproducible because the bank supervision wants to have a look at it”, Interviewee1 said.

As far as completeness is concerned, Interviewee1 states: “In controlling in turn it does not matter. If I have calculated 90 or 95 percent, the result does not differ that much. For a regulatory requirement, 100 percent need to be calculated and it has to be auditable in a way that at the bottom, a complete portfolio goes in and a complete portfolio comes out at the top” (Interviewee1). In scenarios other than a regulatory-driven one, a good performance might be of higher value and importance than complete calculations or complete data. While performance is valuable in a regulatory context, the completeness is a compulsory and obligatory requirement. Interviewee1 puts it in that way: “If you compare a marketing warehouse with a regulatory-driven warehouse: In marketing it is important that 80 to 90 percent of the customer data is in there and that the performance is high. Performance has to be high for regulatory requirements as well, but here you have much stronger completeness requirements”.

Furthermore, for regulatory-driven calculations, separate calculation engines are used (“There we have different calculation engines. For regulatory requirements, it is a separate one”, Interviewee1). Regulatory-driven calculation engines are standard products providing all necessary calculations and ensuring an easier approval by supervisory authorities. In addition, they are cheaper than in-house developments (“Well, with regulation I would say: classic standard software for several reasons. Quality?, approval, and it is simply cheaper”, Interviewee4; “Then you mostly choose a standard product from the market, that is specialized in programming only that calculation engine”, Interviewee2). Those standard calculation engines “[…] allow for certain configurations and global parameterization and have a defined deployment process”, Interviewee1 said.

Testing

Testing the fulfillment of regulatory requirements does not require a different methodology (“But methodologically I would not say that there is a difference”, Interviewee3). However, it is more rigorous compared to non-regulatory requirements testing (“And of course a much higher accuracy has to be fulfilled during the testing, compared to non-regulatory requirements”, Interviewee1). The tests are more sophisticated and extensive in a manner that every possible path has to be tested and not just a particular number of tests to reach a predefined level of certainty. Interviewee3 said that “more extensive audits and more extensive tests [are required]. So that [I test] every path that I can somehow go”. In addition, Interviewee1 agrees that the tests for regulatory-driven requirements are more intense. According to Interviewee4, this is because quality requirements for regulatory-driven data are well defined, in contrast to non-regulatory data (“And it is more accurate because I have clear requirements. At first, this is a raise of quality. I have a clear picture in mind for a good process. In my opinion, I do not see such clarity in other projects. […] Is the data good enough for Basel? This question I can answer”, Interviewee4). In this part of the design, the customer is closely involved in the conceptual design and test case development, especially tests that contain real data (“With this procedure, [...] the customer knows about the test case development and test case matrices in great detail, […]”, Interviewee2).

Regulatory-driven DWH development processes contain “[…] longer test phases and a better documentation, since they will most probably be inspected”, Interviewee1 said. Overall, the testing is only finished when the system runs as expected or respectively, all differences to the expectations can be explained which leads to longer testing periods (“The green light in regulation is set only if what was reported before can now be reproduced by the new data warehouse reporting software or the existing differences can be explained.”, Interviewee1).

The extensive and sophisticated testing is not a voluntary process required by the banking customers. Rather, it is “simply the requirement of the supervision that motivates a bank to do something more about it”, Interviewee2 said. Interviewee4 said; “there is at least a pressure to test” because the consequences of failed authority audits are penalties that affect the board of the bank and are embarrassing in general (“[…] the related penalties that strike the executives if they do not comply with certain things lead to doing some phases more accurately and spending more money for the whole thing”, Interviewee2; “[…]
and maybe it is even embarrassing if you do not get the clearance […]”, Interviewee4). Overall, testing is important for DWH regulatory-driven development processes. “About 40 percent of the effort for the customer is testing. It is very important”, Interviewee2 said.

**Figure 2: DWH Process Differences Due to Regulatory Influences**

Another impact of regulations can be observed in the planning and execution of the tests. The test scope can be determined in advance because the test cases are known and can be categorized by products and re-used for different customers (“[…] in regulation it pays off to use test case sets […]”, Interviewee4). This allows for a quicker definition of test cases, and the test coverage and test execution can be accelerated. The reusability aspect is limited to the test case groups because the test cases themselves need to be created individually based on the institution’s portfolio (“On the one hand you have to put the portfolio, on the other hand you have to write the cases. Then you basically pick up an account for every such thing. For a regulatory environment, it is likely that you can reuse that”, Interviewee4). The reason is that real data cannot be simulated (“But the test cases themselves are individual per institution, especially if I do real data tests”, Interviewee4). In contrast to the before mentioned statement of longer testing phases, Interviewee4 states that the described possibilities of test case groups re-use as well as predefined test cases “shorten the process”.

Figure 2 summarizes the findings and illustrates the points of regulatory influence on the DWH development process. It indicates the activities that produce additional, less or generally different effort. In the following, we briefly discuss our findings and derive an agenda for further research in the DWH development process.

**DISCUSSION**

The identified differences outlined in Figure 2 can probably be largely ascribed to the characteristics of the financial sector or to the limitation of research on this sector. The business models have very similar characteristics and regulations apply in a similar manner to most institutions, which therefore demands for similar information and calculations. In DWH projects with regulatory influences, a few additional differences can be noted. During the pre-study and requirements engineering, even more detailed lists of requirements can be used due to the experience with regulation at CONSULTANCY. However, the analysis of the legislative texts to extract the detailed lists with requirements has to be accounted for and the repeated usage of such lists only applies to the situation of CONSULTANCY as it is able to conduct multiple projects with comparable settings. As reporting requirements are quite well defined by governing bodies, the structures of the data warehouses are
similar between projects. This allows CONSULTANCY to apply patterns to data modeling activities, the ETL processes, and testing.

The case studies – and therefore the results of the paper – presume that regulatory requirements are well-defined in legal documents. Current examples of bank regulation, e.g. the execution of Basel III in Germany, draw a different picture. In this case, the requirements are specified on a very generic and high level and the banks have to suggest a solution for execution before the governing body approves the suggestion or not. In this case, the workload for requirements elicitation (here: the interpretation of the high-level regulation) remains within the bank. It can be assumed, that this phenomenon will be ruled out, as more banks successfully roll-out Basel III solutions and requirements become more explicit.

The analysis is based on a very limited sample of case studies and focus on financial services. In less regulated industries, the results may be different. In addition, the applied snowball sampling method may result in some bias due to the potential sharing of a similar viewpoint. Therefore, the results only give first indications on how regulations affect the DWH engineering process. Yet the results provide a solid foundation for the research agenda sketched in the next section.

CONCLUSION AND RESEARCH AGENDA

The paper presents a case study-based analysis of the regulatory influences on the DWH engineering process. It is shown that legal requirements account for differences – in terms of effort – within all phases of the process. In order to increase the efficiency and effectiveness of the process, the scientific community of DWH-research can contribute to numerous aspects:

1. End-to-end methods for DWH engineering: So far – as reflected in the literature review in this paper – most research focuses on singular aspects in DWH engineering. Even comprehensive works, such as the one of Jarke et al. (2003) do not provide solutions for regulatory requirements. In order to make research results applicable in real world situations, more guidance for the integration of existing method fragments in a regulatory environment is needed. Yet, there is no dedicated scientific guidance available on how to adapt existing approaches.

2. Development of modeling techniques for regulatory-driven DWH engineering. Regulatory requirements are special and not necessarily multidimensional. In order to develop conceptual models for such specific requirements, new or extended and adapted modeling techniques need to be developed.

3. Reference models: The case study results indicate that the reuse of test cases and transformed regulatory requirement lists is beneficial for consulting companies. In order to explicate such knowledge, reference models either for test cases or for regulatory requirements are necessary. While XBRL is a first step towards a harmonized data exchange (Bergeron 2004), further conceptual reference models for supporting communication processes between stakeholders in a regulatory environment are required. Existing work on reference modeling (e.g., Goeken and Knackstedt 2007) may serve as foundation for this research area.

4. DWH engineering patterns: In order to increase efficiency and effectiveness within the process, a high rate of reusability is desirable. The analysis has shown that, e.g., requirements elicitation and test case definitions can be highly standardized and reused. It is also discussed that certain standards are proposed by the governing bodies (e.g., XBRL taxonomies, IFRS). In order to create an (1) integrated and (2) customizable process, research on how to integrate existing standards and how organizational and technological interfaces must be designed needs to be conducted. Existing conceptual pattern research from software engineering offers a generic starting point for this research area.

REFERENCES