# Steering the road of Data Quality with a Dashboard


*Assessment of the Data Quality approach in Shell International Exploration & Production*

**PUBLIC VERSION**


**Author:** Mark van der Hoorn
**Faculty:** School of Management and Governance
**Master:** Industrial Engineering & Management

University of Twente
Enschede - The Netherlands

# Master Thesis

Steering the road of Data Quality with a Dashboard: *Assessment of the Data Quality approach in Shell International Exploration & Production*

October 2008 – June 2009


**Author**

Mark R. van der Hoorn
Tuin 10
1431 LW Aalsmeer, the Netherlands
Student-nr: 0073997
✉ m.r.vanderhoorn@gmail.com

**University of Twente**

Universiteit Twente
*de ondernemende universiteit*

University of Twente
Faculty of Management & Governance
Capitool 15
7521 PL Enschede, the Netherlands

Examining Committee
Dr. A.B.J.M. Wijnhoven
Capitool Building, Room C-109
P.O. Box 217
7500 AE Enschede, the Netherlands
✉ a.b.j.m.wijnhoven@utwente.nl

Dr.ir. H.G.A. Middel
Capitool Building, Room E-206
P.O. Box 217
7500 AE Enschede, the Netherlands
✉ h.g.a.middel@utwente.nl

**Royal Dutch Shell**

Shell International Exploration & Production
Kesslar Park 1
2288 GS Rijswijk, the Netherlands

Examining Committee
P.Lesslar (SIEP-EPT-IT)

R.R.F. Meiburg (SIEP-EPT-IT)

# Preface

Herewith I would like to present the public version of my master thesis. With this graduation project I am pursuing my Master of Science degree in Industrial Engineering & Management at the University of Twente. In October 2008 I have been offered the opportunity to join the internship program with Shell International Exploration and Production (SIEP).

From the start I have been assigned two projects, respectively to improve the Global Data Quality Dashboard practice (80%) and to develop a global data standard for the subsurface activity Vertical Seismic Profiling (20%). This standardization effort is necessary to be able to globally measure, calculate and analyze data quality metrics in the Global Dashboard. Since the global data standard has no scientific relevancy, my work on this has not become part of this report.

I would like to take this opportunity for expressing my gratitude to a number of people who have supported me in writing this thesis. First of all, I would like to thank my supervisor Philip Lesslar and mentor Ron Meiburg for providing me the opportunity to join the assessed internship program at the IT Subsurface & Wells department. The past six months were a tough road on which I was challenged to showcase my skills and competences. After a couple of struggles, there came light at the end of the tunnel and which resulted in an amazing experience full of learning and merriment!

Second, I highly appreciate the help and explanations I received from Jan Eikelboom, Hans Dijkerman and Hans van Smoorenburg to understand the complex subsurface activities and put together the global data standard for Vertical Seismic Profiling. I also would like to thank Gerrit Louwaars, Albert van Os and Han Tan for their willingness to help and support me during the implementation of the prototype in Spotfire.

Third, I would like to express my gratitude to my supervising professors Fons Wijnhoven and Rick Middel for their commitment and contributions to make this project a success for all parties.

Finally, I am very grateful for the mental support I have received from my family and friends during my startup struggles. Also highly appreciated are the contributions I received from my fellow Shell-interns and colleagues from the IT Subsurface & Wells department.

Mark van der Hoorn<br>*June 2009*

# Management Summary

‘REMOVED’

For proper data quality management Wang et al. argue, by means of their so-called Total Data Quality Management theory, to implement continuous improvement cycles. With the iterative process of Define, Measure, Analyze and Improve, this methodology should ensure reliable delivery of high-quality information products. This is important as high-quality data has become a strategic resource to make well-informed decisions in today’s fast changing business environments. In fact, this calls for a theoretical framework that can help companies to address and manage their data quality issues more sustainably.

Despite the availability of the Total Data Quality Management (TDQM), it does not operationalize its variables in terms of implementation and management. After all, data quality managers should be able to monitor and steer their processes with a set of instruments and controls. Therefore, this study proposes a research model that operationalizes the TDQM method and extends it with management dashboard theory. Namely, this latter enables the efficient integration of information systems and corresponding performance indicators to support managers monitoring and steering their data quality throughout the continuous improvement cycles.

For the theoretical model is a corresponding benchmark developed. It is based on statements from scientific literature and used to assess the current practice in Shell International Exploration & Production (SIEP). The outcomes of this benchmark bring a number of recommendations to improve the data quality approach in SIEP. The implement and validate these recommendations, a set of requirements is translated into a working prototype. The interactive character and strong visualization capabilities of this tool allow analyzing data quality in different formats and views.

‘REMOVED’

After implementation of the requirements in a full-functioning prototype, its effectiveness and efficiency are validated with a quasi experiment. The Technology Acceptance Model is used to validate the prototype against. This causal model is relevant as it links system features to cognitive responses and actual usage of the prototype in the end. Empirical testing in the SIEP business shows that the prototype is a significant improvement over the current situation. In other words, this research contributes to SIEP by delivering a prototype and to the academic world by delivering a validated research model that integrates the TDQM theory with a management dashboard.

# Table of Contents

# Chapter: 1 Introduction

This chapter contains the introduction of this master thesis project in Shell International Exploration & Production (SIEP). Section 1.1 starts with a brief discussion of Royal Dutch Shell and its core business processes. Section 1.2 then converges to the problem context and explains the current practice in SIEP with regard to data quality management. Based on these outcomes, Section 1.3 defines the research questions for this project. Finally, Section 1.4 concludes with the main research topics and the corresponding structure of this report.

## *1.1 The Shell Organization*

For most people Shell is a well-known organization, especially because of its service stations. Though it is mainly perceived as a petroleum company, it offers a much wider range of energy solutions and hydrocarbon chemicals. For example, investments are also made in renewables and lower-carbon energy sources. Around the globe, Shell companies work in partnership with industry, government and society to deliver what is expected from them in terms of economical, social and environmental contributions. Some short facts:

- About 102.000 people employed in over 100 countries
- Besides oil & gas interest in bio fuels, hydrogen, wind- and solar power
- Daily production is approximately 3.2 million barrels of oil equivalent
- World's largest retail network with 45.000 service stations
- Annual R&D investment of $ 1.3 billion (2008)
- One of the largest independent oil & gas enterprises in the world
- Committed to social and environmental sustainability

Figure 1 gives a visual representation of the core process in Shell:

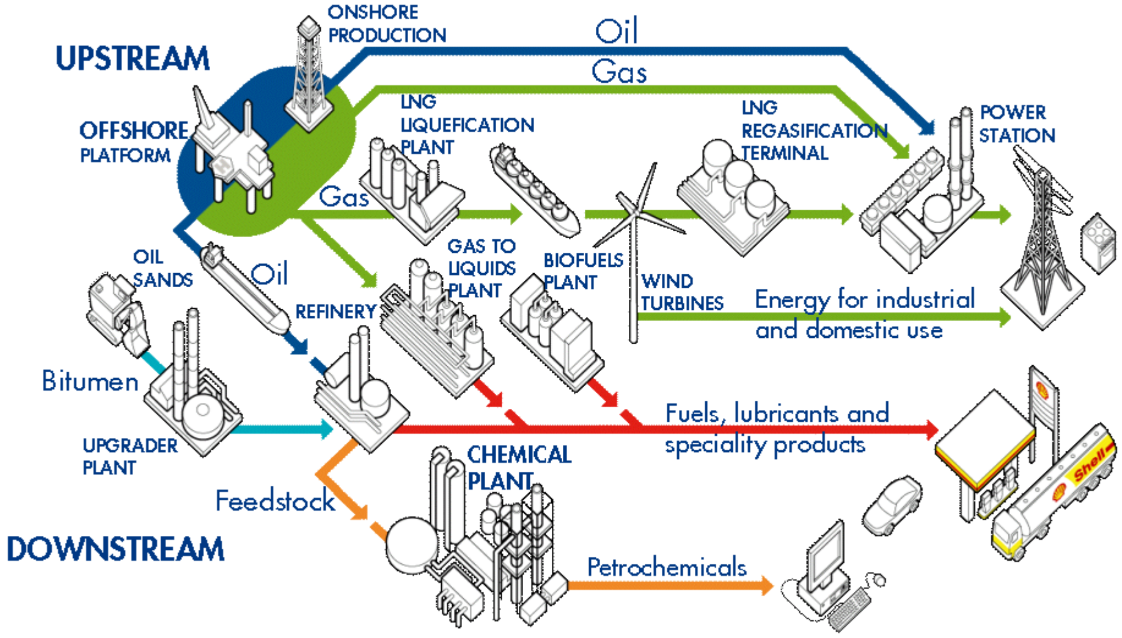



**Figure 1: The Business of Shell [113]**

## 1.2 Problem Context

‘REMOVED’

## 1.3 Research Question

‘REMOVED’

The scientific literature argues yet since 1986 to use continuous improvement cycles for quality management. Despite many of these theories are available for quality management in general (e.g. Six Sigma and TQM), there is hardly any comprehensive framework available for data quality in particular [1][5][7]. A non-pragmatic theory that can be of use is Total Data Quality Management (TDQM) by Wang et al. [2][3][5][11][19]. He argues that data quality should be managed through a continuous improvement cycle of Define, Measure, Analyze and Improve. To fit SIEP’s desire to improve on its existing methodologies used, the following research question is addressed in this paper:

**Which improvements should be made to the Global Data Quality Dashboard in order to better support the continuous improvement cycles of Total Data Quality Management?**

The following sub questions should be answered to deal with the main question:

A. Which gaps become clear from a benchmark between SIEP’s data quality approach and arguments in theory on TDQM and Management Dashboarding?
B. Which requirements and recommendations can be defined for the Dashboard?
C. What are a new dashboard design and corresponding prototype that incorporate these requirements and recommendations?
D. Which conclusions and final recommendations can be made from validating this prototype in practice?

## 1.4 Project Structure

The previous sections explained the focus of studying the Dashboard support in SIEP's data quality approach. Scientific literature is required to assess the current situation in terms of data quality management with continuous improvement cycles and management dashboarding. *Chapter 2* contains a presentation of all relevant theories currently available in literature on these topics. Then *Chapter 3* gives an explanation about the methodology used for this research. Together with SIEP's data quality approach, the theoretical foundation forms the input for a gap analysis in *Chapter 4* – thus addressing **Question A**. Next, *Chapter 5* uses the identified gaps to define new requirements and recommendations for improvement (**Question B**). The next step is to translate these into a solution artifact, which in this case consists of a prototype. The development process and final product delivery address **Question C** and are presented in *Chapter 6*. Obviously the prototype has to be tested and validated in practice, which is explained in *Chapter 7*. Also conclusions are drawn here and the final recommendations formulated to improve the Dashboard (**Question D**). Finally, *Chapter 8* contains the overall Discussion and suggestions for further work. Apart from the Introduction (*Ch.1*), Methodology (*Ch.3*) and Discussion (*Ch.8*), the project structure can be summarized as:

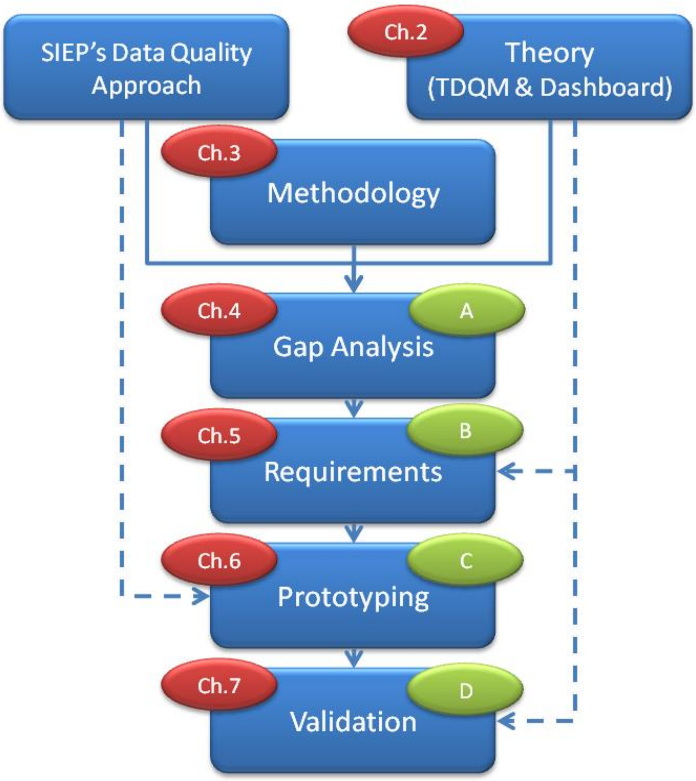



**Figure 2: Project Structure**

# Chapter: 2 Theory

This chapter concerns a theoretical discussion on the key subjects of this research, being data quality management with continuous improvement cycles in relation to management dashboarding. First of all, section 2.1 presents the findings of a structured literature review. Section 2.2 discusses data quality management in general, with subjects like the information product and continuous improvement cycles (TDQM in particular). Section 2.3 continues the discussion on dashboards, with for example possible features and relevant stakeholders. Finally, section 2.4 concludes with the development of a new research model that is used to structure the assessment in Chapter 3.

## *2.1 Literature Review*

The two following sections explain the methodology and results of a structured literature review. After all, 'relevant literature is an essential feature for any academic project' [45].

### 2.1.1 Methodology

For a structured literature review the following 4-step methodology should be applied [45]:

1) Since it is most likely that the major contributions are published in leading journals, they are started with. The search engines Scopus and Web of Science cover 92% and Inspec 88% of the Top 25 IS journals (see Appendix B) [44]. Due to the coverage and user convenience Scopus is used as search engine. Obviously all databases should come up with the same results, except for the excluded journals of course. To complete the findings from Scopus, the missing articles were manually searched in the respective journals, in this case being the 'Journal of MIS' and 'Communications of the AIS'. The following inclusion/exclusion criteria were used to select the relevant articles:
    - **Key words**: data quality, data quality management, TDQM, quality dimensions, stakeholders, management dashboard, functionality, design, implementation
    - **Management summary**: assessed on significance, novelty and generality
    - **Figures and content**: checked usefulness and reliability of the used theory

2) After this forward sampling, also backward or so-called snowball sampling is applied. This means that the reference list of earlier found articles is scanned for interesting publications. Possible articles are next searched and selected similar as in Step 1.

3) Iterations of this back and forth sampling are conducted until no new concepts are found. Using such a systematic approach should 'ensure a relatively complete census of all relevant literature' [45]. As such, a comprehensive collection of relevant articles becomes available that represent the current state of theory development.

4) A literature review should be concept-centric, especially because the author-centric method fails to synthesize the literature [45]. The articles are logically grouped, key concepts identified and compiled in a so-called concept-matrix. This table has in the rows the list of relevant articles and columns the overview of identified concepts. Every article is scored on whether it addresses a particular concept. A concept-matrix can serve as a framework for further review or structured discussion of the relevant theory. The concept-matrixes for this research can be found in Appendix C.

### 2.1.2 Search Results

This research concerns two main subjects, respectively Dashboarding and TDQM. As a result also two searches have to be done. The first search used 'data quality dashboard' and 'management dashboard' as key words to search the Top25 IS journals. Unfortunately this hardly led to any useful results, only two articles were found that were not relevant. A more extensive search covering all journals led eventually to respectively nine (for data quality dashboard) and seventeen (for management dashboard) relevant articles. Obviously some articles appeared in both searches, after selection 22 relevant and unique articles remained.

The second search used 'Total Data Quality Management' or TDQM as key words. Also here this resulted in only a few relevant articles published in the Top25 IS journals. After snowball sampling a number of other articles were added, for example of the TDQM founders Richard Y. Wang and Yang W. Lee. Although the additional articles do not mention TDQM explicitly, they contain interesting backgrounds to better understand TDQM and its context of data quality management. In order to assure the quality of these new articles, the references of the actors were checked. It appeared that some articles are published in the Top25 IS journals, respectively the Communications of the ACM, Journal of MIS and Information & Management. In addition, the number of citations gives a reliable indication for the quality as well, for some publications this exceeds a number of hundred. After selection a total of 21 relevant articles remained for this topic.

In summary, after forward and backward sampling of the main subjects, two comprehensive sets of respectively 22 and 21 useful articles became available. They give a representation of the current theoretical developments in data quality management and dashboarding. Reading and studying them more in depth led to the identification of different concepts. The aggregation of these concepts can be displayed in a matrix, the so-called concept-matrix [45]. Please refer to Appendix C for the matrixes of this research.

## *2.2 Data Quality Management*

The objective of this project is to improve the sustainability of the Dashboard by assessing it on the different continuous improvement cycles of the Total Data Quality Management theory. To get a grasp on the context of continuous improvement cycles, this section discusses the concepts as identified in the literature review – see concept-matrixes in Appendix C. Then managing information as a product and the information manufacturing system are explained. The different stakeholders are presented and the institutionalization of information quality argued. Finally, Total Data Quality Management and Aim Quality (AIMQ) as a supporting methodology are explained.

### 2.2.1 The business case

With the present trend of high customization, increased need for agility and global sustainable solutions 'high quality data has become a baseline for managing strategic corporate capacity and assets beyond operational necessity' [20]. Nowadays data is also supposed to be accessible anywhere, anytime and changing business environments require decision makers to react faster to their decision tasks [2]. This can be explained with a case study in the medic field of radiology [25][34].

Digitalized workflow models, exploding volumes and narrowing time constraints created a complex environment that made it hard for the radiologists to make well-informed and optimized workflow decisions [25]. As also in general, high quality information becomes more and more critical to every organization [1][2][15][18]. Nevertheless, in practice managers experience difficulties with effectively utilizing their information and face serious consequences as a result of low-quality information [10][13][18][19][17][38].

Taking a closer look at data quality issues in companies, the following root causes can be identified. First of all, data quality is often not part of the company's culture and it is not integrated with the daily operations [24]. As a result employees are also unaware, unwilling or unable to take care of it [27]. Second, in many companies information is dispersed across the organization. It resides in different divisions, geographical locations, data marts or paper files [28][38]. Therefore managers have to manually extract and assemble reporting information from back-end data sources [38]. Third, problems occur with information quality definition, measurement, analysis, and improvement. Without these in place it is also difficult to consistently embed quality in tools, methods, and processes [5]. Also, in many cases quality success is not defined and stakeholders have different views on victory. It can be concluded that many organizations face many problems in proper data quality management. In fact this calls for a theoretical ground that can help companies to better address data quality issues.

### 2.2.2 Definition and Views

In literature many views on data quality exist, hence no generally agreed upon definition is at hand either [4]. In many occasions data quality is defined as **fit for use** [3][5][10][17][20][21]. Although this definition might capture the essence, in practice it appears difficult to operationalize [7]. Other distinguished definitions contain [1][7][9][18]:

- Excellence
- Value
- Conformance to specifications
- Meeting consumer expectations

The **Excellence** definition is perceived as a subjective approach to assess quality, it lacks direction for improvement and possible high costs incurred are left out [7]. The **Value** definition imposes a balance between excellence and costs, hereby ignoring the importance to consumers and opting for affordable excellence [7]. The third and fourth definitions might be more practical and are therefore used for this research. Namely, **Conformance to specifications** can be defined and measured, primarily by establishing and operationalizing specifications [1][7][9]. The final view, **Meeting consumer expectations** states that information must be useful and should have an added value to its consumers [1][7][18].

Another question is the organization's perception on information management; should it be considered a delivery of service or product? [1][7] The **service** view implies an action to experience, use or consume information. For instance, the transformation of data into information can be seen as a service. This is traditionally facilitated by an IT function that provides tools and assistance to the business. **Service quality** is defined as the 'dimensions related to the service delivery process as well as addressing the intangible measures like ease of manipulation, security, and added value of the information to consumers' [7].

But this definition is a vague since it can be hardly operationalized. In my perspective service quality should be more concerned with transforming and tailoring it to the customer needs. When information is perceived as **product** mainly the production, storage and utilization is concerned [10]. In addition, adopting a product view emphasizes the delivery of valuable information to customers [11]. **Product quality** includes 'dimensions related to product features, and involves tangible measures of accuracy, completeness, and free of errors' [7].

On the road to operationalize data quality, it is often viewed as a multidimensional concept in literature [4][5][6][8]. Commonly used dimensions like accuracy, completeness and consistency were distinguished based on intuitive understanding, industrial experience and theoretical study [4]. Nevertheless 'the problem with these approaches is that they focus on the information product in terms of development characteristics instead of its use characteristics' [17]. In fact, they lack quality attributes that are important to and focus on the customer. For a data customer orientated approach 'high-quality data should be intrinsically good, contextually appropriate for the task, clearly represented and accessible' [17]. Based on empirical research, categories are established containing the data quality categories and dimensions [5][17].
They primarily seem applicable to the product view on information:

- **Intrinsic**: data should have quality in its own right.
  - accuracy, objectivity, believability & reputation
- **Contextual**: data quality must be considered within its task context.
  - value-added, relevancy, timeliness, completeness & amount of data
- **Representational**: data should be presented in an intelligible and clear way.
  - interpretability, ease of understanding, representational consistency & concise representation
- **Accessibility**: extent to which data are available and obtainable.
  - accessibility & security

Although these four categories (intrinsic, contextual, representational, and accessibility) are often used in quality discussions, these do not incorporate the information consumer needs [1]. That is why the so-called PSP/IQ model (Figure 3) has developed its dimensions from the information consumer and quality decisions point of view. With application of this PSP/IQ theory, organizations are supposed to better manage the quality of information [7]. This matrix model has in the columns the two views on data quality (conformation to specifications and meeting consumer expectations) and the rows the information product and information service quality. As a result four quadrants can be identified, respectively sound, dependable, useful and usable information [1][7].

A case study with three large healthcare organizations showed that they provide *useful* and *dependable* delivered information. Nevertheless, the *usability* and *soundness* of information scored below average. This is a common complaint, summarized by information consumers as "what we have, we use if we can. But we know it's no good" [7]. Experience proves that the soundness quadrant is the main focus of most organizations. This stems from the fact that data quality is still largely considered the responsibility from the IT function. These tend to focus on relatively quantitative and easy to measure soundness dimensions [7].

| | Conforms to specifications | Meets or exceeds consumer expectations |
|---|---|---|
| Product Quality | Sound information<br>IQ dimensions<br>Free-of-error<br>Concise representation<br>Completeness<br>Consistent representation | Useful information<br>IQ dimensions<br>Appropriate amount<br>Relevancy<br>Understandability<br>Interpretability<br>Objectivity |
| Service Quality | Dependable information<br>IQ dimensions<br>Timeliness<br>Security | Usable information<br>IQ dimensions<br>Believability<br>Accessibility<br>Ease of operation<br>Reputation |

**Figure 3: The PSP/IQ model [1]**

**Sound information**
Contains the extent to which characteristics of the supplied information meet IQ standards.
**Dependable information**
Verifies whether the process of converting data into information meets the standards.
**Useful information**
Verifies whether the supplied information meets the consumer's task needs.
**Usable information**
Contains the process of transforming data into information and the extent it fulfills consumer needs.

Be aware that a stakeholder might perceive the importance of dimensions differently than others. For example, in many occasions IT departments are very much concerned with delivering 100% accurate data, while management requires it to be 100% timely and complete [19]. Therefore communication and coordination among stakeholders is key to align expectations of delivering high quality information.

### 2.2.3 Information Product Management

Traditionally for many IT departments the 'focus is on systems and events that produce information instead of the information itself'[18]. This so-called By-Product approach controls individual components and manages the IS lifecycle. In fact, information is often perceived as a By-Product instead of the critical deliverable. Another view contains the Information Product view, which focuses on fulfilling customer needs and managing the information as a product itself [21]. Such an Information Product (IP) is defined as 'a collection of data element instances that meet the specified requirements' [12]. And a 'data element is a basic unit that has meaning in the context of the operational environment' [12]. This Product approach manages the information in an integrated way and concerns the information product lifecycle instead of the IS lifecycle. See Table 1 for further analogies.

| | Product | By-Product |
|---|---|---|
| **What is managed?** | • Information<br>• Information product life cycle | • Hardware and software<br>• Systems life cycle |
| **How is it managed?** | • Integrated, cross-functional approach<br>• Encompass information suppliers, manufacturers, and consumers | • Integrate stovepipe systems<br>• Control of individual components<br>• Cost controls |
| **Why manage it?** | • Deliver quality information products to consumers | • Implement quality hardware and software systems |
| **What is success?** | • Quality information product continuously delivered over the product life cycle<br>• No garbage-in, garbage-out (GIGO)! | • The system works<br>• No bugs<br>• Short-term perspective |

**Table 1: Comparison between product and by-product view [18]**

A great advantage of the Product view is that information quality can be sustained and safeguarded on a continuous base instead of only ad-hoc attempts (By-Product). In fact, this approach manages the information product in an integrated way and concerns the entire information product lifecycle. Which contains the 'stages through which information passes: introduction (creation), growth, maturity and decline' [18]. Research from the past decade favours that information should be managed as a Product [2][3][5][12][13][18][19]. The previous section emphasizes the importance of managing information as a Product rather than the traditional approach of managing merely hard- and software (By-Product). It is also argued that the production process of an information product should be regarded similar to manufacturing tangible products [10][11][13][17][18][21]. In fact, an analogy can be made between manufacturing a tangible product and information product (Table 2).

| | Product Manufacturing | IP Manufacturing |
|---|---|---|
| **Input** | Raw Materials | Raw Data |
| **Process** | Materials Processing | (Requirements) Processing |
| **Output** | Physical Products | Information Product |

**Table 2: Analogy between tangible and data products**

The entire set of information systems that produce information products is referred to as the Information Manufacturing System (IMS) [12]. As in physical manufacturing systems, different process steps can be recognized that should be fully understood for proper quality management. Generally speaking these steps encompass the 'collection of raw data, storage and maintenance of data, and data utilization' [21]. As in normal production management, these processes should be properly defined and controls like quality assurance, inspection and time management implemented [18]. Figure 4 shows these processes in relation to stakeholders, their responsibilities and quality problems that can occur [10].

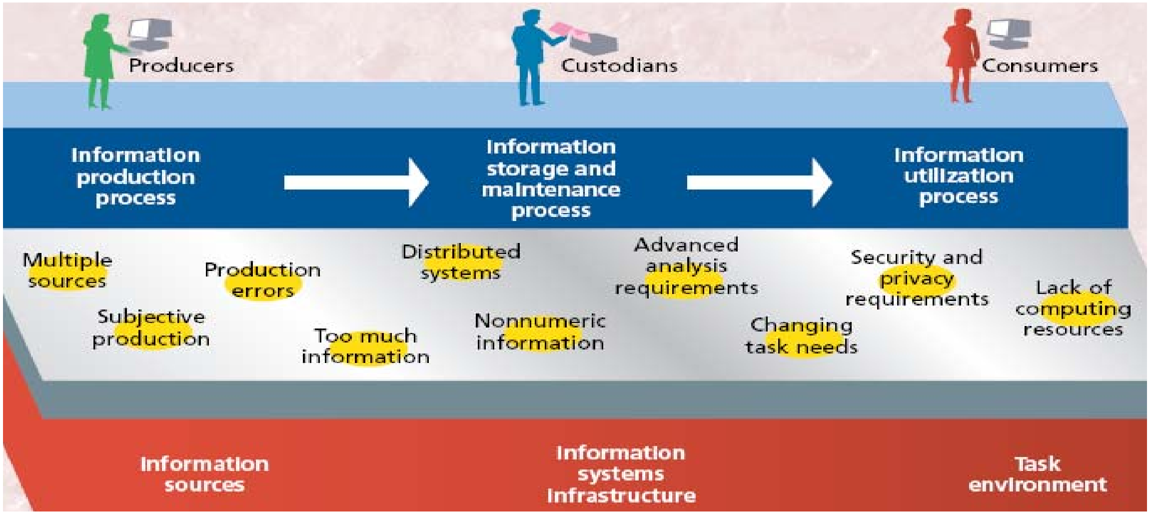



**Figure 4: Overview of manufacturing process [10]**

So quality management not only concerns correcting values, but also managing deficiencies in the production process, technical issues (storage and access) and consumer needs (utilization). Companies can better anticipate and deal with quality problems if they are able to recognize and categorize them on time.

Related to these three categories, ten causes in the information delivery process can be distinguished [10]:

- **Information production**
    - Multiple sources of the same information produce different values
    - Information is produced using subjective judgments, leading to bias
    - Systemic errors in information production lead to lost information
- **Information storage**
    - Large volumes of stored information make it difficult to access it in a reasonable time.
    - Distributed heterogeneous systems lead to inconsistent definitions, formats and values.
    - Large sets of non-numeric (qualitative) information are difficult to index.
- **Information utilization**
    - Automated content analysis across information collections is not yet available.
    - Consumers' needs for information changes and are not recognized.
    - Easy access may conflict with requirements for security, privacy and confidentiality.
    - Lack of sufficient computing resources limits access.

As in tangible production, the quality of an information product depends on the input's quality - high quality delivery therefore calls for raw data of high quality too [14]. Furthermore, the bullwhip effect explains that 'the earlier quality is considered in the production cycle, the less costly it is in the long run. Upstream defects cause downstream inspection, rework, and rejects' [9]. Managing this properly leverages product quality and intellectual capital, which 'avoids expensive mistakes, allows for faster product development, provides better customer service, creates better process management and results in more robust and reliable products' [28]. In addition, it is argued that value and quality of the information products can be enhanced by (re)designing and improving the information manufacturing system [9][11][13]. To do so it is possible to 'transfer knowledge from the field of regular production management to the field of manufacturing quality information products' [11]. This approach gains acceptance in organizations because existing experience on physical product quality (e.g. storage, processing and TQM) can be leveraged in this less developed field of data quality [2][13].

### 2.2.4 Stakeholder management

Stakeholders are an important factor in the product manufacturing process; therefore do the following discussed stakeholders correspond only with the Information Product view. The combination of data collectors, data custodians, and data consumers should monitor and evaluate the information production system [12]. In this production process 'data collectors should ask *why* people need data; data custodians ask *what* data should they be storing; and data consumers ask *how* to use these data' [21]. The theory distinguishes the following roles, corresponding tasks and relevant quality dimensions [6][7][9][10][19][20][21]:

- **Data collector/ producer –** generating and providing the information input
    - **Task**: data production process
    - **Quality dimensions**: accuracy, completeness, accessibility and relevancy
    - **Remarks**: The role of data collector is concerned key for data quality. The data collector should have a complete understanding in order to collect and select the appropriate data for production and utilization.

- **Data custodian –** providing and managing the computing resources
    - **Task**: data storage, maintenance and security processes
    - **Quality dimensions**: accuracy, completeness and timeliness
    - **Remarks**: The main concern of the custodian is to ensure that all attributes contain accurate values and processing completed on time.
- **Data consumer –** accessing and using the information for their task
    - **Task**: data utilization process (also retrieve, aggregate and integrate)
    - **Quality dimensions**: relevancy
    - **Remarks**: In fact the consumer of the data is only supposed to evaluate whether the data is relevant for his task or not.
- **Data quality administrator –** ensure that data conforms to the requirements
    - **Task**: data monitoring, controlling and reporting on quality

Though these stakeholder roles are important, they can result in different views on data quality requirements and standards [9][19][21]. Poor quality is 'commonly caused by lack of coordination and sharing of knowledge among the information consumers, producers, custodians and suppliers [18]. This can be explained with the following three premises:

- **User specificity of quality attributes**: 'Quality parameters and quality indicators may vary from one user to another. An example: for a manager the critical quality parameter for a research report may be cost, whereas for a financial trader, credibility and timeliness may be more critical' [9].
- **User different quality standards**: 'Acceptable levels of data quality may differ from one user to another. An investor loosely following a stock may consider a ten-minute delay for share price sufficiently timely, whereas a trader who needs price quotes in real time may not consider ten minutes timely enough [9]'.
- **Non-uniform quality attributes and standards**: A single user may have different quality attributes and quality standards across databases, entities, attributes, or instances. For example, a user may need higher quality information for a telephone number than the weather [9].

In addition, it can be assumed that users of a dataset know its quality. But in the information product approach, data is transferred and combined with other domains in the value chain. As a result the quality in the next phase may become unknown and different views on required quality attributes among users occur. Therefore coordination and collaboration between the different roles is required to have a mutual understanding of consumers' interests and the process of information production and maintenance [5]. In order to better facilitate coordination and deliver quality information, a new role of Information Product Manager (IPM) is suggested [7][18]. The IPM is responsible for coordination of data quality management activities among stakeholders. By application of an integrated and cross-functional approach the IPM orchestrates the fulfilling of information consumer needs. In fact the IPM is also responsible for monitoring the changing expectations and management of the continuous improvement cycle [18]. Also, the organization should be aware that not only the IT related employees are responsible for data quality, but also the functional personnel in their daily operations [19]. Finally organizations should enforce a mind shift from a strong technology focus to a wider and business supportive orientation [21].

### 2.2.5 Institutionalization

The shortcomings of the conventional data quality approach, primarily focusing on hard- and software instead of information, are yet argued. After all, a set of control technologies is used that only address data storage in terms of accuracy and integrity [6]. Many organizations have 'only piece-meal, ad hoc techniques available to measure, analyze, and improve information quality' [1][20]. In the change from by-product to product view, the focus should also shift from temporarily and ad hoc solutions to a continuous and process oriented approach [6]. To do so, solutions for data quality problems should be embedded in an overall data improvement process [1][3][8][19]. This change is also referred to as institutionalization of data quality in an organization [19]. This can be realized by embedding rules in routine work procedures, software codes, system processes and integration of the various information systems [20][36]. Nevertheless, one should be aware that assessment of data quality is a complex on-going effort that requires fundamental principles and a solid assessment structure.

Often mentioned and well-known methodologies in this field are Business Performance Management (BPM) and Business Activity Monitoring (BAM). Namely, BPM is concerned with the management, modelling and automation of business processes to increase the enterprise's agility and operational performance [39]. BAM is an enterprise solution that aggregates, analyzes and presents the performance information in a real-time manner [31]. Important to remark is that the information is reported only in business terms [31], rather than also technical aspects. Therefore recently so-called Event Processing (EP) was introduced, which integrates the technical and business layers – see Figure 5. It does so by integrating the different technologies like BPM, BAM and SOA [39]. Eventually this set of technologies allows for describing and monitoring the business performance on a continuous base.

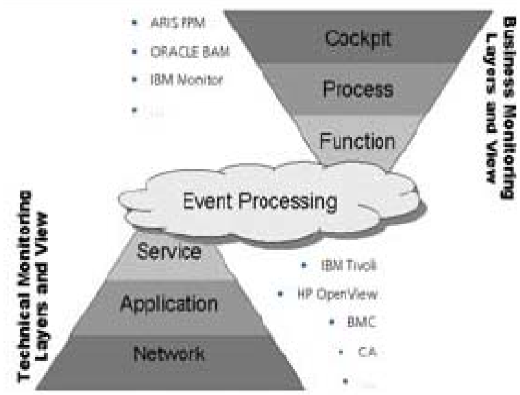



**Figure 5: Event Processing [39]**

### 2.2.6 Total Data Quality Management

The need for information product management and institutionalization of data quality has been explained in previous sections. Also the importance of shifting from ad-hoc measurement to quality management on a continuous base is stressed. In theory this is referred to as the implementation of continuous improvement cycles. Quality management through on-going cycles is already acknowledged in manufacturing theories since 1986. In that time William Deming developed Total Quality Management (TQM) as a pragmatic method for quality improvement. He believed that adopting a continuous improvement cycle of Plan, Do, Check and Act could help companies to better address and improve their production quality. Later also other methodologies like Six Sigma were based on this principle. Nevertheless, these are not specific and applicable for data quality management.

Therefore the basic cycle steps (definition, measurement, analysis and improvement) were tailored in 1992 by Wang et al. and the Total Data Quality Management (TDQM) method was born [2][3][5][11][19]. This theory aims for improvement of fundamental problems in data quality [11] in order to 'better deliver high-quality information products to information consumers' [5]. This can be realized by adopting the process of 'defining, measuring, analyzing, and improving data quality through multiple, continuous improvement cycles' [3] – see Figure 6. The success of TDQM can be explained by a case at S.C. Johnson Wax [19]. The goal of the program here was to 'deliver the right business information to the right person at the right time' [19]. Institutionalizing information quality formed the basis for continuous improvement cycles, global reporting capability and performance measurement. Johnson shifted from project-based efforts to company-wide efficient data quality management.

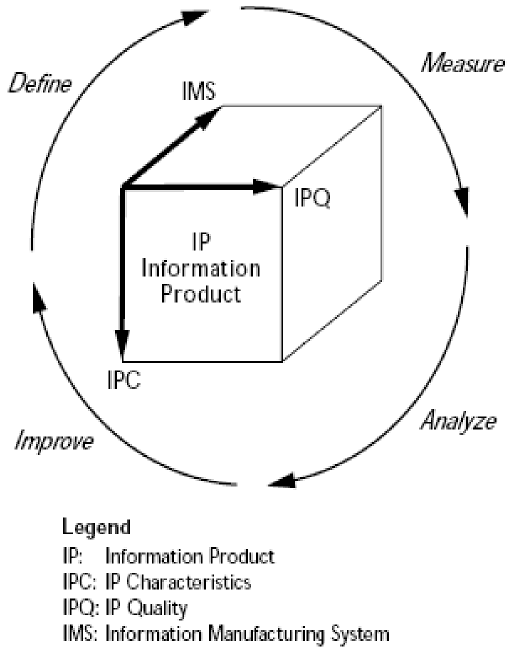



**Figure 6: The TDQM Methodology [5]**

When an organization applies the TDQM it must [5]:

(1) Clearly articulate the Information Product (IP) in business terms
  a. Define the characteristics for the IP
  b. Assess the IP's information quality requirements
  c. Identify the information manufacturing system for the IP
(2) Establish an IP team consisting of
  a. A senior executive as the TDQM champion
  b. An IP engineer who is familiar with the TDQM methodology
  c. Members who are information suppliers, manufacturers, consumers, and IP managers
(3) Teach quality assessment and management skills to all the IP constituencies
(4) Institutionalize continuous IP improvement

As a supplement to the TDQM theory, the Aim Quality (AIMQ) methodology was designed. Which provides 'a rigorous and pragmatic basis for information quality (IQ) assessment' [1]. It contains a set of pragmatic tools to identify problems, prioritize IQ improvements and monitor them over time. An important construct here is the PSP/IQ model (see Section 2.2.2). The AIMQ method has been successfully applied in various sectors, among them finance, healthcare and manufacturing [1]. The methodology is constructed of three main components that support the first three TDQM elements of Define, Measure and Analyze. The four TDQM steps combined with the assessment methodology of the AIMQ methodology results in the following descriptions:

## Step 1: Define

This first step starts with the definition of the Information Product (IP) in terms of characteristics, functionalities, components and relations [5]. Also the corresponding requirements and the Information Manufacturing System (IMS) should be determined. Finally, relevant and specific quality dimensions should be established that display information quality meanings to both information consumers and managers. These dimensions are necessary to measure and monitor the quality in later stages. The AIMQ methodology argues to use the PSP/IQ model (see section 2.2.2) as point of departure here [1]. After all, this covers important quality dimensions for information consumers.

## Step 2: Measure

In the second step quality metrics are developed to measure the current state. The set of relevant metrics should be implemented in the Information Manufacturing System as add-on routines [5]. The AIMQ method uses a questionnaire to measure the quality along the PSP/IQ dimensions [1]. In fact, this calls not only for objective assessment (like completeness and consistency), but also a subjective assessment of information quality.
This enables an organization to study also the non-quantifiable dimensions in terms of current value, their importance and (changing) consumer needs. The quality results are determined per dimension, averaged and mapped on each of the quadrants in the PSP/IQ model (section 2.3.1). The list of relevant questions is context dependent and should be developed based on the organizational characteristics [1].

## Step 3: Analyze

This third step is based on measurements and statistics to study the root causes for quality problems [5]. Also the impacts of deficiencies are calculated using a set of tools. The implications of poor-quality data can be addressed by [2]:

- Evaluating the impact of delays in one or more manufacturing stages.
- Tracing a quality problem in an IP to the manufacturing stage(s) that caused it.
- Predicting the IP impact by quality issues identified at some manufacturing steps.

The AIMQ method describes two analysis techniques for interpretation of the questionnaire and to focus on improvement efforts. The first technique 'compares the quality to a benchmark from best-practice organizations. The second technique measures the distance between the assessments of different stakeholders of an information production system' [1]. These gap analysis techniques assess the information quality in the four PSP/IQ quadrants and form the foundation for IQ improvement efforts. In case of low scores, the root causes are investigated and improvement projects started.

## Step 4: Improve

The final component argues that areas for quality improvement should be identified based on the root cause analysis from the previous step. Unfortunately the AIMQ method does not prescribe any technologies for improvement, except the fact that the quality improvements should be prioritized. The TDQM theory argues that the information product should be (re)aligned with the workflow and its characteristics redefined according to the business needs [5]. Also should the data integrity rules and standards be revised in accordance with the latest insights and developments [5].

Despite various technologies are available that address monitoring and improving operating performance, like BPM and BAM, it remains a persistent problem in organizations [31][37][39]. Often quality metrics are implemented, but it is hardly believed that these are effective and aligned with the strategic goals. In fact, the metrics often lack relevance and usefulness, leading to the following set of mistakes [37]:

- **Vanity -** Many times metrics only measure values that make performance look good. This means that for example only metrics are used that give a score above 95% or last promise date instead of customer request date in logistics.
- **Provincialism –** Refers to the fact that often metrics pertain to a specific discipline or group. As a result the processes are sub-optimized on a local scale, rather than addressing it from the organization as a whole.
- **Narcissism –** This mistake is the result of an internal view, meaning that organizations measure what is important to themselves rather than for their customers. 9 out of 10 can mean for the company 90% but for a customer 0%.
- **Laziness –** Results from jumping to conclusions and measuring what is easy to measure. Important to consider is what is really important and relevant to the customer.
- **Pettiness –** Companies tend to measure only small components rather than put it in its wider context. For example, moving production to cheap labor countries may in the end not outweigh the additional cost for logistics and lower quality.
- **Inanity –** Often metrics are implemented while management is not aware of its consequences on employees and the organization as a whole. Slightly in line with some other mistakes, metrics are implemented that are considered important for compensation reasons while the big picture is neglected.
- **Frivolity –** Where the other mistakes are sins of intellect, this worst of all and last mistake, is related to company culture. Bottom-line is that the importance of metrics and root causes are neglected, stakeholders blame each other rather than shouldering responsibilities and opinions are more convincing than objective data.

## *2.3 Management Dashboard*

Having discussed the data quality management with continuous improvement cycles, this section continues the discussion with management dashboards. As in section 2.2 also here the concepts identified in the literature review are taken as point of departure. The first section explains the business case for this tool. The second section gives a general overview of different features and requirements as discussed in literature and case studies. Section 2.3.3 gives an introduction in dashboard design and visualization considerations. Finally stakeholder management is addressed here as well.

### 2.3.1 The business case

Formerly IT staff manually collected the performance information piecewise from different data sources. Obviously this process was ‘slow, tedious and error-prone, and needed to be performed frequently for different business users’ [38]. To increase business values, IT solutions were developed that are able to monitor the business operations [36]. These efficiently and effectively integrate information systems [25] to monitor performance indicators and optimize their decision-making based on real-time information [34]. This continuous control tool makes it possible to early identify and resolve quality issues [22]. This so-called Dashboard technology continuously monitors the business performance.

Then it presents essential information to business users in a fast, on demand, accurate, easy to read and interpreting way [24][27][36][38]. A dashboard is defined as a 'concise, interactive and context-specific display of key metrics for quick evaluation of multiple subsystems' [25]. As a result, employing a Dashboard can support performance management and reduce costs of faulty decisions [26]. Namely, with personalized and timely information at hand, faster and more-well informed decisions can be made [38]. As such it 'empowers knowledge workers to make more informed decisions by convenient access to real-time summary of important system metrics' [34]. It consolidates operational data, presents business results and alerts when a metric drops below his threshold [38]. In addition, the understanding of information is improved by data clustering, knowledge codification and visualization of deficiencies and impacts [29]. These can be displayed in different formats, for example 'general-purpose lists, tables and specialized visualizations like trees, graphs, charts, or maps' [22]. Proper visual displays, like tachymeters, are critical to make decisions in a complex environment with large datasets [24].

Furthermore, a set of integrated dashboard features allows users to continuously monitor, assess and improve quality in an easy, transparent and meaningful manner [22][24]. It also enhances the visibility and communication of problems, which is considered a critical prerequisite for TDQM [3]. A dashboard serves both management by providing project control and the quality assurance staff by in-depth quality analysis [22][32]. Among other features, the dashboard can be used to interpret operational data, conduct trend analysis and tracing quality issues [24][38]. Bear in mind a dashboard should measure what is important to know for the business rather than what is easy to measure or looks fancy [22].

### 2.3.2 Requirements

The previous section shortly introduced a couple of basic characteristics for a management dashboard. This section elaborates these more in-depth and gives a brief overview of possible features derived from theory and case studies. The following key requirements form the foundation to put continuous control into practice [22]:

## Integration

- <u>Aggregation and Visualization</u>
  A dashboard is intended to present quality metrics in a comprehensive and appropriate manner, which for example can be done by categorization or aggregation of basic operators [22]. Often powerful visualization is required to display large data sets. So-called Statistical Processing Control (SPC) can be employed to visualize these and quantitatively distinguish between variations. A set of charts (Pareto, polar, bubble & spider), scatter graphs and 3D-displays can be used [24][26][33]. Another well-known metaphor is the traffic light, which is perceived a good way to summarize a system status according to 85% of the users [34].
- <u>Analysis</u>
  In many cases it is hard to recognize quality deficiencies based on a single and ad hoc base. Therefore it is required to monitor the changes over time and conduct a trend analysis. Useful to incorporate are historical information, actual values and the target measures simultaneously [23]. In some cases, also predictive reports are made available to better support decision-making [26]. Also, some metrics should be assessed with a relative scale rather than absolute measures [22].

## Functionality

- Stakeholder views
  The dashboard is used by different stakeholders that each have their own interest. As such the dashboard should be customizable to each participant's need. For instance, management requires a high-level overview in which quickly problems can be identified. In contrast, quality staff and developers demand the ability to drill-down and conduct in-depth analysis of specific quality issues [22].
- Diversity & Presentation
  Because data quality concerns many different aspects and dimensions, a control tool should be able to address many factors and artifacts as well. Not only source code must be analyzed but also measures for models, build scripts or stored information should be implemented.
  Furthermore, because the variety of stakeholder and corresponding views on quality levels, the dashboard should be able to support different analysis tools on different granularity levels [22]. Users should be able to measure specific indicators and aggregate results in groups like clients and products [33].
- Customizability
  Besides being able to address different stakeholder needs, the dashboard should also be customizable in terms of requirements, processes, analysis tools and results presentation. After all, the 'quality requirements are highly project specific due to differences in target systems, applied tools and processes, involved technologies, and project participants' [22].
- Drilling-down
  Usually analysis starts with a general overview, after which the user will zoom, filter and searches detailed information [24][38]. This process from big picture to detailed data is called drilling-down [29]. It allows users 'to drill indicators down to specific metrics or a next hierarchical level' [33]. For example, multi-layer graph-based techniques [29] or mouse-scroll-over sensitive objects can be used here [33]. With these advanced technologies a user can quickly obtain detailed information by hovering over the graph [24] or clinking on a specific bar to open a report [33].

## Management

- Workflow monitoring
  A management dashboard can play an important role in taking timely, information-driven and more confident decisions [26][28]. Monitoring the workflow and providing support in terms of consolidation, workload distribution and urgency evaluation makes a strong contribution to this [25]. As case studies showed, collected performance data and further analysis, helps to monitor the workflow more accurately and improve the process performance in the long term [26][31].
- Alerts & Warnings
  In order to trigger corrective actions, stakeholders should be alerted when a quality metric drops below his target threshold, or performance gaps and undesired trends occur [23][38]. In case of a radiology dashboard, alerts were generated with personalized, predefined and context-specific thresholds [25].

- Accountability

  In case managers have identified a bad performing quality metric, they should be able to track who is accountable and responsible for the data and solving the issue [33]. This should also help to improve the communication between the different stakeholders. After all, in case an employee faces a problem he should know who is responsible for addressing the data quality issue.

## Technology

- Quality metrics

  An important construct for a management dashboard is a set of quality metrics. These should reflect the business objectives and goals that should be monitored in the tool. A metric must be 'easy to measure, readily accessible, objective, and clearly understood' [23]. For each metric a set of queries must be developed that is linked to the data source and a proper time interval determined to review and update the metric.
- Web-based

  In order to allow for easy deployment, company-wide accessibility and maintenance it is argued to have the dashboard web-based [26][28]. In this way the information is also uniformly presented, centralized and reliable stored, more easy scalable and security more flexible [28].
- Extensibility, autonomous operation, and performance

  Since a dashboard can never provide full support initially and customer needs change over time, the tool should allow for adding new modules, for example statistical analysis or assessment modules. Secondly, the dashboard should be autonomous and not take too much effort to operate. In relation with performance, a dashboard 'must be capable of analyzing large systems within a reasonable amount of time' [22].

### 2.3.3 Dashboard Interface

The development and design of a management dashboard is extensively discussed in a couple of case studies. Important for the design is that a dashboard should be presentable and appealing to use, but especially the content is critical to the success [23]. In order to also buy-in top management, high-level information should be presented in 'a graphically rich, easy to use format and on a real-time basis' [28]. Furthermore, it should have 'a very intuitive and consistent interface for turning data into information, and information into decisions' [28].

This also relates to hierarchical presentation of quality metrics, preferably depending on the type of user. An example case presented the information corresponding to its geographical structure [38]. Obviously the highest overall overview is only accessible by a top-manager role. The next layer concerns several regions, which can be accessed by both top-management and regional managers. Then, each region has several representatives that are able to only monitor their own performance. Another case study shows three alert levels, respectively User, Division and System [34]. The first level only addresses issues pertaining to a single user, meaning deficiencies of particularly dataset under his responsibility. On the divisional level alerts are displayed that pertain to an entire group, containing issues that cannot be related to a particular individual but need group attention. The final level concerns an entire department or organization, containing for example alerts with regard to unplanned downtime or enterprise-wide issues.

In many examples the dashboard is constructed of a couple of sections/ segments, each having its own objectives and functions [29][33]. The first section usually contains at-first-sight performance and shows the overall current, historical and forecasted performance. A second part can display aggregated data clusters (e.g. per product or region) or specific indicators. A third segment is reserved for a targeted analysis, for example to study trends or conduct a statistical assessment. In some cases a data-mining engine is implemented here as well [29]. Another section can be used for expert opinions, related publications or suggestions for improvement [29]. Next, a part can be reserved for real-time profiling of assets or customer satisfaction [26]. A final section can contain a forecast engine with different statistical tools available to make quantitative future predictions [29]. Figure 7 contains an example dashboard. Features like Alerts, Tachymeters, Trends and visualization of Contract Type can be recognized in this screenshot.


**Figure 7: An example from dashboardMD [52]**

Another interesting case concerns a radiology dashboard [34]. Here an empirical study was conducted to study the impact of dashboards on radiologists' behavior. After all, it appears that in many cases employees are unaware, unwilling or unable to take care off quality issues [27]. With three different pilot dashboards, the researchers investigated a possible change in report signing behavior, a task that is often neglected but important for quality assurance. The most sophisticated dashboard, containing a direct actionable link to sign, appeared to have the biggest positive impact on the radiologists' behavior. Thus, a dashboard should be designed to support the easy execution of tasks and herewith enforce behavioral change to increase the awareness and willingness to take care of quality issues.

Finally, important to bear in mind during the design is of course the target audience. After all, an implementation can only be successful when the tool is aligned with the business, rather than being technological advanced and looking fancy [22]. A team should be build that represents different aspects of the business, including leadership, business representatives, quality practitioners and the IT department [23]. A dashboard should help the business to 'systematically identify, monitor, and address data quality problems in order to assure data is fit for use and meet its decision-making requirements' [12]. Questions should be asked like who are the users, what do they need and how can it support their daily operations [23]. Be aware of the fact that hereby employees might perceive a dashboard implementation a threat as they could feel being spied upon [24].

## *2.4 Research Model*

Section 2.2 discusses the criticality of using continuous improvement cycles (TDQM) to deliver high-quality information products. Hereby an iterative process of Define, Measure, Analyze and Improve should be followed. Though hereby all individual steps are extensively explained, the authors do not operationalize these in terms of how to implement and manage these processes with tools [50]. After all, in general project managers need a set of instruments and controls to be able to monitor and steer their processes. Therefore, the researcher suggests the application of a management dashboard in the field of data quality management. Namely, as Section 2.3 explained, a management dashboard allows managers to monitor performance indicators and optimize decision-making by efficiently and effectively integrating information systems. As such they are also referred to as Software Project Control Centres (SPCC) or Project Management Offices (PMO) [32]. In fact, a dashboard serves as a management information system (MIS) that is designed to support managers in their tasks and steer in case necessary.
In conclusion, application of these two independent theories leads to the conclusion that a management dashboard can be employed to monitor and steer data quality throughout the TDQM cycles. It is proposed that embedding the management of Define, Measure, Analyze and Improve in a dashboard enables a data manager to have real-time access and controls available to better manage the delivery of high-quality information products. This results in the development of the following improved research model (Figure 8):

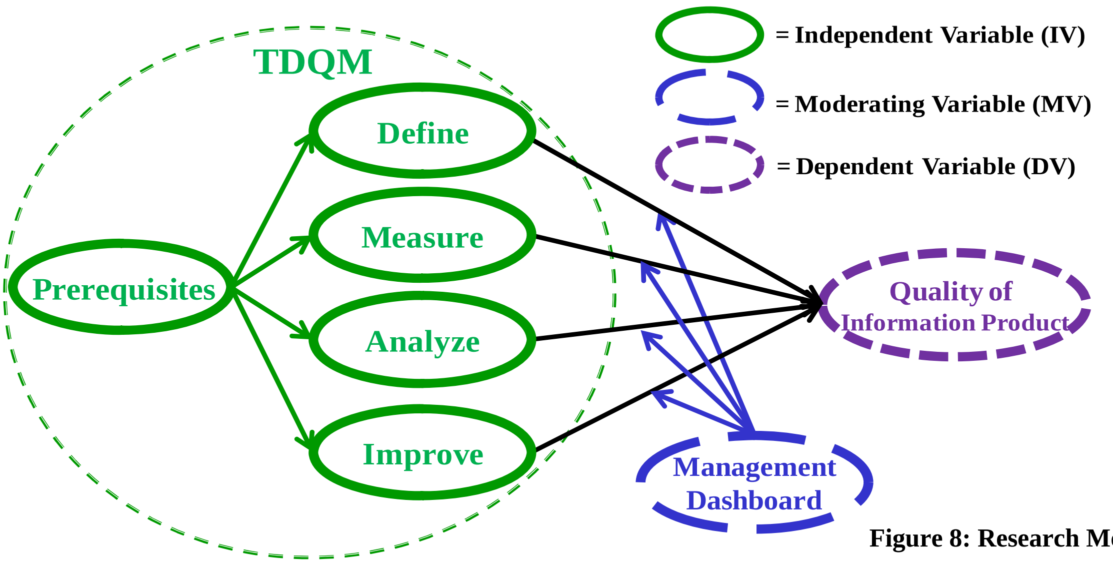



**Figure 8: Research Model**

Each of the variables and their relations are further explained here. The operationalization (IV and MV) in terms of measures is incorporated in the Gap Analysis in Section 4.2.

## Variables

- **Prerequisites (IV), Define (IV), Measure (IV), Analyze (IV) and Improve (IV)**

This set of independent variables together represents the Total Data Quality Management theory – Section 2.2. Next to the four known cycle steps (Define, Measure, Analyze and Improve), the theory also makes statements that cannot directly be related to either of these. Therefore a new independent variable is introduced, being **Prerequisites**. This variable primarily encompasses the organizational aspects that need to be in place before the other cycle steps can be implemented. Hereby one can think of institutionalization of data quality, stakeholder management and organizational commitment to data quality.

- **Management Dashboard (MV)**

This moderating variable comprehends the application of a management dashboard (Section 2.3). In general, a moderating variable means that it affects the strength of a relation between an independent and dependent variable. In this context this means, as the researcher suggests, linking a dashboard to the TDQM cycle steps (independent) and the quality of information products (DV). Namely, application of a dashboard can improve the monitoring and management of data quality throughout the cycles. Which in the end strengthens the relation between TDQM and delivery of high-quality information products.

- **Quality of Information product (DV)**

The objective of data quality management is to deliver high-quality information products. As the research model shows, this depends on the success of the four independent cycle steps from TDQM and application of a dashboard to enhance this delivery. The quality of an information product can be measured in various ways. According to theory the PSP/IQ model (Section 2.2.2) is the best to use, especially since the dimensions are developed from the information consumer and quality decisions point of view.

## Relations

- **Define (IV) – Management Dashboard (MV)**

The first step in TDQM requires the definition of data quality, the Information Product, quality dimensions and corresponding metrics. A management dashboard ensures the implementation of these and forms the communication mean of how data quality should be managed in a uniform and consistent manner.

- **Measure (IV) – Management Dashboard (MV)**

The next step in TDQM measures quality performance indicators against the defined standards. The management dashboard serves as the reporting tool in which the query results from the databases are aggregated and displayed to the management.

- **Analyze (IV) – Management Dashboard (MV)**

The Analyze step continues with visualizing the data quality indicators in various formats and calculating the business impact. The dashboard assists management with instruments and controls (e.g. trend analysis and drilling-down) to identify outliers and support in-depth analysis of data quality issues.

- **Improve (IV) – Management Dashboard (MV)**

This final step concerns the identification and initialization of data quality improvements. Hereby the dashboard serves as a control centre that allows prioritization of the identified improvements, tracking down the responsibility for an issue and monitoring the progress until the follow-up meets the target again.

# Chapter: 3 Methodology

This chapter contains an outline of the study used to develop improvements for the Global Dashboard and especially its support in TDQM's continuous improvement cycles. Generally, two types of studies can be distinguished, respectively a *Behavioural* science or a *Design* science [48]. The first paradigm concerns theories that explain and predict behaviour of humans or organizations. Hereby three additional types of study are distinguished, being explorative, explanatory and descriptive [46]. An *Explorative* study is intended to discover an ambiguous problem or develop a new subject. An *Explanatory* study has a clearly defined problem, looks for an explanation of why something has happened and addresses different dimensions of the phenomena. In a *Descriptive* study one is aware of the problem and the current state of a set of variables is formulated in an accurate way, which means without the intention for improvement.

Nevertheless, none of these seem applicable given the problem statement of this project. The objective is namely to improve the efficiency and effectiveness of data quality management in SIEP. In contrast to the *Behavioral* sciences, a *Design* science is fundamentally a problem-solving paradigm that creates new and innovative artifacts to improve current practices. Applying these two types of research results in the conclusion that a Design science is the most applicable for this case. The benefit of this study is twofold; 'namely application of academic insights should deliver concrete value for practice, and this gives feedback to the academic work' [50] in return. As such, designing a solution solves the identified problems in SIEP (practice) and implementation leads to useful experience to expand the existing scientific knowledge base (academic).

In order to conduct this problem-solving Design science, a set of seven guidelines has to be followed [48]. These guidelines are intended to assist researchers in effective and rigorous design-science research. Each of them is applied to this project in the following sections.

## 3.1 *Guideline 1:* Design as an artifact

The first requirement of design-science is the creation of a purposeful and innovative IT artifact that solves an important organizational problem. This artifact can be in different forms, for example a construct, model, method or instantiation [48].

The gaps from this benchmark serve as point of departure to define the requirements (construct/model) for the solution artifact. In addition, case studies are used (Chapter 2) to identify additional requirements and best practices from other dashboard projects. As not all can be implemented in this project, this set of requirements and recommendations needs to be prioritized based on relevancy and feasibility. This is determined in collaboration with the representatives in SIEP to ensure alignment with the business. Next, the selected requirements and recommendations are taken as starting point to develop the solution artifact. This artifact consists of developing a new dashboard design and next operationalizing it into a prototype (instantiation). It is expected that implementation of this prototype contribute to more efficient and effectively managing data quality in SIEP.

## 3.2 *Guideline 2:* Problem Relevance

The solution artifact that is designed should be relevant for the organizational problems identified. Therefore, design-science research opts to acquire knowledge and insights that support the successful development and implementation of technology-based solutions in the problematic business area. In this way the designed solutions can change and improve the occurring phenomena [48].

The necessity for proper data quality management has become clear in the problem statement. On the way to monitoring and controlling data quality, the importance of information systems can nowadays not be neglected anymore. In this reasoning, successful implementation of data quality tools has become a vital part of data quality management. Prescriptions derived from theory (Chapter 2) are used as a benchmark to assess SIEP's data quality approach…..

'REMOVED'

## 3.3 *Guideline 3:* Design Evaluation

A crucial part of a design science is evaluating and demonstrating the design via well-executed methods. This is important to check its components on issues like quality, utilization and efficiency in the business environment. Furthermore, designing is an iterative and incremental process in which feedback is essential to (re)align the solution artifact with the business requirements [48].

The design artifact is evaluated in different stages and with different stakeholders. First of all, since theory and a practice have to be developed simultaneously, the research scope and focus of the artifact are discussed with SIEP's business representatives and supervising professors. Based on the findings from the Gap Analysis, requirements are identified and translated into a prototype. The effectiveness and efficiency of this solution artifact is subsequently (internally) validated with the Technology Acceptance Model [43]. This theory relates system features to the perceived usefulness and perceived ease-of-use, which in the end contributes towards a positive attitude and eventually actual use of the new information system. This theory is developed to understand and improve the user acceptance by the design of new information systems. It is based on an experiment during implementation of an electronic mail system. This study showed a causal relationship to exist between information system features and psychological responses, like the user's attitude and behavior. As this theory is well-known and established nowadays, its key constructs can be applied here to evaluate the prototype on its improved ease-of-use and usefulness.

Next to the methodology for the design evaluation, important to discuss here as well are the new insights gained and the feedback for the proposed research model. As this study concerns a design-science, theory has been developed (academic) and applied to the SIEP business (practical). With regard to the new insights it can be concluded that management dashboards are yet introduced in the field of data quality management. Also, it appeared that SIEP's data quality approach has similar elements as argued in the Total Data Quality Management theory. In fact, this confirms for practice as well as the academic world, there is a mutual agreement on the data quality management approach.
Overall, this research helped SIEP to improve their data quality practice with recent academic insights. On the other hand, SIEP served as a suitable case to test the proposed research model, which resulted in valuable experience for the scientific knowledge base.

## 3.4 *Guideline 4:* Research Contributions

An effective design-science research should deliver clear contributions in at least one of the following three areas: novelty, generality and significance. The key question to be asked is 'what are the new and interesting contributions of this research?'[48]. An important remark is that an artifact should be implementable (evaluated by instantiation and experiment) and deliver value to the business by solving the identified problems.

Because integrated theories on data quality management with continuous cycles and application of a management dashboards is lacking, it can be concluded that this research applies to all three contribution areas:

- **Novelty –**

'REMOVED'

- **Generality –** Since no mechanism is available to validate the management of data quality with a dashboard, this research develops a model with new measures and evaluation metrics that are non-context specific. As such the outcomes can be applied both in SIEP internally and externally in other organizations with data quality practices.
- **Significance –** As explained, theory has no framework available to connect the TDQM theory to a dashboard. This design-science integrates both these topics to develop methods and instantiations that contribute to existing foundations and knowledge.

## 3.5 *Guideline 5:* Research Rigor

Rigor corresponds with how the research is conducted; meaning that for construction and evaluation of the design artifact rigorous methods should be applied [48]. As such it relates to the use of appropriate data sources (Section 3.3.1) and techniques to develop the theory and artifact. These have impact on the applicability and generalizability of the solution. Be aware that 'an overemphasis on rigor can lessen the relevance' [48]. Important to also incorporate in the rigor discussion are the reliability and three types of validity, being the construct, internal and external validity [47][51].

First of all, this SIEP project concerns application of developed theory in a real-life practical case, which makes this a **field study** rather than a laboratory setting. Second, to form a strong theoretical base, a structured literature review has been conducted according to a strict methodology (Section 2.1.1).

This ensures a comprehensive set of scientific articles that represents the current state of developed theory for both TDQM and management dashboards. The third construct for this research is the description of SIEP's data quality approach, which has been defined using multiple data sources. Mainly internal archives & documentation, colleagues and direct observation have been used to obtain a complete picture of the SIEP's data quality approach. Using a set of different data sources is important for triangulation, which requires different data sources to be used in order to check consistency between the sources (see explanation in *Guideline 6*). Using for example only direct observation might lead to a biased view of the researcher to the problem. As such, using multiple data sources ensures an accurate and valid description of the SIEP approach. Finally, the recommendations and the prototype are validated by means of a quasi experiment.

### 3.5.1 Validity

As this design-science develops a solution, an important aspect of the research rigor is *validity*. Namely, it is stated that 'a solution design is **valid** if the designed solution is expected to reduce the gap between experiences and desires that it sets out to reduce' [51]. As such, discussing the construct-, internal- and external validity and the reliability of this study helps to predict the extent to which the gap between desire and experience is expected to be bridged by the proposed solution – see Discussion in *Chapter 8*.

**Construct**

The construct validity concerns the use of a transparent methodology to operationalize and measure the research objects. This type of validity can be increased by using multiple sources of evidence and let key informants review the design and findings regularly [46].

For this research, the construct validity applies to its two constructs: the TDQM framework and SIEP's data quality approach. Namely, a theoretical framework (combining data quality management and dashboard theory) was needed to operationalize the (in)dependent variables and benchmark the current practice in SIEP. To obtain a comprehensive set of scientific articles, a formal methodology is used by means of a structured literature review [45]. For example, the search engine, key words and selection criteria are explained in Section 2.1. With this scientific basis a research model could be developed, integrating TDQM and Dashboards, to subsequently measure the variables. Next, the combination of (internal) documentation, informal discussions with business representatives and direct observation ensured a consistent and comprehensive description of SIEP's data quality approach. In this way, both the theory (framework) and practice (SIEP) constructs are operationalized in a rigorous and transparent manner.

In addition, the following conclusions can be drawn with regard to the research process. First of all, throughout the study there was continuous collaboration and alignment between both the scientific world (professors) and practice (SIEP representatives). Second, the guidelines for design-science have been effectively applied to obtain a rigorous research design [48]. Finally, multiple data sources are used to verify the consistency of the findings and strengthen the validity of the constructs used for this research.

**Internal**
The internal validity concerns whether the proposed solution solves the identified problems of the project. In fact, it reflects the level of confidence in the conclusions and the causal inferences of the academic study. Also part of the internal validity is the verification of other alternative explanations for the observed improvements [51].

The main question is whether the developed prototype (solution artifact) indeed solves the identified problems – Section 1.2. As explained for construct validity, a rigorous research design is applied to ensure the use of consistent and complete research objects. This results in identifying the root problems, understanding the gaps from the benchmark and selecting feasible recommendations in a reliable way. The next step contributing to the internal validity comprehends the operationalization into a prototype and evaluation of the recommendations by means of a quasi experiment. The Technology Acceptance Model is used to validate the prototype in the end-user community. Based on the findings from the empirical study, it can be concluded with confidence that the proposed solution solves the initially identified problems.
Nevertheless, a **quasi** experiment has been applied for validation, meaning that no control group is used, the treatment group is not randomly selected and the independent variable is not manipulated. Also given the number of respondents, it must be concluded that the degree of internal validity is average and can be improved by running a more rigorously designed experiment. In summary, it can be concluded that the research objects and their causal relations are reliable, but additional research is necessary to increase the confidence and the strengths of the correlations.

**External**
The external validity represents the extent to which the solution and findings can be applied to other domains [46]. In fact, it addresses the question whether the conclusions can be generalized. But, it is argued that attaining external validity is difficult with a single case research. After all, generalizing the findings of one situation might not result in reliable theories or experiences for all other domains [47].

Since this problem-solving design-science contains both developing and evaluating academic insights [50], the (causal) inferences in the research model are not domain specific. After all, a framework is developed that argues to employ a dashboard for data quality management. This generally applicable concept is constructed on an independent scientific knowledge base. Next the benchmark measures are applied in the SIEP domain to identify specific improvement areas. This obviously makes the outcomes of the assessment, requirements, prototype and validation domain-specific topics. Thus, to strengthen the external validity the generally applicable framework should also be used to benchmark the data quality practices in other domains and sectors.

### 3.5.2 Reliability

The reliability of a research depends on the extent that the design study is repeatable. This depends on whether the procedures are properly documented and the use of formal protocols demonstrated throughout the research phases [46]. Different methodologies are used for this research, for which the groundwork is explained in Chapter 3.

Details are discussed in the respective sections for the structured literature review (Section 2.1), assessment (Section 4.2), prototype (Section 6.3) and validation (Section 7.1). These comprehensive explanations allow for repetitiveness in terms of searching relevant theory, developing the research model and defining the benchmark. With regard to the data collection in SIEP, for confidentiality reasons the names of specific documents, employees and data repositories are not documented. Nevertheless, all these are tracked and available with SIEP's business representatives on request. With the theoretical and practical research objects available, the benchmark can be performed to identify requirements. Though no formal protocol is used to select the requirements, develop the design artifacts and operationalize these in the prototype, the procedures are transparent and clearly documented. The experiment setup in terms of the research model, questionnaire and test group makes it possible to repeat the validation of the prototype as well. It can be concluded that, based on the explained methodology and procedures, the research is repeatable and should result in similar conclusions as drawn in Chapter 8.

## *3.6 Guideline 6:* Design as a search process

Generally design sciences are inherently subjected to iterative development. It is stated that 'problem solving can be viewed as utilizing available means to reach desired ends while satisfying laws existing in the environment' [48]. These means contain the actions and available resources required to design the solution.

For this project there is not a straightforward approach of applying theory on a business case or vice versa. Therefore, a project structure is defined in Section 1.4 that phases this research project in four key steps: Gap Analysis, Requirements, Prototyping and Validation. Throughout these stages collaboration is needed between the business and the scientific field in order to iteratively determine the project scope, develop the design artifacts and implement the prototype. An important prerequisite is the availability of a theoretical framework. But as in Guideline 4, there is no research model in theory that connects TDQM to a management dashboard. Therefore, the researcher has an **active** role in developing a research model that combines TDQM with dashboarding (theory building) and changing SIEP's business (problem-solving design science).

Given this project requires both theory- and practice building, case sampling is required. This project does not concern *Multiple cases*, namely theory (TDQM and dashboarding) is applied only on the SIEP case. Since a design-science is problem-solving, it is important to validate the effectiveness and efficiency of the solution artifact. This requires the measurement to be at least two time instances (before and after). This makes **Longitudinal case** sampling a more appropriate option than a *Single case* (single moment in time).

Theory discusses different types of data sources, but not all are useful and applicable in every science study [46]. The next overview explains the relevant data sources that are used to obtain an objective and consistent picture of SIEP's data quality approach:

- **Archives and documentation**
  The main source used to find relevant information on SIEP's data quality process and the Dashboard consist of documents and presentations. SIEP has a large digital repository available in which files are stored. Searching for data quality management related files resulted in an extensive set of archives and presentation packs. The reports are mainly on the system's development process and the presentations concern primarily management information.

Next to searching the repository, colleagues showed additional papers or magazines that supplemented the researcher's findings. Eventually this resulted in an extensive collection of documents. Next these were subjected to selection on relevancy and reliability, which was based on the accuracy, publication date and author.

- **Colleagues**
  In addition to the written documentation, discussions and informal conversations were held with all the team members. Namely, the archives do not always provide sufficient information (or are unclear) and do not capture tacit knowledge. Therefore colleagues were questioned to supplement the archives with their experience, insights or opinions. This has brought up specific issues and explained why or how specific decisions were made.

'REMOVED'

- **Direct Observation**
  This data source contains an observation of the research objects by the researcher himself. For this research that means the SIEP approach is studied in terms of the different tools, data quality processes and controls. This allows gaining a firsthand impression and experience of the problems by actually using the different tools. In addition, this allowed for verification of the written documentation as well.

## *3.7 Guideline 7:* Communication of Research

This final guideline concerns effective presentation of the development and design artifact to both technology- and management-oriented audiences. This means that the communication (report/paper) should contain sufficient detail for the technical staff to construct and implement the solution in the business context. For management it is important that the presentation discusses the novelty and effectiveness of the design artifact and in which way it solves the business problem. Also the research processes should be reported clearly so the study is repeatable and allows further research by IS researchers.

The main mean of communication for this research is this master thesis report. The backgrounds of the different audiences are considered. The subjects are presented as much as possible in general business terminology and in case background information is required explanations are incorporated in the Appendixes. As such, the report is constructed in such a way that it serves both the technical staff and management in SIEP. Besides the delivery of this report, also two internal presentations have been given to communicate the main findings of the study and present the developed prototype to the key stakeholders in SIEP. For successful implementation in the longer term, it is useful to give further explanations and support by facilitating interactive (internal) workshops.
Next to the SIEP stakeholders, as important are IS researchers that can use this project as an explorative study on combining data quality management (TDQM) with management dashboards. To communicate and contribute the findings to the existing knowledge base, the key outcomes and conclusions will be summarized in a scientific article, which hopefully will be published in one of the Top 25 IS journals.

# Chapter: 4 Gap Analysis

This chapter contains the assessment of data quality approach in Shell International Exploration and Production (SIEP). As point of departure for this discussion, the theoretical framework is taken (see Section 2.4). The first section discusses the SIEP data quality approach in the light of the variables from the research model. The relevant constructs from the approach are explained per variable. Section 4.2 then uses this description to conduct a gap analysis, meaning that the SIEP approach is assessed on what is argued in theory (Chapter 2). Finally, Section 4.3 concludes with a summary of the Gap Analysis.

## 4.1 SIEP's data quality approach

This section explains the relevant constructs of SIEP's data quality approach. The structure of the discussion corresponds with the research model (Section 2.4); TDQM Prerequisites, Define, Measure, Analyze, Improve and Management Dashboard – see Figure 9. For consistency, this structure is also used in the Gap Analysis in Section 4.2.

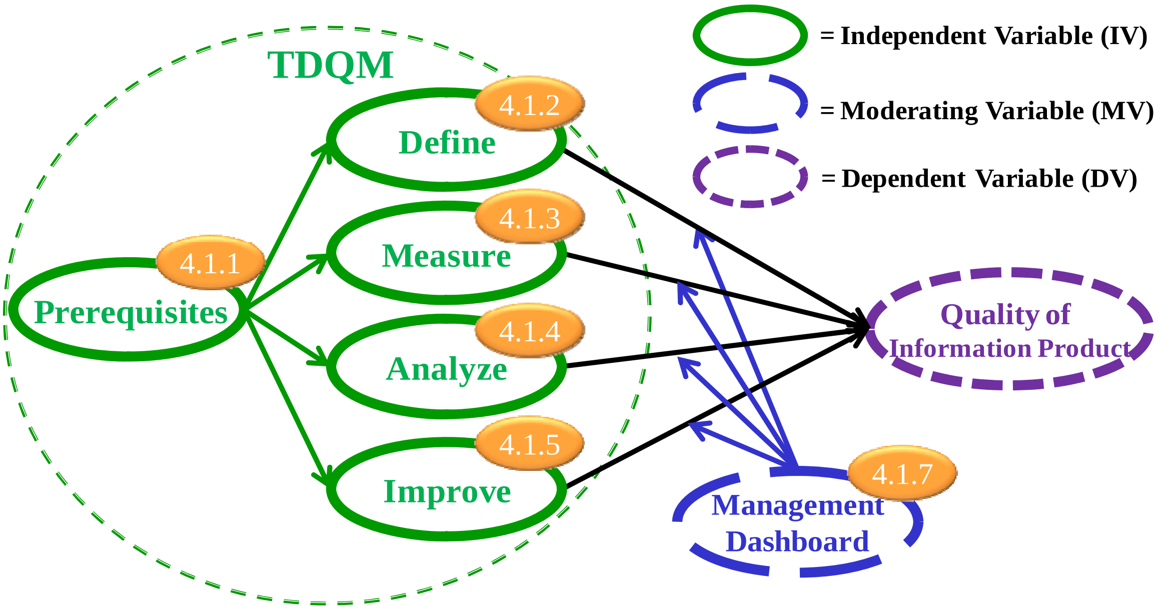



**Figure 9: Gap Analysis structure**

'REMOVED'

## 4.2 Assessment

The previous section explains the SIEP's data quality approach in terms of the variables from the research model; the Prerequisites for TDQM, Define, Measure, Analyze, Improve and Management Dashboard. This section presents the results of the benchmark, which applies statements made in Chapter 2 on SIEP's data quality approach. The same structure is used as Section 4.1 and the results of the assessment are presented in a matrix like this:

| "Variable" | Proposition | Yes | Partly | No |
|---|---|---|---|---|
| 1 | "Statement" | X | | |
| 2 | "Statement" | X | | |
| 3 | "Statement" | | X | |
| .. | .... | | | X |

The rows list the propositions made in theory corresponding to the particular variable. The columns contain the fields **Yes**, **Partly** or **No**. These are used to score SIEP's data quality approach on a statement. For example, the paragraph on the Define step states that an organization should 'define the characteristics and requirements of the information product' [3,5]. Based on the whether the practice is recognized in SIEP's data quality approach, the statement is assessed with a cross in either of the columns. In case the description in Section 4.1 seemed not sufficient to address a statement, these have been discussed with the relevant business representatives. The sub-sections give the benchmark results for each variable by the showing the table and an explanation of the individual statements. Finally, for reader convenience the propositions are showed according to the assigned value. Theory does not make any distinction between the importance or priority of the statements.

'REMOVED'

## 4.3 Summary

'REMOVED'

With this Gap Analysis available, the next step is identifying the requirements to improve the current practice in SIEP. Therefore the propositions scoring **Partly** or **No** are subjected to selection based on feasibility and business impact in Chapter 5. Also discussed are the translation and categorization of the requirements for the solution artefact. Eventually new designs are delivered that form the starting point to develop the prototype in Chapter 6.

# Chapter: 5 Requirements

This chapter identifies and selects requirements for the solution artifact. Section 5.1 starts with selecting the recommendations from the Gap Analysis (Section 4.3) based on feasibility and priority. After this selection, Section 5.2 continues with the operationalization of these requirements. Then Section 5.3 translates these requirements in three role-based dashboards, respectively for Global, Regional and Discipline managers.

## *5.1 Recommendations*

This section takes the statements on which SIEP's data quality approach scores either **Partly** or **No** in the Gap Analysis. This set of recommendations is taken as point of departure and subjected to prioritization and selection. A list of feasible requirements remains that serves as starting point for operationalization in Section 5.2.

### 5.1.1 Categorization

The first step in selecting recommendations is to categorize them. The summary of Section 4.3 is taken to derive the improvement areas. As derived from theory on Requirements Engineering, the following requirement types are used to categorize the recommendations: Business, User, Functional or System [51]. Some recommendations are related to two types of requirements, of which the second and less relevant is indicated with a **(X)**. The values are assigned based on logic reasoning and discussion with the SIEP representatives.

'REMOVED'

### 5.1.2 Selection

'REMOVED'

## *5.2 Requirements Engineering*

The previous section selects the improvement areas based on their relevance and feasibility. The causal relationships among these areas are modeled using a means-end graph. This section operationalizes the identified end-goals, which serve subsequently as input to develop the three solution artifacts in Section 5.3.

'REMOVED'.

## 5.3 Design artifacts

‘REMOVED’

## 5.4 Prototype requirements

Based on selection of the recommendations identified in the Gap Analysis (Section 4.3), a number of them have been elaborated in Section 5.2 and Section 5.3. Collating and summarizing these recommendations results in a list of requirements that can serve as basis to develop the prototype in Chapter 6. It is important to bear in mind that these requirements mainly address the issues identified in the Problem Context (Section 1.2).

‘REMOVED’

# Chapter: 6 Prototype

Chapter 5 translates the outcomes from the Gap Analysis into a set of User, Functional and System requirements. The first two have then been used to design three new role-based dashboards. The next step is to operationalize and implement these design artifacts into a functioning prototype. Therefore, Section 6.1 gives an introduction into the used application and Section 6.2 shows the instantiation by means of different screenshots.

## 6.1 Software application

'REMOVED'

## 6.2 Development

After familiarizing with the application, the next step is to operationalize the requirements as identified in Section 5.4. Namely, the requirements should be translated into prototype features. The following overview contains a trace table in which the requirements (Section 5.4) are mapped against the prototype features. Next, explanations are briefly given in order to understand the realization of the requirements. These are more extensively discussed by means of screenshots of the prototype in Section 6.3.

'REMOVED'

## 6.3 Screenshots

This section explains how the features are operationalized by means of screenshots. In this way it should become clear what the prototype looks like in practice. For the reason of confidentiality, the regions, query sets and query names are blanked. This should not make a significant difference in understanding the functional descriptions given.

'REMOVED'

# Chapter: 7 Validation

With the prototype available from Chapter 6, this chapter continues with validating its efficiency and effectiveness in the SIEP business. Section 7.1 starts with introducing the experiment in terms of the theoretical model and methodology. Section 7.2 then analyses and discusses the outcomes of the quasi experiment.

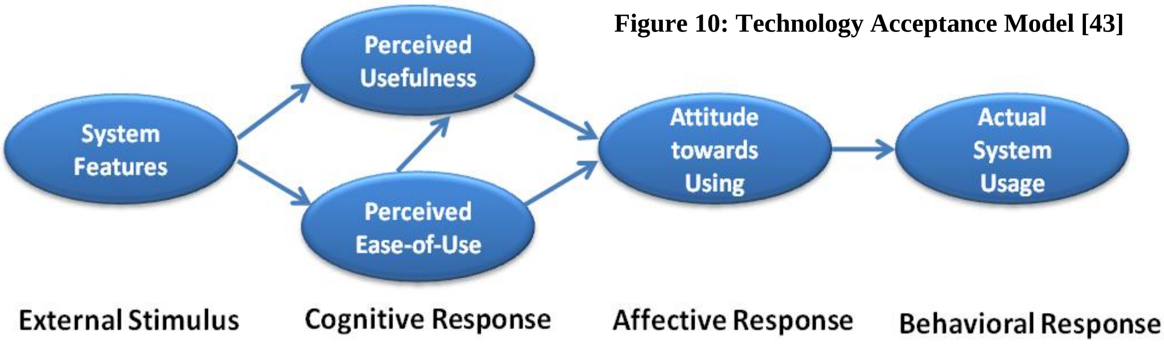
## 7.1 Experiment

To validate the prototype, empirical testing can be conducted by means of an experiment. As in Section 3.3, when the researcher has no control over all factors that can be of influence, the study is referred to as a quasi experiment. This is different from a normal experiment because there is no rigor design and control group used, the treatment group is not randomly selected and the independent variable not manipulated. To validate the prototype, a theoretical model is required that can measure the new system design. Next a corresponding questionnaire is developed and the test group determined.

### 7.1.1 Technology Acceptance Model (TAM)

To ensure successful implementation of the design artifacts it is important to validate the prototype against a research model that links system features to user acceptance ultimately. A recognized and well-established theory is the Technology Acceptance Model (TAM) [43]. This causal model is developed to understand and improve user acceptance by the design of the information system. After all, the 'lack of user acceptance has long been the impediment to the success of new information systems' and thus 'has become a pivotal factor determining the success or failure of IT projects' [43]. The author has combined information systems research with psychological responses to develop a causal relationship between system features, perceived usability, attitude and behavioral actions (Figure 25).

**Figure 10: Technology Acceptance Model [43]**




'REMOVED'

### 7.1.2 Methodology

‘REMOVED’

## *7.2 Quantitative Analysis*

This section contains the statistical analysis of the experiment. To be able to analyze the retrieved answers, the statistical program SPSS was used. Since the respondents filled out their answers in a Word-document and were sent piecewise by email to the researcher, they had to be manually populated into SPSS.

‘REMOVED’

## *7.3 Qualitative findings*

Besides the quantitative measures in Section 7.2, the respondents were also questioned for their opinions on the prototype. This section presents the findings of this qualitative analysis. It concludes with summarizing the overall findings and drawing final conclusions for this research.

‘REMOVED’

## *7.4 Conclusion*

‘REMOVED’

# Chapter: 8 Discussion & Future Research

This last chapter draws the final conclusions of this research, discusses the corresponding recommendations and finally comes up with suggestions for future research; both for the academic field and SIEP. Section 8.1 starts with answering the research question in terms of the realized improvements for the Dashboard. Also the guidelines for design-science are evaluated. Then, Section 8.2 takes the conclusions as point of departure in order to formulate the final recommendations. As such it discusses the identified problem statements (Section 1.2), the developed solutions and finally an action list to successfully implement these. Finally, Section 8.3 presents suggestions to continue this work. A distinction is made between further research for the scientific world and SIEP.

## *8.1 Conclusions*

To benchmark the data quality approach (tools, controls and processes), a framework has been developed to assess the current situation based on statements from scientific literature. As a result, feasible improvement areas have been selected and validated in order to improve the current practice. Therefore, findings and conclusions should be discussed for both the SIEP business (practice) and academic world (new theory).

### 8.1.1 SIEP business

'REMOVED'

### 8.1.2 Academic world

Next to the findings for the SIEP business, a number of important conclusions can be drawn with regard to implications for the academic world. As such, the findings with regard to TDQM, management dashboards and the proposed research model are discussed.

- **Total Data Quality Management**
  The TDQM theory by Wang et al. has played a major role in this study. After all, the use of continuous improvement cycles (TDQM) has proven to be important for the delivery of high-quality information products. An important conclusion of this research is that data quality management in theory and in practice have quite some commonalities. In fact, this confirms for both practice and the academic world, there is mutual agreement on data quality management. On the other hand, subjects like the Information Manufacturing System seem easy to realize in theory, but global standardization projects in large organizations like SIEP require a lot of effort. Theory often has a green field assumption, whereas practice has to deal with legacy first.
- **Management Dashboard**
  The advantages of a management dashboard become more and more clear these days. Namely, literature contains a lot of case studies whereby organizations monitor and steer their process with a set of instruments and controls. As most articles are about case studies, kernel theories for a management dashboard are still lacking.

As such there is no framework or manual available on what a dashboard should look like or how it should be implemented. It might be due to the fact that a dashboard is very context specific, which makes it hard to pre-define its context and design. For this study a number of theories are combined that together formed a general dashboard design (Section 5.2). Finally, as with the TDQM theory, it can be concluded that the management dashboard arguments are recognized in practice and they provided useful input to assess and improve the practice in SIEP.

- **Research Model**
  Whereas the TDQM and Dashboard theory existed for a couple of years, no framework was available integrating both. As can be derived from Section 2.4, this research proposes to connect a Management Dashboard as moderating variable to each of the continuous improvement cycles; Define, Measure, Analyze and Improve. This allows data quality managers to monitor performance indicators and optimize their decision-making by efficiently and effectively integrating information systems.
  As this research concerns a problem-solving design-science, this research model is operationalized and applied to benchmark the current practices in SIEP. As most of the improvements could be realized in the Analyze and Improve step, the developed prototype has a specific focus on supporting these. The validation (Chapter 7) showed significant improvements with regard to the usefulness, meaning the support in SIEP's data quality approach. This shows that the Dashboard has a direct relation with data quality management and can make contributions to improve the data quality practice. Although the impact of the prototype on the delivery of high-quality data is not measured in this research, the users have indicated it increases the productivity, control, quality and effectiveness of their data quality job (Section 7.2).
  Nevertheless, it also has to be concluded that a management dashboard cannot be employed to support all the tasks and activities in TDQM. Especially the role in the Define step is questionable. After all, the organization should first define vision, quality dimensions, metrics and stakeholders before a dashboard can even be implemented. Furthermore, the dashboard cannot support all the tasks for the Improve step either. For example, it can help to identify and prioritize improvements but it is up to the data quality manager to follow-up on the issues and to correct the data himself.

### 8.1.3 Design-Science

Another important aspect in the discussion of this research is the set seven guidelines to conduct a problem-solving design science (Chapter 3). As such each guideline is reflected and evaluated here separately [47]:

- **Guideline 1: Design as an artifact**
  As the prototype has significantly proven improvements over the current practice, it can be concluded that this artifact is a purposeful and innovative IT solution. Next to the development of the prototype (instantiation), this research also delivers a new theoretical model (method) to benchmark the data quality management practices in an organization. After all, it can be concluded that two important artifacts (instantiation and method) are designed and delivered by means of this research.

- **Guideline 2: Problem relevance**
  The second guideline requires the delivery of a solution that is relevant to the organizational problems. Both the quantitative and qualitative findings of the quasi experiment show that the prototype solves the identified problems.

  ‘REMOVED’

  Overall, it can be concluded that the deliverables of this research are relevant to the problem and solve the identified issues.

- **Guideline 3: Design evaluation**
  The design artifact is validated both quantitatively and qualitatively. A quasi experiment in combination with the well-established Technology Acceptance Model was used to evaluate the effectiveness of the solution artifact. Nevertheless, to obtain more reliable outcomes a more rigorous experiment should be applied and the efficiency of the prototype measured in the business environment. After all, despite improvements can be realized in the design evaluation, it can be concluded that the used methodology is sufficient to obtain reliable results.

- **Guideline 4: Research contributions**
  Logically every research should have a contribution, either in terms of its novelty, generality and significance. As Section 3.4 explained, this research basically contributes to all these three areas. *Novelty* because a newly designed prototype is delivered to the business (practical). And a new theoretical model combining TDQM with management dashboard theory is build (academic). *Generality* because the developed benchmark and dashboard design principles are applicable to any other domain as well. *Significance* because the outcomes of this research contain important findings to extend the existing knowledge bases. Overall, it can be concluded that this research has significant contributions to each of the three areas.

- **Guideline 5: Research rigor**
  The fifth guideline concerns the methodology used to conduct the research, meaning the extent to which rigorous methods are applied to construct and evaluate the solution. To assure a rigorous research design and process, a number of formal methodologies is used; like the guidelines for a structured literature review (Chapter 2), design-science (Chapter 3) and validation (Chapter 7). In addition, different data sources are important for triangulation, which requires several sources to be used in order to check consistency between the sources. The internal-, external- and construct validity and reliability play an important role in the research rigor as well. As in Section 3.5, the internal validity can be improved by running a rigorous experiment rather than a quasi experiment. Also, the external validity of the scientific conclusions can be improved by applying the research model to more cases than only this single one. Furthermore, this study is reliable as all the methods and processes are extensively documented, allowing for repetition of this research. Thus, despite a couple of improvements available, it can be concluded that rigorous methods are used to conduct this research.

- **Guideline 6: Design as a search process**
  Because design-sciences are subjected to iterative development, continuous alignment between stakeholders and revision of scope are important considerations. As explained in Section 1.4, a proper project structure is used to guide the search and development process. In addition, a variety of data sources (archives, documentation, colleagues and direct observation) are used to obtain a comprehensive and consistent picture of the SIEP practices. Since all stakeholders are closely involved and strict guidelines implemented, it can be concluded that the right actions are taken and resources utilized to design and deliver the solution artifact.

- **Guideline 7: Communication of research**
  This final guideline requires effective presentation of the research outcomes and developed solution artifacts. The most important communication means are this master thesis and the final defense at the university. Next, two presentations in SIEP were given to communicate the main findings of the benchmark and present the developed prototype to the key stakeholders. Finally, the key outcomes and conclusions will be summarized in a scientific article, which hopefully will be published in one of the Top 25 IS journals. After all, it can be concluded that the findings of this research are communicated through a variety of means and domains.

### 8.1.4 Overall

To start with the previous discussion, it can be concluded that this research fulfills the guidelines, apart from a couple exceptions, that are required for a proper design-science. As such, the findings and conclusions are considered reliable and based on scientific grounds.

‘REMOVED’

On the academic side, SIEP served as a suitable case to test the proposed research model. The validity and the importance of the theory are confirmed in this case study. Although the external validity is not very strong yet (single case), the findings contain important contributions to the existing scientific knowledge base.

## *8.2 Recommendations*

This section discusses the recommendations that are the result of this research. First of all, the initial problem statement is repeated to refresh the memory of the situation this study tends to improve (Section 1.2). Then the two solutions are presented and elaborated in more detail. This section concludes with a number of considerations for this research.

### 8.2.1 Problem statement

‘REMOVED’

### 8.2.2 Solutions

‘REMOVED’

### 8.2.3 Action list

‘REMOVED’

### 8.2.4 Considerations

Despite a rigorous methodology (Chapter 3) is applied for this research, a couple of considerations have to be discussed. First of all, the developed theoretical model is only tested with a single case study at SIEP, which implies it is difficult to attain external validity. This means that one should be careful with generalizing the findings from one situation to reliable theories or application in other domains. Furthermore, the theoretical model is not empirically tested and quantitatively validated. This means that this model needs further research before it can be reliably applied to other organizations. With regard to the validation of the prototype, a quasi experiment is used rather than the more rigorous experiment. To increase the confidence and strengths of the correlations, the following issues should be considered: use of a control group, random selection of the treatment group and manipulation of the independent variable.

In addition, an experiment is always based on opinions of users in this case; this might affect the outcomes as they might answer with a biased view rather than based on reality [51]. Not all experimental subjects are a professional in the data quality management field, which means that they might lack expertise and knowledge to give reliable answers [51]. Finally, the fact that a prototype is developed rather than a real system might affect the validity and reliability of the findings for the longer term as well [51]. Overall, it can be concluded that the outcomes of this study are valid and reliable within the SIEP domain. But additional research and empirical study is necessary to generalize the findings to reliable theories and apply them in other domains.

## *8.3 Future Research*

Based on the considerations from the previous sections, this final section discusses suggestions for future research more concretely. The sub-sections present directions to respectively improve the SIEP practice and the academic work.

### 8.3.1 SIEP business

‘REMOVED’

### 8.3.2 Academic world

This study has developed new academic insights that can be extended by conducting research in the following directions:

- **Business Case development**
  According to theory, managers often assess data quality based on their intuitiveness, familiarity and experience with their data sets [14]. Furthermore, they are unaware of their role in the value chain (e.g. bullwhip effect) and are not committed if their value delivery and return are not equivalent [49].

  ‘REMOVED’

- **Governance**
  Despite different stakeholders are discussed for both data quality management and dashboard, there is no commonly agreed governance structure for the here proposed integrated approach. A clear decision- and responsibility structure is necessary to govern the processes and systems corresponding to managing data quality with a dashboard. After all, for data quality follow-up and improvement it is important to know who the data owners, administrators and custodians are. As these can be different units or employees, there needs to be a consistent and clear governance structure.
- **Business Process Management**
  Once Total Data Quality Management is successfully implemented, the next step would be to embed data quality in other programs like Business Activity Monitoring [31] and Business Performance Management [39]. Then data quality can become an integral part of overall business performance measurement. Interesting to research would be the integration of data quality in existing programs like (Lean) Six Sigma or Operational Excellence. Whereas this research has shown the success of employing a dashboard for TDQM, it would also be interesting to study the role of a dashboard in these programs.
- **Generalization and Certification**
  As discussed in the external validity, generalizing the findings in this research requires replication and testing in other domains. For example, to be done by studying whether the conclusions apply to information management as well or develop kernel theories for data quality management [50]. The latter prescribe which choices to make given a given set of (organizational) conditions.
  As this requires multiple case studies, but also in general, supplementary research should investigate the current data quality practices in other organizations or sectors. With a benchmark being developed in this study, it should be more convenient to conduct this benchmark now. Taking this idea to a further level brings us to data quality certification. Although certificates exist for quality management topics, like the ISO 9000 series and Balridge National Quality Program, there is no well-defined certificate for data quality yet.

## Scientific References


1. Lee, Y. W., Strong, D. M., Kahn, B. K., & Wang, R. Y. (2002). AIMQ: A Methodology for Information Quality Assessment. Information & Management, 40(2), 133-146

2. G. Shankaranarayan, M. Ziad & R.Y. Wang (2003). Managing Data Quality in Dynamic Decision Environments, Journal of Database Management, 14(4), Oct-Dec 2003, 14-32

3. Y. W. Lee, L. Pipino, D.M. Strong & R.Y. Wang (2004). Process embedded data integrity. Journal of Database Management, 15(1), 87-103, Jan-Mar 2004

4. Y. Wand & R.Y. Wang (1996). Anchoring data quality dimensions in ontological foundations, Communications of the ACM 39 (11), 1996, pp. 86–95.

5. Wang, R.Y. (1998). A product perspective on total data quality management. Communications of the ACM 41, 2 (1998), 58–65.

6. D.M. Strong, Y.W. Lee & R.Y. Wang (1997). Data quality in context. Communications of the ACM 40 (5), 1997, pp.103–110.

7. B.K. Kahn, D.M. Strong & R.Y. Wang (2002). Information quality benchmarks: product and service performance, Communications of the ACM 45 (4ve), April 2002, pp. 184–192.

8. Pipino, L., Lee, Y., & Wang, R. (2002). Data Quality Assessment. Communications of ACM, (April), 211-218.

9. Wang, R. Y., Kon, H. B., & Madnick, S. E. (1993). Data Quality Requirements Analysis and Modeling. Proceedings of the 9th International Conference on Data Engineering, Vienna.

10. D.M. Strong, Y.W. Lee & R.Y. Wang (1997). “10 putholes in the road to information quality”. Computer 30 (8), pp. 38-46

11. R.Y. Wang, V.C. Storey, and C.P. Firth, “A Framework for Analysis of Data Quality Research,” IEEE Trans. Knowledge and Data Eng., Aug. 1995, pp. 623-640.

12. Davidson, B., Lee, Y.W., Wang, R.Y. (2004). “Developing data production maps: meeting patient discharge data submission requirements”. International Journal of Healthcare Technology and Management 6 (2), pp. 223-240

13. Ballou, D.P., Wang, R.Y., Pazer, H. and Tayi, G.K. (1998) 'Modeling information manufacturing systems to determine information product quality', Management Science, Vol. 44, No. 4, pp.462–484.

14. Ballou, D.P., Chengalur-Smith, I.N., Wang, R.Y. (2006). "Sample-based quality estimation of query results in relational database environments". IEEE Transactions on Knowledge and Data Engineering 18 (5), pp. 639-650

15. Ballou, D., Madnick, S., Wang, R. (2003). "Special Section: Assuring Information Quality". Journal of Management Information Systems 20 (3), pp. 9-11

16. Wang, Richard Y., Reddy, M.P., Kon, Henry B. (1995). "Toward quality data: an attribute-based approach". Decision Support Systems 13 (3-4), pp. 349-372

17. R.Y. Wang & D.M. Strong (1996). "Beyond Accuracy: What Data Quality Means to Data Consumers," J. Management Information Systems, pp. 5-34

18. Wang, R.Y., Lee, Y.L., Pipino, L. and Strong, D.M. (1998) 'Manage your information as a product', Sloan Management Review, Vol. 39, No. 4, pp.95–105

19. J. Funk, Y.W. Lee, and R.Y. Wang (1998). "Institutionalizing Information Quality Practice," Proc. 1998 Conf. Information Quality, vol. 3, pp. 1-17

20. Y.W. Lee (2003). "Crafting Rules: Context-Reflective Data Quality Problem Solving". Journal of Management Information Systems 20 (3), pp. 93-119

21. Y.W. Lee & D.M. Strong (2003). "Knowing-Why About Data Processes and Data Quality". Journal of Management Information Systems 20 (3), pp. 13-39

22. Deissenboeck, F., Juergens, E., Hummel, B., Wagner, S., Parareda, B.M., Pizka, M. (2008) Tool support for continuous quality control. IEEE Software, 25 (5), pp. 60-67

23. Khorasani, R. (2008). Setting Up a Dashboard for Your Practice. JACR Journal of the American College of Radiology, 5 (4), p. 600

24. Nagy, P.G., Konewko, R., Warnock, M., Bernstein, W., Seagull, J., Xiao, Y., George, I., Park, A. (2008). Novel, web-based, information-exploration approach for improving operating room logistics and system processes. Surgical Innovation, 15 (1), pp. 7-16

25. Morgan, M.B., Branstetter IV, B.F., Mates, J., Chang, P.J. (2006). Flying blind: Using a digital dashboard to navigate a complex PACS environment. Journal of Digital Imaging, 19 (1), pp. 69-75

26. Rosow, E., Adam, J., Coulombe, K., Race, K., Anderson, R. (2003). Virtual instrumentation and real-time executive dashboards. Solutions for health care systems. Nursing administration quarterly, 27 (1), pp. 58-76

27. Bitoun, M. (2002) Show them the data. What do trustees need to know and how do they need to know it? Trustee, 55 (8), pp. 18-20, 33

28. Mettas, A., Rock, D. (2002). Intellectual capital: Utilizing the web for knowledge management and data utilization in reliability engineering. Proceedings of the Annual Reliability and Maintainability Symposium, pp. 379-385

29. Wan Mohd, W.M.B., Embong, A., Zain, J.M. (2008). A knowledge-based digital dashboard for higher learning institutions. Lecture Notes in Computer Science, 5212 LNAI (PART 2), pp. 684-689

30. Workman, M. (2008). An experimental assessment of semantic apprehension of graphical linguistics. Computers in Human Behavior, 24 (6), pp. 2578-2596

31. Dang, J., Tokiu, C., Hampel, K., Enke, U. (2008). Human workflows via document-driven process choreography. Proceedings-2008 International MCETECH Conference on e-Technologies, MCETECH 2008, art. no. 4483415, pp. 25-33

32. Heidrich, J., Munch, J. (2008). Goal-oriented setup and usage of custom-tailored software cockpits. Lecture Notes in Computer Science, 5089 LNCS, pp. 4-18

33. Rimini, G., Roberti, P. (2008). Business process monitoring: BT Italy case study. IFIP International Federation for Information Processing, 280, pp. 227-234

34. Morgan, M.B., Branstetter IV, B.F., Lionetti, D.M., Richardson, J.S., Chang, P.J. (2008). The radiology digital dashboard: Effects on report turnaround time. Journal of Digital Imaging, 21 (1), pp. 50-58

35. Williford, J., Chang, A. (1999). Modeling the FedEx IT division: a system dynamics approach to strategic IT planning. Journal of Systems and Software, 46 (2), pp. 203-211

36. Palpanas, T., Chowdhary, P., Mihaila, G., Pinel, F. (2007). Integrated model-driven dashboard development. Information Systems Frontiers, 9 (2-3), pp. 195-208

37. Hammer, M. (2007). The 7 deadly sins of performance measurement. MIT Sloan Management Review, 48 (3), pp. 19-28

38. Chieu, T.C., Chowdhary, P., Fu, S.S., Pinei, F., Yih, J.-S. (2006). Enterprise telesales opportunity pipelines performance management. Proceedings - IEEE International Conference on e-Business Engineering, ICEBE 2006, art. no. 4031684, pp. 433-441

39. Jobst, D., Preissler, G. (2006) Mapping clouds of SOA- and business-related events for an enterprise cockpit in a Java-based environment. ACM International Conference Proceeding Series, 178, pp. 230-236

40. Tudor, D., Walter, G.A. (2006). Using an agile approach in a large, traditional organization. Proceedings - AGILE Conference, 2006, 2006, art. no. 1667600, pp. 367-373

41. Britt, P. (2005). Finding the formula for successful CM: Managing financial content. EContent, 28 (1-2), pp. 38-42

42. Cassaigne, N. (2002). The dashboard: A knowledge conversion tool. IEEE International Engineering Management Conference, 1, pp. 292-297

43. Davis, F.D. (1993). User Acceptance of Information Technology: System characteristics, user perceptions and behavioural impacts. International journal of man-machine studies 38, pp. 475-487

44. Schwartz, R.B. & Russo, M.C. (2004). How to Quickly Find Articles in the Top IS Journals. Communications of the ACM. February 2004/Vol. 47, No. 2

45. Webster, J. & Watson, R.T. (2002). Analyzing The Past To Prepare For The Future: Writing A Literature Review. MIS Quarterly Vol. 26 No. 2, pp. xiii-xxiii

46. Yin, R.K. (1994). Case study research: design and methods (2$^{nd}$ ed). Thousand Oaks, Sage, cop. CA.

47. Cooper, D.R. & Schindler, P.S. (1998). Business Research Methods (7$^{th}$ edition). McGraw Hill International Edition, Statistical and Probability Series. ISBN 0071181091

48. Hevner, A.R., March S.T., Park, J. (2004). Design Science in Information Systems Research. MIS Quarterly Vol. 28 No. 1, pp. 75-105

49. Wijnhoven, F., Kraaijenbrink, J. (2008). Product-oriented design theory for digital information services – A literature review. Internet Research Vol.28 Nr.1, pp 93-120

50. Wijnhoven, F., Boelens, R., Middel, R., Louissen, K. (2007). Total Data Quality Management: A study of bridging rigor and relevance. Proceedings of the 15$^{th}$ European Conference on Information Systems (ECIS2007), pp. 925-937

51. Wieringa, R.J (Winter 2007-2008). Lecture Slides ‘Problem Analysis and Solution Requirements‘– 232080. University of Twente, the Netherlands

52. http://www.dashboardmd.com/2_revenue_cycle/s-billingdashboard.htm

## Shell References

‘REMOVED’

# Appendixes

## *Appendix A: Project positioning*

‘REMOVED’

## *Appendix B: Top 25 IS Journals*

The following table gives an overview of the world ranking of the IS journals, as constructed by [44]. Usually the Top25 is used for master graduation projects.

| WORLD RANK | TITLE | WORLD RANK | TITLE | WORLD RANK | TITLE |
|---|---|---|---|---|---|
| 1 | *MIS Quarterly* | 18 | *Communications of the AIS* | 35 | *Journal of Information Systems* |
| 2 | *Communications of the ACM* | 19 | *IEEE Computer* | 36 | *The Information Society* |
| 3 | *IS Research* | 20 | *Journal of Strategic IS* | 37 | *Journal E-U Computing* |
| 4 | *Journal of MIS* | 21 | *Admin. Science Quarterly* | 38 | *Info Resources Mgmt Journal* |
| 5 | *Management Science* | 22 | *Academy of Mgmt Review* | 39 | *Interfaces* |
| 6 | *IEEE Transactions* (various) | 23 | *Int'l Journal of E-Commerce* | 40 | *EM - Electronic Markets* |
| 7 | *Harvard Business Review* | 24 | *ACM Computing Surveys* | 41 | *Journal of CIS* |
| 8 | *Decision Sciences* | 25 | *Accounting, Management & IT* | 42 | *European Journal of OR* |
| 9 | *Decision Support Systems* | 26 | ACM SIG Publications | 43 | *Operations Research* |
| 10 | *Information and Management* | 27 | *IT and People* | 44 | *Int'l Journal of H-C Studies* |
| 11 | *European Journal of IS* | 28 | *IBM Systems Journal* | 45 | *Journal of the ACM* |
| 12 | *Sloan Management Review* | 29 | *OMEGA* | 46 | *Australian Journal of IS* |
| 13 | *ACM Transactions* (various) | 30 | *Journal of the AIS* | 47 | *Org. Behavior and Human Dec.* |
| 14 | *Data Base* | 31 | *Journal of Org., Comp. and EC* | 48 | *Behavior and IT* |
| 15 | *Organization Science* | 32 | *Human-Computer Interaction* | 49 | *Scandinavian Journal of IS* |
| 16 | *Information Systems Journal* | 33 | *Information Systems Management* | 50 | *Computer Journal* |
| 17 | *Academy of Management Journal* | 34 | *Int'l Journal of Man-Machine Studies* | | |

**Figure 11: World Ranking IS journals [44]**

## *Appendix C: Concept matrixes*

Since two types of searches are conducted, also two concept-matrixes have to be made. The first matrix lists the 21 relevant articles regarding Total Data Quality Management and its context in data quality management, mainly by Richard Y. Wang and Yang W. Lee. The second matrix contains the 22 relevant articles relevant for the dashboard subject (data quality and management). Important to mention is that the sequence of both listings is meaningless.

**TDQM**

Table 3 shows the concept-matrix for the relevant articles concerning Total Data Quality Management. By reading the articles more closely it appeared that some especially focus on the TDQM method and its cycle elements. While on the other hand, many articles mention it as important tool for data quality management. From this perspective also other important aspects of data quality are discussed; among its definition, dimensions and stakeholder management. Logically, these different topics have become the concepts.

From the matrix it can be concluded that a lot of attention is being paid to the definition and problems related to data quality. Though these address different areas, they are put together in one concept. Namely, many articles that discuss the definition also elaborate on data quality problems. Another interesting finding is the relation between Information Product, (Process) Management and Anchoring & Cycles. This basically refers to the institutionalization of data quality in the organization. To be realized by taking the product view on information and managing the production processes by implementing TDQM. Another interesting concept is stakeholder management; different roles and responsibilities are distinguished and the related know-how is discussed in the light of data quality.

[See next page]

**Table 3: Concept-Matrix TDQM authors**

| | Definition & Problems | Information Product | Quality Dimensions | System Design | (Process) Management | Anchoring & Cycles | Stake-Holders |
|---|---|---|---|---|---|---|---|
| [1] | | √ | | | | √ | |
| [2] | | √ | | | | √ | |
| [3] | √ | √ | | | √ | √ | |
| [4] | √ | | √ | √ | | | |
| [5] | √ | | | | | √ | |
| [6] | | | | | √ | √ | √ |
| [7] | √ | √ | √ | | | | √ |
| [8] | √ | | √ | | | √ | √ |
| [9] | √ | | | √ | | | √ |
| [10] | √ | | √ | | √ | | |
| [11] | | √ | | | √ | | |
| [12] | | √ | | | | √ | |
| [13] | | √ | | | √ | | |
| [14] | | √ | | | √ | | |
| [15] | √ | | | | | | |
| [16] | | | √ | √ | | | |
| [17] | √ | | √ | | | | √ |
| [18] | | √ | | | √ | | √ |
| [19] | | | | | √ | √ | |
| [20] | √ | | | | √ | | |
| [21] | | | √ | | | | √ |

## Dashboarding

Table 4 displays the concept matrix for the second search, respectively data quality- and management dashboards. Since Shell's Global Dashboard mainly concerns data quality, first only this key word was search for. But with insufficient results, the search was generalized by also incorporating management dashboards. As with the TDQM articles, some articles discuss data quality dashboard itself while others address it from a broader data quality management perspective. For example, topics like the (dis)advantages, tools & features and architecture.

A remarkable conclusion from the table is that many articles encompass a case study. An implemented dashboard is usually taken as case study and is used in order to elaborate or explain some particular aspects. Theories like project management and system design are discussed and applied on the case study. Because the variety of dashboards, the subjects differ and the focus is context dependent. Unfortunately there is no framework available that can be used for discussion or benchmarking. It can be concluded that most of the scientific literature is concerned with case studies instead of developing kernel theories that are generally applicable.

**Table 4: Concept-matrix Dashboarding**

| | (Dis-) Advantages | Project Management | Design | Tools & Features | Case Study | Products (packages) | Architecture |
|---|---|---|---|---|---|---|---|
| [22] | | | | √ | | √ | |
| [23] | | √ | | | | | |
| [24] | | | √ | | √ | | |
| [25] | | | √ | | √ | | √ |
| [26] | | | | | √ | | √ |
| [27] | √ | | | | √ | | |
| [28] | √ | | | | √ | | |
| [29] | | | | √ | √ | | √ |
| [30] | | | √ | | | | |
| [31] | | | | √ | | √ | √ |
| [32] | | √ | | | | | |
| [33] | | √ | | | √ | | |
| [34] | | | | | √ | | |
| [35] | | | | √ | √ | | |
| [36] | | √ | | | √ | | |
| [37] | √ | √ | | | | | |
| [38] | | | √ | | √ | | √ |
| [39] | | √ | √ | | | | |
| [40] | | √ | | | √ | | |
| [41] | | | | | √ | | |
| [42] | | √ | √ | | | | |

## *Appendix D: Critical Business Activities*

‘REMOVED’

## *Appendix E: Quality Dimensions in SIEP*

‘REMOVED’

## *Appendix F: System Architecture*

‘REMOVED’

## *Appendix G: Questionnaire*

| Perceived Ease-of-Use | Strongly disagree | | Neutral | | Strongly agree |
|---|---|---|---|---|---|
| 1. I find the data quality dashboard cumbersome to use. | 1 | 2 | 3 | 4 | 5 |
| 2. Learning to operate the data quality dashboard is easy for me. | 1 | 2 | 3 | 4 | 5 |
| 3. Interacting with the data quality dashboard is often frustrating | 1 | 2 | 3 | 4 | 5 |
| 4. I find it easy to get the data quality dashboard to do what I want it to do. | 1 | 2 | 3 | 4 | 5 |
| 5. The data quality dashboard is rigid and inflexible to interact with. | 1 | 2 | 3 | 4 | 5 |
| 6. It is easy for me to remember how to perform tasks using the data quality dashboard | 1 | 2 | 3 | 4 | 5 |
| 7. Interacting with the data quality dashboard requires a lot of mental effort. | 1 | 2 | 3 | 4 | 5 |
| 8. My interaction with the data quality dashboard is clear and understandable. | 1 | 2 | 3 | 4 | 5 |
| 9. I find it takes a lot of effort to become skillful at using the data quality dashboard | 1 | 2 | 3 | 4 | 5 |
| 10. Overall, I find the data quality dashboard easy to use | 1 | 2 | 3 | 4 | 5 |

| Perceived Usefulness | Strongly disagree | | Neutral | | Strongly agree |
|---|---|---|---|---|---|
| 1. Using the data quality dashboard improves the quality of the work I do | 1 | 2 | 3 | 4 | 5 |
| 2. Using the data quality dashboard gives me greater control over my work. | 1 | 2 | 3 | 4 | 5 |
| 3. The data quality dashboard enables me to accomplish tasks more quickly | 1 | 2 | 3 | 4 | 5 |
| 4. The data quality dashboard supports critical aspects of my job | 1 | 2 | 3 | 4 | 5 |
| 5. Using the data quality dashboard increases my productivity | 1 | 2 | 3 | 4 | 5 |
| 6. Using the data quality dashboard improves my job performance | 1 | 2 | 3 | 4 | 5 |
| 7. Using the data quality dashboard allows me to accomplish more work than would otherwise be possible | 1 | 2 | 3 | 4 | 5 |
| 8. Using the data quality dashboard enhances my effectiveness on the (data quality) job | 1 | 2 | 3 | 4 | 5 |
| 9. Using the data quality dashboard makes it easier to do my (data quality) job | 1 | 2 | 3 | 4 | 5 |
| 10. Overall, I find the data quality dashboard useful in my (data quality) job | 1 | 2 | 3 | 4 | 5 |

Additional Questions Asked

1. What is your first impression (is it an improvement)?
2. Does the dashboard support you better in prioritizing quality improvements?
3. Do you have any suggestions for further improvement?
4. Do you have final comments you would like to make?

## *Appendix H: Stakeholders*

'REMOVED'