Optimizing data collection for cost effective control of assets

Master of Science
In
Asset Management Control (AMC)

International Master’s School
&
Hogeschool Zeeland

AMC-student: Anko Brink
d.d.. 21 December 2007
Acknowledgements

My Business Supervisor:
Name: MSc W.A.P. Polle
Company: RNLN/Maintenance Establishment (Marinebedrijf)
Position: Head of Engineering Department Sub Warfare Systems

My Course Supervisor:
Name: MSc M.M.R. Kuijer
Company: Copernicos Groep BV
Position: AMC Change Agent

Michel and Wim, thanks a lot for supporting and inspiring me during my research!
I would like to say thanks to all the persons, especially my employer RNLN/ME, Hogeschool Zeeland, lecturers of the AMC-education and John Stavenuiter, who made it possible for me the last four years to perform this very interesting master study and final research.
Also thanks to the test team (Elselina, Koen, Loet, Mario, Michel, Noud and Ruud).
They have done good preliminary work, in advance of the research.
The research could not have been completed without sufficient response of the interviewees of the two surveys. I will say it again: thank you.
I am very thankful to Noud Goossens. He is a great colleague and was always very helpful in bringing me needed information.
The tremendous patience and understanding of my wife were important attributes to fulfil the AMC-study and this thesis. It should be a hopeless mission when these attributes were missing. Gerie, from now I have more time and you will get compensation enough. I promise!
Also thanks to my family, friends, colleagues and my four wonderful children Fabian, Ellen, Mario for your bibliography support and Ruud: I am very grateful for your hints, knowledge and experience in performing research.
Last, but not least, Mark, your English is better than my Dutch!
Executive Summary

Aim
The aim of this dissertation is to demonstrate the possible, different ways of data collection in relation to management of cost effectiveness during life cycle. The most important aim is to point out an easy (less complicated), reliable and efficient way of data collection.

Objectives
Two very important motives to write this thesis are:
1. A proven lack of data quality during data management activities;
2. The Stavenuiter (2002) case study, which clearly shows that it is difficult to obtain reliable system data.

The intended objectives are translated into the sub questions and are related to the methods of collecting data for cost effectiveness and its particular details about quality/cost of collecting data and effectiveness for management control.

Sub questions
The sub questions are formulated as follows:
Which companies are familiar with ILS (Integrated Logistic Support), LCM (Life Cycle Management) and CE (Cost Effectiveness)?
   a. How can quality of cost data and performance data be measured?
   b. Which methods are used in the researched companies to generate cost- and performance data?
   c. Are these methods appropriate and cost effective?
   d. Why are these methods not appropriate and cost effective?
   e. Is there a correlation (and if so, what is the relation) between cost of data collection and quality of data?
   f. What are alternative methods to generate reliable cost- and performance data?
g. Which factors do influence reliability of data significantly at the point of registration?

h. Which opportunities are available to optimise the generation of appropriate, reliable cost effectiveness data?

**Research approach**

An elaborate research at the field of data quality on a broad scale is needed in literature and companies with similar ILS/LCM-view as used in Royal Netherlands Navy/Maintenance Establishment (RNLN/ME), i.e. supporting and maintaining capital assets, especially maritime objects. The purpose is to get information about the ways and possibilities of collecting necessary and high-quality data in the most cost effective way.

Primarily, a literature review was performed. All existing information on a broad scale on the field of data management, data quality, cost effectiveness etc. was researched.

Secondly, after literature review the selection of companies to interview was started. Maintenance, repair and logistic support of capital assets (aviation, ship, tank etc.) are essential elements.

A quick scan based on existing knowledge of comparable companies shows that there might be a little number of companies; in the worst case none, which are working with LCM/ILS such as RNLN/ME does. This should be taken into account in the research methodology.

All companies do need management information, the same or different does not matter. The goal of the research is to define or demonstrate the possible, different ways of data collection in relation to manage LCM or cost effectiveness during life cycle. When a company does not manage cost effectiveness, but another key performance indicator (KPI), it is relevant to know how the company deals with data collection and reliability or quality of data in that particular area.

Several research methods are available (Baarda et al, 1997). One of them is participating observation which is not chosen, because this requires being present with data collection activities.
The selected companies were approached by means of e-mail and/or telephone in order to get an e-mail address from a relevant interviewee for performing the objected research via Internet. The interviewees are responsible for, involved with or well-informed about collection and registration of the data.

Pre-structured interview questions are selected as primary research method. The interviews were arranged as much as possible in accordance with the SPIE-method (NIVE, 1999). The questions have a relation with the research topic in the following category:

- Situation
- Problem
- Implication
- Efficiency

Finally, the interviews were carried out via Internet (Surveyworld.net) in order to get information.
# Table of Contents

Acknowledgements .................................................................................................................. 2  
Executive Summary .................................................................................................................... 3  
Aim ........................................................................................................................................ 3  
Objectives ................................................................................................................................. 3  
Sub questions ............................................................................................................................ 3  
Research approach .................................................................................................................. 4  
Table of Contents ...................................................................................................................... 6  
1. Introduction .......................................................................................................................... 8  
   General .................................................................................................................................. 8  
   Clarification of title of the dissertation .................................................................................. 8  
   Definitions .............................................................................................................................. 9  
   Conceptual Framework ......................................................................................................... 10  
   Asset Management Control ................................................................................................. 11  
2. Research Methodology ........................................................................................................ 13  
   Problem description ............................................................................................................... 13  
      Examples of bad data quality ............................................................................................ 13  
   Organization within RNLN/ME ............................................................................................ 15  
   Data collection within RNLN/ME ......................................................................................... 17  
   Product Data Management .................................................................................................. 20  
   Integrated Logistic Support ................................................................................................. 21  
   LCM-information systems within RNLN/ME ...................................................................... 22  
   Data quality within RNLN/ME ............................................................................................. 25  
      Example of collecting bad data within RNLN .................................................................. 27  
   ERP-implementation within RNLN/ME ............................................................................. 28  
   Overview of the problem ...................................................................................................... 30  
3. Theoretical research ............................................................................................................. 31  
   Data ...................................................................................................................................... 31  
   Hierarchy of Data .................................................................................................................. 31  
   Entities, attributes and relations .......................................................................................... 32  
   Database-models ................................................................................................................... 33  
   Database management system (DBMS) .............................................................................. 34  
   Data management ................................................................................................................ 35  
   Data quantity ....................................................................................................................... 37  
   Data quality, general view ................................................................................................. 38  
   Data quality requirements ................................................................................................. 46  
   Data collection ..................................................................................................................... 49  
   Enterprise Resource Planning (ERP)-system ....................................................................... 53  
      Advantages of ERP Systems ............................................................................................ 55  
      Disadvantages of ERP Systems ....................................................................................... 55  
   Data warehousing .............................................................................................................. 56  
   Cost of data collection ......................................................................................................... 59  
   Overview of the theoretical research .................................................................................. 62  
4. Practical research ................................................................................................................ 63  
   Preparations ......................................................................................................................... 63  
   Performance ......................................................................................................................... 65  
   Results .................................................................................................................................. 66
Overview of the practical research ................................................................. 69
5. Characteristics of a cost effective data collection ........................................ 70
6. Summary ........................................................................................................ 72
7. Conclusions ..................................................................................................... 73
8. Recommendations .......................................................................................... 76
9. Suggestions for further work ................................................................. 77
10. Abbreviations ................................................................................................. 78
11. Bibliography .................................................................................................. 80
   Internet sources ................................................................................................ 89
   Appendix A. Research questions ....................................................................... 90
   Appendix B. Results of research ..................................................................... 94

List of Figures

Figure 1. Conceptual framework ................................................................. 10
Figure 2. Organization DMO ........................................................................ 15
Figure 3. Organization RNLN/ME ................................................................. 15
Figure 4. ILS - mutual agreement ................................................................. 17
Figure 5. Life cycle phases of an asset ........................................................... 18
Figure 6. PDM-system RNLN ....................................................................... 21
Figure 7. ILS-database ...................................................................................... 22
Figure 8. The conceptual system LCA data set ............................................. 23
Figure 9. Cost Effectiveness model RNLN/ME (old situation) ....................... 24
Figure 10. Hierarchy of Data .......................................................................... 32
Figure 11. Attributes and entities .................................................................... 33
Figure 12. Relational database (Failure Database, example, partial) ............... 34
Figure 13. Data identification process ............................................................ 36
Figure 14. Feedback information system (Orr, 1998) ..................................... 44
Figure 15. Data quality requirements (CIHI) ................................................ 47
Figure 16. Work required to access seven units of information (Inmon, 1997) .. 57
Figure 17. Architecture of a distributed data warehouse (Bose, 2006) .......... 57
Figure 18. Data management process (Stair, 2006) ........................................ 58
Figure 19. Data quality cost ............................................................................. 61
Figure 20. Average data quality per respondent ........................................... 67

List of Tables

Table 1. Products vs. information manufacturing (Wang, 1998) ...................... 43
Table 2. Dimensions of high-quality data ..................................................... 48
Table 3. Requirements for reliability databases ............................................ 50
Table 4. Quality dimensions question 8 ....................................................... 64
Table 5. Overview of responses ................................................................. 66
Table 6. Overview of possible causes of poor quality data ......................... 68
1. Introduction

**General**

Assets of Royal Netherland Navy (RNLN) are being maintained during life cycle by Royal Netherland Navy Maintenance Establishment (RNLN/ME, in Dutch: Marinebedrijf). These assets are capital investments and an economical topic. From economic and operational point of view it is extremely important to support these capital assets as cost effective as possible. Operation and support cost during life cycle can be a multiple of acquisition cost (Jones, 1995).

The main objective of cost effective maintenance of assets during the life cycle is maximizing system effectiveness in relation to minimizing life cycle cost (Stavenuiter, 2002).

It is extremely important to arrange well-based support decisions. These decisions depend on information, which is processed out of data. The challenge is to obtain high-quality data in a cost effective way.

**Clarification of title of the dissertation**

Cost effective control of assets requires cost and performance information during total life cycle, from asset need to asset disposal. This information is created out of the data collection of data registration systems. In line with managing assets, collection of supporting data should be cost effective and optimised, because this is also an economical topic.
Definitions

There is a difference between data, information and knowledge. These are not interchangeable terms. Definitions from Albrecht (2001):

Data is the atomic raw material of human craft. It’s the irreducible symbolic level, where alphanumerical encoding allows us to move and store data without regard to its meaning.

We distinguish two sorts of data: master data is static and transactional data is dynamic. Master data may include data about organization, employees, customers, products, materials, configuration, inventory, suppliers, etc. which often turns out to be non-transactional, persistent in nature.

Information is the meaningful arrangement of data, which creates patterns and activates meanings. It’s the words, pictures and sounds rather than the data. Information is dynamic. It exists at the point of human perception.

Knowledge is the value-added content of human thought, derived from perception and intelligent manipulation of information. Knowledge is transcendent; it exists uniquely in the mind. It is the basis for intelligent action.

Cost effectiveness refers to the comparison of the relative expenditure (cost) and effect (performance), in this case concerning capital assets maintained by RNLN/ME. As a result of the above-mentioned, cost effectiveness will be higher when support cost will decrease or performance will increase, emphasizing the relationship between cost and performance.
Conceptual Framework

The conceptual framework for data collection looks like this:

Concerning the asset, resources are coming in and performance is coming out. Resources will result in financial cost and performance is what the customer wants and responds to its operational needs.
We want to know information about resources and performance and therefore we collect data to measure the effectiveness of our asset.

We collect all sorts of data about the asset in our Database Management System (DBMS). Data collection is the beginning of the decision making process. We bring the master data and transactional data (cost & performance data) into a data warehouse or Management Information System (MIS). We want to process the data to our information needs and we perform an analysis.

After having the objected information we obtain the knowledge to support decisions. This can be done in a Decision Support System (DSS), also mentioned Decision Tool. It is extremely important to use this knowledge as input for asset management. Knowledge will easily be exchanged with communication in all phases, will enforce team building and finally will positively influence cost effectiveness.

The research is focused on the way of data collection and its impact on decision making. The conceptual framework supports the structure of this dissertation. Along the indicated path in this framework points were placed to measure certain feelings, thoughts and opinions.

**Asset Management Control**

Asset Management Control (AMC) enables us to reach the target of cost effective control of assets.

Asset management is a management approach to manage all company processes (specify, design, produce, maintain and dispose) needed to achieve a capital asset capable to meet the operational need in the most cost effective way for the customer/user (Stavenuiter, 2002).

The target will be to reduce the resources and to bring the performance to a level which has been set as base-level or higher if possible. In other words: the target is to maintain or support the asset as cost effective as possible during life cycle.

We want answers on questions like (ref. Blanchard B., 1998a):

- What is the *true* performance of an asset during a certain period?
- How much did logistic support *really* cost during a certain period?
- Which corrective and preventive maintenance activities are to be expected?
- Are the initially specified requirements being met?
Is the system cost effective?

Being able to give answers to these questions, we must assemble information about performance and cost and at least collect the following data from the asset:

1. Master data
   a. Configuration data
      • Asset (system)
      • Equipment (installation)
      • Parts/components
   b. Prices
      • Acquisition cost
      • Hourly pay
   c. Baseline data for performance and cost (STORM)

2. Transactional Performance data (= system effectiveness data)
   a. Availability = \( \frac{U}{U+D} \) ……most used formula for continuous operation
      • \( U \) = Uptime (h) = \( \sum \) (Operation Times)
      • \( D \) = Downtime (h) as result of failure = \( \sum \) (WT, ST, RT)
         • WT = Waiting time (h)
         • ST = Search time (h)
         • RT = Repair time (h)
   b. Reliability = \( \exp(-\lambda \cdot T) \) ……assuming \( \lambda \) = constant (!)
      • \( \lambda \) = failure rate = \( \frac{1}{MTTF} \) = \( \frac{N}{U} \)
      • \( T \) = mission time (h)
      • \( N \) = \( \sum \) failures
         • Failure criticality
         • Failure effect
         • Failure mode
   c. Capability = \( f \) (installation test)

3. Transactional Cost data = \( \sum \) Cost (Materials, Wages, Services, Remaining)
   a. Materials cost = \( \sum \) Cost (failed parts for consumables)
   b. Wages cost = \( \sum \) Cost (repair hours * hourly pay)
   c. Services cost = \( \sum \) Cost (service wages, hours, parts etc.)
   d. Remaining cost
2. Research Methodology

Problem description

Just as most companies, RNLN/ME has a data quality problem, see for details: Data quality within RNLN/ME.

This research was carried out in order to approach the data quality problem and to try to bridge this problem.

Data quality is a hot issue. Data quality is often subject of talking and reading, but mostly in the negative way. There are many examples of the negative effect of bad data. Below is listed a number of examples to enlighten the problem and to give an impression about the size of the problem.

Examples of bad data quality

- The most important Dutch Controlling Governmental Department did perform an investigation to the reliability of management information about operational readiness of Dutch Defence-assets (Algemene Rekenkamer, 2006). It seemed that there are shortcomings in the way the data is collected. There are different collection methods with ambiguous criteria as result of a lack of cooperation, definitions and system development. The data measuring systems contain wrong or incomplete information because of dysfunctional ICT-systems. Data is so dirty that it is not possible to produce reliable monthly reports.

This situation is not new: 15 years earlier (in 1991!) the Algemene Rekenkamer already reported that operational readiness of Land-Systems was hard to measure because of a lack of a good working information system (ref. Defensiekrant).

- According to many data warehouse industry experts, anywhere from 5% to 50% of the data in company databases is missing or inaccurate (Inmon, 1999).

- 60% of data-integration and data-warehouse projects get delayed, go over budget or fail because of poor data quality. 60% of invoices contain errors, and 30% of customer data is inaccurate. "Most errors exist because of poor standards at the data-entry point (Fest, 2005).

- 53% of the 750 IT professionals and business executives surveyed by the Data Warehousing Institute last year said their companies had experienced problems and suffered losses or increased costs because of poor-quality data, up from 44% in a similar survey in 2001 (Whiting, 2006).

- Almost two-thirds of 501 medium and large companies recently surveyed by Computerworld reported problems resulting from inaccurate outdated or missing data (Knight, 1992).

- Experts estimate that anywhere from 10% to 30% of the data flowing through corporate systems is bad, inaccurate, inconsistent, formatted incorrectly, entered in the wrong field, out of a value range, and so on (Goff, 2003).
Bad data has a great influence on waste of resources and performance:

- Bad data refers to incorrect or duplicated customer information which cost companies more than $600 billion annual (!) (Fitzgerald, 2005).
- According to a General Accounting Office report, taxpayers paid $2 billion because information provided by the nation's banks on defaulters had not been entered properly into a centralized database at the Department of Education (Knight, 1992).
- Inaccurate data regularly forces planes to take off one-half to one-third full, costing airlines tens of millions of dollars (Knight, 1992).

When you know your data is bad, you can avoid the problem and create for instance a fictive world:

- In the movie War Games, set in the U.S. Strategic Air Command Centre under a mountain near Colorado Springs, there is a critical scene: on a huge map of the globe are arrows indicating large numbers of nuclear missiles approaching the U.S. from the Soviet Union. The military official in charge is trying to decide if he should ask the President for permission to launch a retaliatory attack because the technicians are telling him the threat is real. At this point, the scientist who originally designed the system (and who knows that something is wrong with the system itself) says: "General, what you see on that board is not reality; it is a computer-generated hallucination!" (Orr, 1998).
**Organization within RNLN/ME**

To understand the problem of data collection within RNLN/ME we will first describe the organization. RNLN/ME is part of the department “Sea Systems” within the Ministry of Defence. The Ministry of Defence is structured as shown in the figure below:

![Figure 2. Organization DMO](image)

The department “Sea Systems” is structured as shown in the figure below:

![Figure 3. Organization RNLN/ME](image)
After decision-making on a high level of Ministry of Defence (MoD) acquisition of assets takes place within DMO (Sea Systems). The relation with the maintainer (RNLN/ME) is in this part of the life cycle not very strong, but Lobregt et al (2004) recommended to involve RNLN/ME more in the beginning of the project, because all decisions at the start of the life cycle have impact on the maintainer during life cycle. Also much collected data is important for both parties. Communication and joint use of database systems can improve asset support.

RNLN/ME is maintaining and supporting many different and complex systems which are difficult to manage during life cycle. Greengard (1998) comments that the more difficult it is to manage systems and the more systems there are to manage, the greater the odds that data errors creep into the picture and systems become more expensive to operate.

Outsourcing, competition and public private cooperation has risen over the past years, but real competition on private market is not yet present. As long as this situation exists, there is no real motivation or drive to improve data and information quality.

This must be applied of course in a competitive environment. To be ready for the future we should be prepared to operate in a more competitive environment and work with high-quality data.

This means that the organization should be changed from a scientific management with a hierarchical structure to a rational process model with a performance-based strategy. It must change in a more efficient, professional and competitive way (market-culture ref. Cameron, 1999). In order to be cost effective and competitive to other organizations the output of all employees must be measured and eventually rewarded if needed (“earned”).

Drive and motivation for user to pay attention to cost is often not high enough. The operational user (RNLN) is still only interested in operational availability of the assets and cost is less important for them. Management, however, is aware of cost effectiveness but lack of competition-drive will not stimulate this view.

But with an inappropriate and ignored data quality the cost might be very high. Cost is even higher than needed when assets are maintained “as new”, sitting on the safe side. A certain deal of work is “insourced”, especially from ICT-services, for a great part Defence Telematics Organization (DTO).
CPIM is a division of RNLN/ME and is the department where cost effectiveness data is being gathered\(^1\).

Data Management delivers the structure of the web portals and arranges, groups and presents the available product-related data, also by insourcing of DTO. Systems Management concerns plans on system/ship-level to develop LCM-information out of available data.

**Data collection within RNLN/ME**

RNLN/ME is a company for logistic and technical support of Defence (mainly naval) material and a department of “DMO/Sea Systems” within the Ministry of Defence. RNLN/ME is performing asset management with the Integrated Logistic Support philosophy (ILS-Handbook, 2004). With this approach life-cycle support of assets can be managed. Generating data starts with a need as a result of mutual agreement between designer, maintainer and user (see figure below).

![Figure 4. ILS - mutual agreement](image)

User, maintainer and designer deliberate during the life cycle of an asset and analyse collected asset data for information about performance and cost. ILS starts with the acquisition planning of an asset and continues throughout its useful life and must be performed as cost effective as possible.

\(^1\) Gathering data is defined as making downloads from legacy-databases
Cost effective control of assets requires cost and performance information during total life cycle, from asset design to asset disposal, but especially during the utilization phase. Data collection of expensive, failure-sensitive and mission-critical parts is most important.

This information is created from the data which is gathered out of data registration systems. Oxbrow (2006) says: “Good asset performance means bringing together data for the whole life cycle of the asset and the customers it serves”.

The life cycle phases of an asset can be represented by the following sequence (RNLN, 2004, ILS-Handbook):

![Figure 5. Life cycle phases of an asset](image)

Life cycle starts with the operational need for an asset with certain previously settled requirements and specifications.

During the life cycle the asset must be supported in a certain, structured way, because there is a need to control cost effectiveness of capital assets. Only in a structured way it is possible to approach this problem.

There is another important distinguishing aspect of LCM-/cost effectiveness-data: not only the availability of high-quality data must be ensured, but collection of this data should be cost effective in itself! Data collection cost should ultimately result in reasonable increasement of asset performance.

Data collection and production of management information is an economical issue and must be treated as a capital asset (see chapter Data quality, general view).

The life cycle phases of an asset (see figure above) can also be applied to data, from user need to disposal: capture, storage, update, cleanse, transmission, access, archive, restore and deletion (Kim et al, 2003/ Wang et al, 1998).

It must be appropriate, especially available and reliable and last but not least, it should be cost effective, not time- or energy-consuming, to collect and analyse. This means
that the essential data to manage cost effectiveness is available and that this data meets certain previously settled requirements.

Achieving cost effectiveness takes place during utilization phase (operational period) in which an asset has to meet the operational needs.

Specifically during utilization phase there is a need for data and information to achieve cost effectiveness, for instance:

- **Material data (master data):**
  - configuration
  - prices
  - spare parts
  - cost (baseline)
  - performance requirements (baseline)

- **Actual cost data**
  - acquisition
  - maintenance
    - hours
    - materials
  - services

- **Actual performance data (also called System Effectiveness data)**
  - availability
  - reliability
  - capability

Actual cost data and performance data (transactional data) is collected during life cycle to make trends visible.

Physical (digital) and configuration data must be available out of a Material Registration and Configuration System, which should have a relation with other databases, for example logistic and operational databases. All these databases do have a specific function, but in practice they have a certain relation with each other, because all databases do have a certain relation with material.
A vision on Cost Effectiveness of assets requires not only information out of planned data from the maintenance plan but also operational (actual) performance data and cost data from other data collection systems (hours, materials and services) bringing the planned and actual data together in a system overview.

It can be very difficult to achieve reliable and relevant logistic data. This may become a serious threat to successful control implementation. Within LCM-systems design the following safeguards have been included to combat this problem (Stavenuiter, 2002):

- a structured data set with respect to the logistic organization;
- understanding of the importance by the actors;
- clear objectives/targets for all actors concerned;
- data collection is reduced to an absolute minimum;
- online and (almost) real time feedback information is ensured.

Data for cost effectiveness information must be available both as actual and baseline data and are divided into system effectiveness (or performance data), life cycle cost data and material data.

Actual performance data is acquired from naval ships and consists of a failure collection database, called MATRACS. The maintainers on the ships take care of the data collection of failures and uptimes. Out of this data availability and reliability can be calculated (Blanchard, 1998a). Capability was not fixed yet and temporarily set to 100%.

Actual cost data is also acquired from MATRACS; further a logistic database, called VAS/LOGDOC and a production database, called BBS. Collection of VAS/LOGDOC - and BBS-data is mostly done by material planning - and production employees.

Material data is acquired from the logistic database VAS/LOGDOC (articles) and material database MBS (configuration of assets above level of articles).

All before mentioned databases are stand-alone and not or nearly integrated. Data for producing information is downloaded from these databases with a request to the owner of the database.

**Product Data Management**

The function of Product Data Management (PDM) is to support logistics management of complex systems (Jones, 1995).
Downloaded data from legacy-systems is uploaded in PDM. The purpose of PDM within RNLN is to provide more accessibility and control to asset data during the life cycle of this asset.

PDM is based on web-technology and consists out of a three-tier architecture: a user-, business- and data-layer (see figure below).

![Figure 6. PDM-system RNLN](image)

**Integrated Logistic Support**

ILS-handbook (RNLN, 2004) describes an integrated data information system, like MIL-STD-1388 2B and DEF-STAN 00-60 (LSAR), for storage of all support data during the life cycle of an asset.

Integrated Logistic Support (ILS) is a proven and accepted method within RNLN/ME. ILS is defined as the disciplined, unified management of the technical logistics disciplines that plan and develop support for military forces (Jones, 1995). The ILS-database looks like this (see figure below):
Baseline master data (ILS-data) is stored in a relatively new information system for maintenance plans, called STORM (Systems Technical Overview of Reliability and Maintainability), and contains the planned (master) data for performance and cost.

**LCM-information systems within RNLN/ME**

RNLN/ME uses a lot of information systems for support and management of navy-material like logistics, finance, configuration, production, planning, failures and so on. All these information systems are old hierarchical legacy databases, are more or less stand-alone databases, not integrated and they all produce data.

In fact there is at this moment a transition in the data-analysis process for Life Cycle Management.

Brink developed already in 1997 an integrated relational database system for cost effectiveness information where data from different databases were gathered in a stand-alone MS-Access data warehouse, processed and analysed. This was called the Availability Killer/Cost Driver (AK/CD)-information system.

In this situation until 2006 about 1,05 FTE (full time equivalent) were responsible for gathering and processing of available cost effectiveness data. This data was gathered
from downloads of legacy-data and used as management information for engineers and managers. It served also as input for primary versions of AMICO (Asset Management Information & Communication). An “old” primary version of AMICO (Windows-application) is operating since 2002.

The figure below represents the conceptual system LCA data set which has been used in AMICO (ref. Stavenuiter, 2007):

![System LCA Data Set](image)

Figure 8. The conceptual system LCA data set

The database-overview for defining Cost Effectiveness in AMICO (old situation) looks like this:
In the new situation web-portals will be used. Cost effectiveness data is planned to be used in a web-based information system, the follow-up of AK/CD, called CIPIER (Cost Information & Performance Information Effectiveness Report), where information about Performance Killers and Cost Drivers (PKCDs) should be processed and presented. The “new” web-based information systems AMICO and CIPIER are in the development stage and not yet operational.

Davis (1985) formulates for development of new information systems several contingencies (project size, structure, task perception, knowledge) and requirements (technical, economical, motivational, timely and operational). These contingencies and requirements should be used to ensure information quality.
Data quality within RNLN/ME

Legacy-data is invariably in worse condition, as Atre (1998) says, and we know RNLN/ME is not exceptional in its data quality in a broad view.

Data quality is subject of many research-activities within RNLN/ME because the common idea of data quality is not very good. To improve or to correct data quality has often been tried, but without result.


Bais and Kalf (2000), but also Verschuren and Wilms (1994) did research several failure reporting systems like PRAWDA/RAPID within RNLN, the predecessors of MATRACS. They concluded that there was a great doubt about the reliability and usefulness of these systems.

In the meantime MATRACS was introduced in RNLN in 2003. So far as known, official research to the data quality of this information system has not yet been performed, but experience has learned that data quality has made progress.

Calbo and Polle (2005) describe the poor data situation within RNLN/ME and the complexity and problems of data migration towards an ERP-system. Collection of data causes about 30 % time-loss during promotion researches in the past. In many cases the data seemed to be incomplete and inconsistent. Along the model of the NATO Product Data Model (NPDM) the construction for a data warehouse was made visible. The NPDM is a conceptual data model. It defines a common set of data definitions and data structures to support Defence System technical information management, throughout the system life cycle, in the context of NATO nations and NATO industries.

Gameren (1998) describes a possibility to more integration of data within RNLN in a research to improve the user-friendliness and effectiveness of performance-information systems.

Hoogenboom (2002) describes the (un)reliability of production process data (BBS) within RNLN/ME. It seemed that the quality of data is poor, caused by complexity of the
system and lack of responsibility, standards and control. Verification of the data is (almost) impossible which results in manipulating the data.


Schröder (2003) indicates which problems with causes and consequences exist in five legacy databases. He also indicates the problems during the implementation of an ERP-system and in which way data can be selected, cleansed and related for data migration towards an ERP-system.

Stam and colleagues of Engineering Departments already established in 1996 that closer research at data from information systems showed that the consistency and quality of this data is bad.

Stavenuiter (2002) conducted a research within RNLN/ME to “Cost effective management control of capital assets”, a study of an integrated Life Cycle Management approach.

The case study (Stavenuiter, 2002, page 178) shows that it was difficult to extract the necessary data. About half (50 %) the imported legacy data had to be corrected manually, because the sequential management information was questionable. The necessity of improving this data is obvious. In the work of Stavenuiter, however, the aspect of data collection, methods and quality is underexposed. It is very difficult to achieve reliable and relevant logistic data. This may become a serious threat to successful control implementation (Stavenuiter, 2002).

As Stavenuiter says (p.64): “ICT is considered to be the most vital asset management facility. It must provide relevant and reliable information which is attainable for all logistic actors over the life cycle. In fact this implies not only the ICT-infrastructure but also the management of all data concerning the asset, its components and logistic activities”.
Example of collecting bad data within RNLN

Beyond a real occurred example of failure data collection within RNLN:

\[
\text{Failures are recorded with an attribute, called “failure criticality”, which definition is: when a failure has critical impact on a mission-critical function of an asset then it is a critical failure.}
\]

\[
\text{Some operators are not well aware of the criticality definition and record a failure “critical” when only a component (part of installation) fails which has no critical impact on the mission-critical function of the asset.}
\]

\[
\text{In that case, the calculated installation performance will be lower than the real.}
\]

This example emphasizes, that education and training with the software tool are important, but also definitions as for instance the Report of Algemene Rekenkamer (2006) also indicates.

It is clear that the ideal situation of data collection, processing and managing during LCM in RNLN/ME has not yet been reached. This is an understatement, because the ideal situation is far from the existing situation. The existing databases which are in use within RNLN/ME are old legacy-systems without the necessary integration of data-types.

Even in the new (web-based PDM) situation with an advanced solution like three-tier-architecture, Enterprise Application Integration and more sophisticated terms, the actual data will still come from the old legacy-systems. Without further data-cleansing the result will be the same or less (old saying: “garbage in, garbage out”). Besides: data cleansing is not a good solution, because it is an indication of bad data (Redman, 1995)

In industry there are many PDM systems designed for a general-purpose database to support logistics management of complex systems (Jones, 1995). Experiments in the RNLN have demonstrated that the large scale of this database makes it difficult to operate and manage. It contains more tables than the RNLN requires and it asks for data that has already been stored in other databases. For this reason a more flexible information management system is suggested based on distributed database principles (Stavenuiter, 2006).
Even another risk has been introduced. Complexity is related to cost of supporting and managing of the system (Inmon, 1998). The more sophisticated the data collection and reporting system, the more expensive it will be to implement. (DOE/PBM, 2001).

Anyway, besides required resources, the coming years there is a new chance to improve data quality within RNLN/ME!

**ERP-implementation within RNLN/ME**

RNLN/ME is part of the Dutch Defence project and at the beginning of an ERP-implementation (SAP) which will bring a total change in DBMS.

The conversion to an ERP system is not just a data extraction, cleansing, transformation, and populate process to effectively implement an ERP system. An organization also needs a strategy and a plan (Vosburg, 2001). This can be applied to Dutch Defence Companies also. Excessively. A huge number of documents are available, describing strategy, planning, migration methods etc.

For RNLN/ME a Project Initiation Document (PID), version 0.4.0 (21 September 2007) has been produced, which however is not yet accorded.

An important pre-condition for success is mutual cooperation and communication within departments, not only within RNLN, but total Dutch Defence (Schröder, 2003).

The ERP-system is used as supporting tool for finance, logistic and plant maintenance processes. The implementation of the ERP-system is extremely complex because of the many different operating Defence-parts and many different products within those parts. Implementation takes several years and is only at the beginning.

As we have seen before the legacy-data is not clean and must be cleaned before data migration starts. It must directly be corrected in the source (the legacy-databases) or extracted out of the source to a special server, directed into the environment where transformation takes place (cleansed and completed), tested and finally after approval loaded in the ERP-system. This is a labour-intensive and long-term activity where start (data-audit) and finish (data mapping, object model) well must be defined.

The data migration takes place in a “Data Migration Street” (DMS), high capacity computers with a special data migration software tool installed on it (ETL-tool). In this
DMS “data rules” will be developed to help realize and maintain the required data quality for upload to SAP.

The choice for the data migration tool has been made (Informatica).

There are experiences with data-cleansing in several Defence-companies. Besides experience with Informatica within Dutch Defence (DMO) in developing a management information system (MIS/WSM), a growing number of Defence-organizations have experience with ERP-implementation. Examples are the Defence organizations in countries such as the United States, Canada, Australia, New Zealand, Israel, Singapore, Finland, Norway and Denmark.

The German “Bundeswehr”, which comprises the German Armed Forces and its corresponding civil administration, successfully deployed integrated SAP-software from the SAP for Defence & Security industry solution in 2004 to drive human resources, logistics (including supply chain management as well as maintenance) and financial management processes.

Newing (2004) describes the method used in the British Ministry of Defence (MoD), United Kingdom. Data from air, land and sea were brought together and cleansed by The Cleansing Project (TCP). TCP was set up on a high level and started in year 2001. Cleansing was done by data set, not by system.

The necessary information for asset management will not be present in the ERP-system. However there is a Life Cycle Product Management (LPM)-module but the available information will, very likely, not meet with the objected or needed information within RNLN/ME (Perik, 2007).

We should have advantage in ERP-implementation and data migration from these “lessons learned”.
Overview of the problem

RNLN/ME is supporting assets of RNLN with a Life Cycle Management-approach. This requires data for the production of information, but there is a data quality problem. As we expect the data collection problem does not only exist within RNLN but the problem exists worldwide, despite all initiatives for improvement. The data collection problem is one of the main causes of misinformation and huge waste of resources.

There are many ways in which data quality has not the desired level and bringing data quality on this level can cost so much that it is not cost effective. An elaborate theoretical and practical research to the causes, influences and dependencies is needed for a movement to the direction of improvement.

Implementation of an ERP-system could be a possibility and a new challenge for RNLN/ME to reach the desired data quality level.
3. Theoretical research

To analyse the way data is collected it is necessary to review all connected aspects. In this chapter we review aspects like data, data management, data quantity, data quality and the way data can be collected, stored and processed. We discuss influences, impacts and dependencies.

Data

Data is at the core of all business: financial, medical, technical etc. and it stands at the beginning of producing information and making decisions. Mostly it is not enough to receive raw data and take decisions. Data has to be processed, analyzed and presented into information. This can be done for example in a Management Information System (Davis, 1985) or Data Warehouse (Inmon, 1997).

We distinguish two sorts of data: master data and transactional data. Master data is static and transactional data is dynamic. Master data may include data about organization, employees, customers, products, materials, configuration, inventory, suppliers, etc. which often turns out to be non-transactional, persistent in nature. Daily operations, planning, and decision-making functions in organizations are increasingly dependent on transactional (also historical) data. This data is entered electronically and manually and then organized, managed and extracted for analysis and decision-making (Vosburg, 2001).

Metadata is data about data. An item of metadata may describe an individual datum, or content item, or a collection of data including multiple content items. The metadata required for effective data management varies with the type of data and context of use (Wikipedia).

Hierarchy of Data

Data is organized in a hierarchy that begins with the smallest piece of data used by a computer, for example a single character such as a letter or number. Characters form
fields such as names, telephone numbers, addresses, and purchases. A collection of fields makes up a record. A collection of records is referred to as a file, also called table. Integrated and related files make up a database (Stair, 2003).

Below an example of the hierarchical structure of a (failure)-database:

![Diagram of the hierarchy of data](image)

**Example: Failure database**

- **Database files:**
  - Material file
  - Installation file
  - Failure file
  - etc.

- **Material file:**
  - Material Number, Name
  - 4410997165682 Coil pad
  - 4410997165685 Sleeve
  - 441099714301 Outer coil

- **Material record:**
  - Material Number, Name
  - 4410997165682 Coil pad

- **Field Name:**
  - Material Number:
  - 4410997165682

- **Character:**
  - 4

**Entities, attributes and relations**

An *entity* (also called record) is a class of people, objects, or places for which data is stored or collected. Examples include employees and customers. Consequently, data is stored as an entity, such as a material database and a failure database.

An *attribute* (also called field) is a characteristic of an entity. For example, the name of a material is an attribute of a material. A specific value of an attribute is referred to as a data item.

There can be a *relation* between entities and attributes of different tables. The relations can be represented in an entity relation diagram. The relation can be 1:1, 1:n, n:1 or
n:n. By means of this relation we can make a relation between two files and we join to the key field. In the case of the material file the primary key will be the material number. When there is more than one key field, we distinguish primary key, secondary key etc.

<table>
<thead>
<tr>
<th>Material number</th>
<th>Part number</th>
<th>Vendor</th>
<th>Material description</th>
</tr>
</thead>
<tbody>
<tr>
<td>4410997165682</td>
<td>...........</td>
<td>........</td>
<td>Coil pad</td>
</tr>
<tr>
<td>4410997165685</td>
<td>...........</td>
<td>........</td>
<td>Sleeve</td>
</tr>
<tr>
<td>4410997514301</td>
<td>...........</td>
<td>........</td>
<td>Outer coil</td>
</tr>
<tr>
<td>etc.</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Figure 11. Attributes and entities**

**Database-models**

The structure of the relationships in most databases follows one of three logical database models: hierarchical, network, and relational.

- **A hierarchical** database model is one in which the data is organized in a top-down or inverted tree-like structure. This type of model is best suited for situations where the logical relationships between data can be properly represented with the one-parent-many-children approach.

- **A network** model is an extension of the hierarchical database model. The network model has an owner-member relationship in which a member may have many owners, in contrast to a one-to-many-relationship.

- **A relational** model describes data using a standard tabular format. The relational database model is the most popular category. All data elements are placed in two-dimensional tables called relations, which are the equivalent of files. Data inquiries and manipulations can be made via columns or rows given specific criteria.

Network database models tend to offer more flexibility than hierarchical models. However, they are more difficult to develop and use because of relationship complexity. The relational database model offers the most flexibility, and is most popular during the
early 2000s (Kirkwood, 2006). In a relational database in which all data is integrated, data input is only once occurring and chance on dirty data is small.

An example of a relational database model:

![Relational database model](image)

**Figure 12. Relational database (Failure Database, example, partial)**

Note: Logistic Support Analysis Record (LSAR) is a data model described by ILS-Handbook which is not prescribed by the Dutch Defence any more. However some companies use this model because of Defence-contracts in the past.

**Database management system (DBMS)**

A DBMS is needed for data collecting and data management. A DBMS is a computerized software system designed for the purpose of creating and managing databases (Davis, 1985). Physically separated databases, each with another “owner”, in the past resulted in low quality data. Davis indicated already in 1985 that the solution of this problem is to arrange one database in which data has relation to other data. During utilization phase of the asset the database should be filled with reliable data to provide answers to questions mentiones in the *Conceptual framework*. Providing answers to these questions requires a formalized data information feedback subsystem with proper output.
A DBMS consists of a group of programs that are used in an interface between a database and the user, or between the database and the application program (Kirkwood, 2006). DBMSs are classified by the type of database model they support. A relational DBMS would follow the relational model, for example. The functions of a DBMS include data storage and retrieval, database modifications, data manipulation, and report generation. A data definition language (DDL) is a collection of instructions and commands used to define and describe data and data relationships in a particular database. File descriptions, area descriptions, record descriptions, and set descriptions are terms the DDL defines and uses. A data dictionary also is important to database management. This is a detailed description of the structure and intended content in the database. Not only is it possible to collect, store, retrieve and manage data, the DBMS provides also data users a view to data with help of a user-interface and special programmed software.

**Data management**

Data is not information, see Definitions. For data to become information, it requires structure. Good data management provides the structure necessary to transform data into information. Good data management systems are complex programs, but the data management solution should be as simple as possible. Data management is like asset management: wrong design results in high cost and low performance. In other words: cost effectiveness is declining (while data quality is not improving) and that is not the purpose of LCM.

Data processing and data management are critical components of business organizations. Master Data Management (MDM) enables control of master data. MDM involves using a centrally managed database of customer names, product numbers, and other critical data.

A good data management system cannot generate trends without data that was collected for multiple years. Good data management starts with good data. Good data is
data which is not corrupted. Data corruption can occur during collection, transfer, summary or storage. First should be identified what the data management system is to achieve and what output is required (Chalfant, 1998).

![Figure 13. Data identification process](image)

Similar with LCM-processes there is a need, in this case a need for information. The goal depends on the needed information. After identifying the goal we should determine which data is required and from which databases this data should be extracted.

The traditional approach to data management consists of maintaining separate data files for each application. For example, an employee file would be maintained for payroll purposes, while an additional employee file might be maintained for newsletter purposes. One or more data files are created for each application. However, duplicated files result in data redundancy. The problem with data redundancy is the possibility that updates are accomplished in one file but not in another, resulting in a lack of data integrity. Likewise, maintaining separate files is generally inefficient because the work of updating and managing the files is duplicated for each separate file that exists. To overcome potential problems with traditional data management, the database approach was developed (Kirkwood, 2006).

Data management practice can always be improved. Common mistakes include: (Chalfant, 1998)

1. No backup capability or the backup feature is not used. Computer systems rely on electricity to run. Without a perfect electrical supply (safety-preventions, redundancy etc.), fatal failures will occur. All data should be backed up on a regular schedule.
2. Using a spreadsheet. It is not a good idea to base your management systems on a spreadsheet.
3. Not enough room to grow. Your data management system should be designed to meet anticipated needs as well as current needs. A scaleable architecture may be necessary if significant growth is anticipated.
4. Too much data in one field. A good data management system should provide appropriate fields and data segregation. Without appropriate data segregation the system is doomed.

5. Poor data entry validation. A good data management system should include checks on data entry to keep the garbage out.

6. Use of proprietary standards. A good data management system should include the ability to import and export standard data formats (DBF, XLS, ASCII etc.).

**Data quantity**

For 25-30 years we have collected electronic data and the data quantity is constantly increasing.

The amount of data generated and maintained by many businesses doubles every 12 to 18 months (Whiting, 2006).

Collecting data is a relatively simple task and from origin we want more data than we really need.

In the early phases of system development available data may be limited, but quite often there is the tendency to generate too much data and too early, which can be counterproductive and very costly (Blanchard, 1998a).

However, too much data, and especially unused information out of that data, costs money (Dwornick, 1999). A striking example is Total Quality Management (TQM). In the 1990s, TQM was all the rage. Large and small companies embraced it. Unfortunately, many TQM administrators demanded that their employees fill out forms to document and justify every decision as part of the TQM process. It got to the point where employees were spending so much time filling out forms and preparing internal reports that the system collapsed on its own paper weight. People were collecting data for the sake of it without analyzing their findings to improve management practices. When senior management discovered such a waste of resources, the TQM program was shut down (Wei, 2003).

Software applications (for instance ERP) may generate great efficiency, but they also generate great floods of data. So great, in fact, that nowadays Chief Information Officers speak of petabytes of storage rather than mere terabytes (Goff, 2003).
Clearly, one of the consequences of such enormous amounts of data being available is the difficulty in finding important, relevant data and being able to compare similar data from different sources (Orr, 1998).

Sometimes too little data is collected, missing data parts or no data parts at all. This is a severe mistake because a good data management system cannot generate trends without data. Generally, years of results are needed for trends (Chalfant, 1998). Data sets with high-quality data but not complete can be analysed by means of statistical calculations.

There are not so many organizations which collect no data, probably all organizations collect data. Most organizations have too much data available, but often not the data we want (Embledsvag, 2003).

A previously well specified data set is very important. Extending your data set after development is a very costly business, so we should start with a good definition of the user need.

Since we have such a huge amount of available data, we did not pay enough attention to data quality.

**Data quality, general view**

After having secured the availability of data we must ensure to collect the right data i.e. data with high quality. High-quality data is extremely important for the success of a company. High-quality data is important because cost effectiveness of assets can not be controlled without reliable information.

We define high-quality data as data that is fit for use by data consumers, a widely adopted criterion (Strong et al, 1997/ Tayi/Ballou, 1998 and others).

The accuracy of data in the early phases of system development consists of rough estimations and will, in all probability, become more precise during lifetime of the asset (Stavenuiter, 2002).

There can be different reasons for gathered data, which doesn’t fit the purpose. Two of the main reasons are lack of motivation and complex, not user-friendly interface, caused by not gathering the right data in the right way (ref. Lammerse, 1993).
The complexity of an application is no assurance for success. On the contrary, just “keep it short and simple” (KISS-method) has generally spoken more chance for success (Jonsson, 1999).

Redman (1995) describes four indicators of poor data quality:

1. Extensively inspecting and correcting data;
2. Redundant data and processes to create them;
3. Lack of quality data for implementing strategy and/or reengineering;
4. Individual or organizational frustration with data
   (There is a natural human tendency to assume that “if it is in the computer, it must be right”).

Accurate data does not come free. It requires careful attention to the design of systems, constant monitoring of data collection, and aggressive actions to correct problems that generate or propagate inaccurate data (Olson, 2002b).

In a further publication Olson (2007) shows four general areas where inaccuracies occur, each of them with initial causes:

1. Initial data entry:
   a. Mistakes
   b. Data entry processes
   c. Deliberate
   d. System errors
2. Data decay
3. Moving and restructuring:
   a. Extracting
   b. Cleansing
   c. Transformation
   d. Loading
   e. Integration
4. Using:
   a. Fault reporting
   b. Lack of understanding
Chalfant (1998) predicts the cause of general error rate in data during:

- Data field collection: 20%
- Data transfer: 10%
- Data storage: 10%

Inmon (1999) indicates four ways in which dirty data enters the data warehouse:

1. Invalid or incorrect legacy applications that collect the data in an incorrect state. Changing data in legacy is time consuming, expensive and complicated.
2. Improper integration and transformation programs. These activities are many and complex (re-sequencing, conversion, standardization, structuring etc.).
3. Ageing of the data inside the data warehouse. Even with perfect legacy data ageing will destroy the purity of data in the data warehouse.
4. Change of user requirements. Data quality depends on pre-defined expectations of data-users.

Vosburg (2001) comments that misspellings, duplicate records, and inconsistencies were the result of a lack of control over who could add, change, or delete customer master data, of instructions for proper management of the data, and of auditing procedures.

It seems that from all data management activities data collection has relatively a great impact on data quality. Umar (2000) recommends entering any particular data item manually into a single system. Input should be validated using business rules. Data needs to be either rejected or marked as being invalid as soon as it is detected.

Tight control of the reporting system development process is required, and close attention must be paid to collect only what is needed, not all what is available. System developers need to resist the temptation to collect the easiest available data at the expense of data that may be harder to collect but is more valuable to decision-makers (DoE/PBM, 2001).

The collection and report frequency do not have to be the same. Some users may like to see monthly data only once a year, for example, while other users may want trend information such as organizational financial data more frequently.
The best data in any company is the record of how much money someone else owes the company. Data quality goes downhill after that. Frequently a data element would be very interesting if it was of high quality, but it either is not collected at all or it is optional. Optional data is the kiss of death for data quality! (Kimball, 1996).

Often there is a low priority assigned to the aspect of data quality, in contrast with for instance product quality. However Kay (1997) states that managing data quality is the most vexing data warehouse challenge. Wallace (1999) says that data quality improvement is the most important technology priority for companies, even more than high availability.

Data quality is a major problem and it seems that data in data sources are often “dirty”. Dirty data includes missing data, wrong data, and non-standard representations of the same data.

The problem of ensuring data quality is exacerbated by the multiplicity of potential problems with data. There are so many ways for data to be wrong.

Probably, legacy-data is invariably in worse condition than we realize (Atre, 1998). The problem is so worldwide, that there are even special educations, for instance the masterclass “Data Quality Management” Nyenrode University, or more technical courses like Informatica (2007).

The results of analysing a database/data warehouse of dirty data can be damaging and at best unreliable.

Missing or inconsistent data has been a pervasive problem in data analysis since the origin of data collection.

Poor data quality impacts the typical enterprise in many ways. At the operational level, poor data leads directly to customer dissatisfaction, increased cost, and lowered employee job satisfaction (Redman, 1998).

Bad data causes not only dissatisfaction and waste of money, resource, sometimes huge amounts, but on the other hand bad data can cause also loss of performance. Both options must be prevented.

Even worse, decisions based on faulty data can in the long term diminish confidence (and credibility) in the data (Nilsson, 1991). This can not only be very costly, but also fatal. When credibility is missing, quality of data is not important, because there is no confidence in the information build on that data.
If users believe that provided information is incorrect, they are quick to criticize the system (Oman, 1988).

Surprising detail as Hernandez (2006) says in improving data quality is that when an organization implements a solution that measures uptime, downtime, availability etc. productivity also will improve even if nothing else changes. This is only valid for reliable data or else there is no confidence or credibility in information. Operators, technicians, engineers and managers are being aware of bottlenecks and problems. Boelens (2002) also described this phenomenon of awareness and he mentioned this a “chicken-egg” discussion.

Responsibility for data quality should clearly be defined (Nilsson, 1991/ Umar, 2000). Often it seems that nobody is responsible for data quality (Blankena, 2005) and we are pointing at each other when there is a problem, beginning with ICT-departments. In fact there should be someone (or a group or team in large organizations) to manage and coordinate all data quality activities. Data users with knowledge of company processes should bring up the need and indicate the required data quality level. Some companies have established “Enterprise data architecture groups” that watch out for the health of corporate data throughout its life cycle.

Another idea is to treat data management as asset management. The managers are paid based on increase in the value of the asset (similar to other assets in organization) (Umar, 2000).

Davenport et al (2005), Greengard (1998), Redman (1995), Stair (2003) and others comment that ability to establish and maintain a competitive advantage is vital to the success of the company and that bad data can put a company on a competitive disadvantage.

The whole process of production of information can also be seen as an asset: we want minimum spending of costs and an optimal effect of our decisions. Redman (1995) emphasizes that data is valuable and must be treated as an asset. Clikeman (1999) considers reengineering of production processes and determines like Wang that production of information out of data applies with the same quality criteria.
Wang (1998) determines that organizations for increasing productivity must manage information as they manage products:

<table>
<thead>
<tr>
<th>Input</th>
<th>Raw materials</th>
<th>Raw data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Process</td>
<td>Assembly line</td>
<td>Information system</td>
</tr>
<tr>
<td>Output</td>
<td>Physical products</td>
<td>Information products</td>
</tr>
</tbody>
</table>

Table 1. Products vs. information manufacturing (Wang, 1998)

To treat information as a product, a company must follow four principles (Wang et al, 1998):

1. Understand consumers’ information needs;
2. Manage information as a product of a well-defined production process;
3. Manage information as a product with a life cycle;
4. Appoint an Information Product Manager (IPM) to manage the information processes and resulting product.

Tayi (1998) comments that often a low priority is assigned to data quality. In that case a Data Quality Manager (DQM) could be important for managers to make them realize the necessity of ensuring data quality.

Olson (2007) remarks that we need to be proactive for reaching the high-quality data goal. We need to focus on data quality and we need a dedicated, focused group. This means we need an organization that is dedicated to improve data quality. We need trained staff members who consider the skills required to achieve and maintain data quality as career-building skills. We need to educate other company members in the importance of data and what they can do to improve data quality.

Educations and courses should pay more attention to data quality for awakening of company members in relation to the importance and effects on LCM and ILS (Olson, 2002b).
Ultimately, the real difficulty with data quality is change. Data in our databases is static, but the real world keeps changing. Use your data, or lose it. Feedback is necessary to adjust planning and improve objected information (Deming-circle or PDCA-cycle)!

![Feedback information system](image)

Figure 14. Feedback information system (Orr, 1998)

Orr (1998) has formulated some data quality rules which all can be deduced from one point of view: data must be used or it will be worse (“use it or lose it”).

Atre (1998) distinguishes seven steps for cleansing the data:
1. Identify data of interest and the business uses of that data;
2. Analyze data for content, meaning and importance;
3. Determine which data to include in the data warehouse;
4. Write pseudocode procedures for data extraction, conversion and population;
5. Determine the suitability and effectiveness of automated cleansing and extraction tools;
6. Consider an iterative and ongoing data warehouse process;
7. Extract data, populate the data warehouse and work in parallel on your online analytic processing (OLAP) implementation.

Parsons (1993) mentions three methods to ensure good data and reduce waste in data collection:
1. Define information needs;
2. Define the data available;
3. Assess the accuracy of the data.
As catastrophic as the results of bad data can be, there is still hope. QDB-solutions, Inc. President Marc Hansen suggests a broad five-step, “zero defect” data approach for improving corporate data quality:

1. Identify the data to be improved;
2. Measure the quality of the data in the database;
3. Identify causes of poor data quality;
4. Improve systems to prevent data quality errors;
5. Measure your improvements.

Olson (2002b) mentions several reasons why not much has been done about quality problems:

- Low awareness of the costs of data quality problems;
- Low awareness of the potential value of improvements;
- Tolerance for errors is high in primary systems;
- Treatment of known issues as isolated problems instead of symptoms;
- Coping; acceptance of poor quality;
- Business case for improving quality is not as high as for alternatives;
- Scepticism over ability to improve things and get returns.

A lack of understanding about how to collect good data creates serious problems for many teams new to process improvement. Their data collection methods result in waste because their practices are inefficient and ineffective. The cost of being lost and the cost of being wrong can be named as the two sources of waste. The cost of being lost stems from a lack of direction. Without clear direction, teams can spend hours collecting and analysing data that is not needed. The cost of being wrong is driven by inappropriate or inaccurate data (Parsons, 1993).
**Data quality requirements**

In the former paragraph we have determined that the term “data quality” best can be defined as “fitness for use,” which implies the concept of data quality is relative. Often there is no information about the measure of unreliability of data, at most an estimate.

Generally, 100% good data is ideal, but not realistic (Orr, 1998/ Pedley, 2000). Where lies the threshold that we can say: “Data quality is good enough”?

Assessment possibilities are available (Pipino et al, 2002/ Parssian et al, 1999 etc.), eventually with help of control matrices (Pierce, 2004) and also auditing ETL-software (Ascential, Informatica, Ab Initio etc.).

There are many dimensions, characteristics or criteria on which data can be assessed in order to know if data meets with certain quality-requirements. The main quality-requirements consist of dimensions or characteristics.

Although all aspects of data/information quality are relevant, some of them are more important in specific situations than others.

If information quality seems to be a problem, data should be taken as an artefact, checked for quality attributes and these problems should first be fixed (Lillrank, 2003).


CIHI (2005) has defined five dimensions of data quality: accuracy, timeliness, comparability, usability and relevance. All of them are divided into certain characteristics and these are divided in criteria (see figure below).

Informatica (2007) identifies six characteristics for data quality: completeness, conformity, consistency, accuracy, duplication and integrity. Some of them are overlapping other characteristics. An important characteristic like duplication is only defined by Informatica.
Accuracy has the most characteristics (7) and criteria (26) on a total of 19 characteristics and 58 criteria. CIHI is working with ratings and produces assessment reports for data quality.

EEA (1997) has defined five dimensions of data quality: precision, completeness, representative characteristics, consistency and reproducibility. EEA is working with a (simple, not unified) data quality index (DQI = 1 to 5) for indication of data quality; for instance (1,3,2,4,1) indicating that precision is low, the completeness is medium etc.

Parssian et al (1999) researched a methodology to determine two data quality characteristics (accuracy and completeness) that are of critical importance to decision-makers.

Carson (2000) mentions five dimensions of data quality for assessment of statistical data: integrity, conceptual consistency, accuracy, serviceability and accessibility.

It is possible to characterize data with typical dimensions or characteristics; all of them are important, more or less.

The following dimensions or characteristics for data, information and software quality are found in literature:
## Table 2. Dimensions of high-quality data

<table>
<thead>
<tr>
<th>Dimension</th>
<th>Definition</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accessible</td>
<td>Data/information should be easily accessible by authorized users to be obtained in the right format and at the right time to meet their needs</td>
<td>2,5,7,8,9,10</td>
</tr>
<tr>
<td>Accurate</td>
<td>Accurate (meta)-data is free from error</td>
<td>1,2,4,5,6,7,8,9,10</td>
</tr>
<tr>
<td>Amount of data</td>
<td>The extent to which the quantity or volume of available data is appropriate</td>
<td>7,10</td>
</tr>
<tr>
<td>Believable</td>
<td>The extent to which data/information is accepted or regarded as true, real and credible</td>
<td>7,10</td>
</tr>
<tr>
<td>Complete</td>
<td>Complete data/information contains all of the important facts</td>
<td>1,2,4,5,7,8,9,10</td>
</tr>
<tr>
<td>Concise</td>
<td>The extent to which data is compactly represented without being overwhelming (i.e. brief in presentation, yet complete and to the point)</td>
<td>7,10</td>
</tr>
<tr>
<td>Consistent</td>
<td>The extent to which data elements are consistently defined, compatible with previous data, presented and understood</td>
<td>1,2,4,5,7,9,10</td>
</tr>
<tr>
<td>Duplicated</td>
<td>The extent to which data occurs in different forms in the dataset</td>
<td>5</td>
</tr>
<tr>
<td>Easy to understand</td>
<td>The extent to which data/information is clear without ambiguity and easily comprehended</td>
<td>6,7,10</td>
</tr>
<tr>
<td>Easy to operate</td>
<td>The extent to which data is easily managed and manipulated (i.e. updated, moved, aggregated, reproduced, customized)</td>
<td>6,7,10</td>
</tr>
<tr>
<td>Economical, cost effective</td>
<td>The extent to which the cost of collecting appropriate data is reasonable or information is relatively inexpensive to produce</td>
<td>2,3,6,7,8,10</td>
</tr>
<tr>
<td>Flexible</td>
<td>Flexible information can be used for a variety of purposes, not just one</td>
<td>6,8,10</td>
</tr>
<tr>
<td>Integer</td>
<td>Is the structure of the data and relationships among entities and attributes maintained consistently</td>
<td>5,9</td>
</tr>
<tr>
<td>Interpretable</td>
<td>The extent to which data/information is in appropriate language and units and the data definitions are clear</td>
<td>2,7,10</td>
</tr>
<tr>
<td>Objective</td>
<td>The extent to which data/information is unbiased (unprejudiced) and impartial</td>
<td>7,10</td>
</tr>
<tr>
<td>Reputation</td>
<td>The extent to which data/information is trusted or highly regarded in terms of their source or content</td>
<td>7,10</td>
</tr>
<tr>
<td>Relevant</td>
<td>Relevant information is important to the decision-maker</td>
<td>3,7,8,10</td>
</tr>
<tr>
<td>Reliable</td>
<td>Reliable data is dependable, trusted, audited data</td>
<td>2,3,4,6,7,8</td>
</tr>
<tr>
<td>Secure</td>
<td>Information should be secure from access by unauthorized users</td>
<td>6,7,8,10</td>
</tr>
<tr>
<td>Simple</td>
<td>Data/information should be simple to find and understand</td>
<td>8</td>
</tr>
<tr>
<td>Timely</td>
<td>Timely data/information is readily available when needed</td>
<td>1,2,3,4,7,8,9,10</td>
</tr>
<tr>
<td>Traceable</td>
<td>The extent to which data is well documented, verifiable and easily attributed to a source</td>
<td>10</td>
</tr>
<tr>
<td>Valid</td>
<td>Data values fall within acceptable ranges defined by the business</td>
<td>9</td>
</tr>
<tr>
<td>Value-added</td>
<td>The extent to which data is beneficial and provide advantages from their use</td>
<td>2,3,7,10</td>
</tr>
<tr>
<td>Verifiable</td>
<td>Verifiable data/information can be checked to make sure it is accurate</td>
<td>8</td>
</tr>
</tbody>
</table>

References:
1. Ballou/ Pazer, 1985
2. Bose, 2006
3. Clikeman, 1999
4. Dep. of Energy/USA
5. Informatica, 2007
6. ISO 9126
7. Pipino et al, 2002
9. Trembly, 2002 (from Data Warehousing Institute, Seattle)
Most commonly used are accessibility, accuracy, completeness, consistency, economy, reliability and timeliness. These seven dimensions should primarily be taken into account with data collection and development of data warehousing.

Almost none of the companies are working with a DQI which indicates data quality. In relatively simple data quality measurements an indicator like “percentage correct” could be used (Oman, 1988).

When more variables are present, for instance the average of given, same-weight dimensions could be an indication of data quality. However, the weight of dimensions will differ. This subject is important, but falling outside the framework of this research.

Advantage of having an indicator or index is that trends can be presented and the possibility exists for planning of data quality improvement.

**Data collection**

Generally, there are three types of data management systems for data collection: cerebral, manual and electronic (Chalfant, 1998).

1. Cerebral is in the head. Cerebral systems are generally unreliable. The data can easily get lost, become corrupted and/or not be collected at all. Not done.

2. Manual systems have been the norm for hundreds of years. The advantages of a manual system include low cost, user familiarity and ease of implementation. However, manual systems are subject to document loss, document misfiling and limited search capabilities.

3. Electronic data management is an emerging technology and now the standard method. Just about everything in companies is “data based”. Internet and web portal technology are important developments (Caro et al, 2006/ Heina, 2005).

The method of data collection for asset analysis can differ. It can be an automatic data collection or a manual data collection. Continuous automatic data collection is a powerful method to increase the equipment's utilization. However, an automatic data collection system is expensive, complex and the data is collected at an aggregated level. A manual data collection can be very detailed and failures can be carefully examined (Ljungberg, 1998), but practice has shown that this leads to time-consuming analyses (Fermont, 2002).
Master data for reliability requirements like failure rates or MTBF can be obtained from literature search, membership of a database organization or review of existing risk and reliability studies. Data books with asset reliability data are for instance: (Schüller et al, 1997)

1. NUCLARR
2. IEEE Std 500-1984
3. Eireda
4. T-book
5. OREDA
6. RDF 93
7. MIL-HDBK 217F

Data quality will be higher when more of the requirements in table 3 are fulfilled.

<table>
<thead>
<tr>
<th>No.</th>
<th>Requirement</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Component type</td>
</tr>
<tr>
<td>2</td>
<td>Clear description of the failure mode</td>
</tr>
<tr>
<td>3</td>
<td>Description of the component boundary</td>
</tr>
<tr>
<td>4</td>
<td>Mean value</td>
</tr>
<tr>
<td>5</td>
<td>Median value</td>
</tr>
<tr>
<td>6</td>
<td>Uncertainty bound</td>
</tr>
<tr>
<td>7</td>
<td>Description of the component population</td>
</tr>
</tbody>
</table>

Table 3. Requirements for reliability databases

An important source for providing asset reliability data is expert opinion. Several methods are available:

1. The Delphi method: undoubtedly the best known method
2. The Nominal Group Technique
3. The Consensus Group Technique
4. Mean value, with or without weighing
5. Bayes technique

The methods mentioned above are just a first step to produce maintenance plans. Failure and cost data have to be collected from the asset to produce and analyse cost effectiveness trends. Collecting asset specific data is a time-consuming effort.
Transactional performance (reliability) data is actual, less generic data and should contain the following information (Lammerse, 1993):

- System description and relevant components with function, location, dimensions, configuration etc.
- Operational data like workload, maintenance philosophy, environmental conditions etc.
- FMECA information like failure mode, failure effect etc.
- Reliability of available data.

We distinguish two sorts of analysis-systems (Blanchard, 1998):

1. Analysis of *ongoing* data to evaluate and assess cost effectiveness. This information must quickly be present at frequent times of the asset life cycle.
2. Analysis of *historical* data for development of new designed (or to be modified) assets having a similar function and nature.

Elements or factors for data collection systems are described by Blanchard (1998). These factors should be:

1. simple to understand and complete;
2. clear and concise;
3. meaningful and useful.

According to Ishikawa (1982), the reason for collecting data should not be to present neat, coloured, sophisticated figures, but to create a base for action and development of processes. This is very much linked to what data is collected, how the analysis is carried out and how the performance information is used (Jonsson, 1999).

The data collection should be at such detailed level that it fulfils its objectives without being unnecessarily demand of resources. A too detailed data collection may result in unmotivated personnel and reaction against the measurement. Sometimes the process in itself is so complex that it is impossible to avoid a detailed data collection. The data collection should be carried out by personnel that can affect the measured parameters. Nearness is an important aspect in continuous improvement and therefore the result of data collection should not only be summarized to a key figure as a part of the measurement system, it should also be used as input in small group
Optimizing data collection for cost effective control of assets

activities (Jonsson, 1999). This information should be well communicated between involved corporate members.

Data collection takes mostly place with maintenance software. Several software packages are available, for instance Datastream, Maximo, Maintcontrol, Optimizer, Ultimo and so on. Integration of data, ERP (SAP) and development of web-portals are coming up (Heina, 2005).

Have (2007) indicates some resistance against use of maintenance software packages. He postulates that creativity and motivation of technicians is decreasing and that the success of most sophisticated systems depends on data-input. He is pointing at the importance of communication. Motivation can be increased by involvement at choices of data collection procedures and tools.

In many companies resistance exists against data collection from operators and foremen. To succeed with data collection, it is necessary to find a less time-consuming method that is also precise, but also conviction and motivation for the data collection method are important (Ljungberg, 1998).

Heina (2005) comments that web-portals in combination with a quick and flexible data communication are a pre-condition for optimising maintenance.

Management has an important role in data collection procedures (Knight, 2007):

1. Management must ensure that pre-assessment of data is thorough. Because management must recognize that the origins of potential problems come from different places and the omission of any significant causes and effects will obviously impact the analysis.

2. Furthermore, management should understand that good statistical analysis is expensive in both time and actual costs. Spending excessive time and money on the collection of poor or inappropriate data is a fruitless and sometimes even deceiving problem sometimes seen in numerical analysis.

3. Finally, management must emphasize that without their prior agreement on appropriate operational variable definition and suitable measurement techniques and methodologies, conclusions may be dismissed if they are contrary to any preconceived plans or agendas.
The following steps should be part of your data collection strategy (Revelle, 2006):

- Determine the purpose of the data to be collected (is related to the need of data).
- Determine the nature of the data to be collected. Are they measurable (variable or continuous) or are they counted (attribute or discrete)?
- Determine the characteristics of the data to be collected. Can the data be easily understood by people who will evaluate product and process improvement, including customers?
- Determine whether the data can be expressed in terms that invite comparisons with similar processes.
- Determine whether the data places priority on the most important quality influences and whether the data is economical and easy to collect.
- Determine the best type of data gathering check sheet to use: checklists, tally sheets or defect concentration diagrams.
- Determine whether it will be possible to use random sampling or necessary to use 100% data collection.

If we collect data we must be sure that data export to the data warehouse for producing information is possible in an easy way. Information is necessary to manage cost effectiveness information, i.e. performance and costs.

The question is whether the database-architecture during development is originally focused on cost effectiveness or has another purpose. Often the original database and information system is present and the company is trying to use it for the purpose of LCM and cost effectiveness. Data collecting might result in wrong or less efficient data. This is the beginning of a lot of data cleansing and improvisation, but ends in window-dressing to mask reality.

**Enterprise Resource Planning (ERP)-system**

An ERP-system (e.g. SAP for Dutch Defence) include all single parts of the company-process, for instance financials, material & inventory management (warehouse), planning, ordering and plant maintenance. ERP integrates all the data and processes of the company into one single system.
Before the implementation of ERP-systems there might be many single, independent databases within the company. The main goal of ERP-systems is to integrate data and processes from all areas of an organization or company into a unified database system and unify it for easy access and work flow.

So, in this ERP-system the data is integrated and has a relation with other data in all the company processes and during the whole life and process cycle.

A typical ERP system will use multiple components of computer software and hardware to achieve the integration. A key ingredient of most ERP systems is the use of a unified database to store data for the various system modules. Typically, an ERP system uses or is integrated in a relational database system.

Benefits of ERP-systems are consistent processes across all units and gives much better visibility across the company.

ERP is acquired for reliable planning and administration in the company, but the necessary information for managers, engineers etc. is probably not standard present in the ERP-system (Perik, 2007).

Although the ideal configuration would be one ERP system for an entire organization, many larger organizations usually create an ERP system and then build upon the system an external interface for other stand alone systems which might be more powerful and perform better in fulfilling the organizations needs. Usually this type of configuration can be time consuming and does require lots of labour hours.

Besides the financial waste as result of dirty data (Vosburg, 2001), the cost of implementing, using and maintaining an ERP-system are relatively very high (Nah,2002/Calbo,2005).

Implementing an ERP system is not an easy task to achieve, in fact it takes lots of planning, consulting and in most cases 3 months to 1 year or more. ERP systems are extraordinary wide in scope and for many larger organizations can be extremely complex with a high risk and failure rate.

Implementing an ERP system will ultimately require significant changes on staff and work practices. While it may seem reasonable for an in house IT staff to head the project, it is widely advised that ERP implementation consultants be used, due to the fact that consultants are usually more cost effective and are specifically trained in implementing these types of systems.

Before ERP-implementation, the most important factor is that it requires good quality data, and that the (especially master) data needs to be cleansed. If we don’t do that, we
implement a big, expensive IT-project which does not work. Cleansing the data is the hardest part of the process and often this is underestimated (Vosburg, 2001).

There are several ETL-tools available with which we can extract the data from the source, transform them to the required quality and load them into the new system. This special software for data-profiling, data-auditing and data-cleansing is sold by vendors like Ascential, Informatica, Ab Initio (Goff, 2003), but also Trillium, First Logic, Datanomic (Whittle, 2004).

**Advantages of ERP Systems**

There are many advantages of implementing an EPR system; here are a few of them:

- A fully integrated system;
- Higher quality data;
- The ability to optimise different processes and workflows;
- The ability to share data across various departments in an organization;
- Improved efficiency and productivity levels;
- Better tracking and forecasting;
- Lower operational cost;
- Improved customer service.

**Disadvantages of ERP Systems**

While advantages usually outweigh disadvantages for most organizations implementing an ERP system, here are some of the most common obstacles experienced:

Usually many obstacles can be prevented if adequate investment is made and adequate training is involved, however, success depends on skills and the experience of the workforce to quickly adapt to the new system.

- Customisation in many situations is limited, the ERP-process is fixed;
- The need to reengineer business processes;
- Complex structure;
- Expensive and cost prohibitive to install, maintain and run;
- Technical support can be shoddy;
ERP may be too rigid for specific organizations that are either new or want to move in a new direction in the near future.

Data warehousing

Normally it is not enough to receive raw data and take decisions. Data has to be processed, analysed and presented into information. All of these activities, from collecting data to producing information can be done with help of data warehousing.

In a data warehouse data is assembled and analysed and data can only be read by users (Harjinder, 1996/ Inmon, 1997/ Koenders, 1999 etc.). Operational systems are available like OLTP (OnLine Transactional Processing), where data can be read and written by users. We describe data warehousing where data only can be read by users for the scope of this research.

Data warehousing is necessary to manage company data in a better way and it has become the decision support trend (Ma, 2003).

Data warehousing involves taking data from a main computer for analysis without slowing down the main computer. In this manner, data is stored in another database, integrated and analysed for trends and new relationships.

Consequently, the data warehouse is not the live, active system, but it is updated daily or weekly (Kirkwood, 2006). Data in a data warehouse is not being updated, but is saved and new data is added.

A company must have systems that can integrate this data and provide the analysis to support decisions. Without such a strategy, companies are left scrambling for timely and reliable data and may be left behind in the race for profits (Pedley, 2000).

Data from different operational databases is usually directed to a data warehouse where data is assembled and processed to management information. Historic data will also be in the data warehouse for (trend)-analysis.

A data warehouse is a subject-oriented, integrated, time-variant and non-volatile collection of data in support of management’s decision-making process. Infrastructure for data warehousing has to be build only once and makes data available and cheap, see figure below.
The data warehousing architecture commonly used is called a multi-tier warehouse, in which an enterprise warehouse coexists with several data marts:

In the long run, a multi-tier warehouse is the best architecture. It provides a single source of clean, integrated data, as well as local stores tailored to the needs of specific
groups. Because it is more difficult to build and manage, many organizations begin in the short term with isolated enterprise warehouses or data marts. After data warehousing data-users do have the availability of “business intelligence tools” for analysis (for instance OLAP), reporting and data-mining.

After data extracting and data cleansing out of the data source (DBMS) the data is conducted into the data warehouse. This model is widely used:

![Data management process (Stair, 2006) diagram]

The data warehouse stores historical (and sometimes current) business data in an organized manner, allowing specialized queries and data retrievals to be easily performed. The promise of data warehousing is to “get the data out” of the operational systems to help business make better decisions. The challenge is to “get the data out” of the data warehouse and convert it to information that helps businesses make more informed choices which will yield better decisions and create a sustainable business advantage (Harjinder, 1996).
Approaching data warehouses with the same information and (mis)management principles that have produced the disintegrated islands of legacy-data will result in failure. It will fail spectacularly. In fact, it will deserve to fail!

Data warehousing projects fail for many reasons, all of which can be traced to a single cause: *non-quality*. Poor data architecture, inconsistently defined departmental data, inability to relate data from different data sources, missing and inaccurate data values, inconsistent use of data fields, unacceptable query performance (timeliness of information), lack of business sponsor (no data warehouse customer) and so forth, are all components of *non-quality* (English, 1999).

**Cost of data collection**

How good can or must data be? How much costs data collection and data cleansing so that we make reasonable decisions and is support of the asset cost effective? This question is very important, because all the activities must not result in a deficit, but are planned in a positive economic perspective.

A wide area of resources is available to provide good quality data and it costs a lot of money. Bad data not only costs the company (much) more money than needed (Redman, 1995), but on the long term LCM will not be possible or hard to implement.

Management decisions are only as good as the data on which they are based and organizational misdirection due to faulty information can be costly (Clikeman, 1999). Most of the cost associated with dirty data cannot be measured in terms of euros. If this cost could be quantified the management would be shocked (Vosburg, 2001).

Obviously, our goal is cost effectiveness, so the investment in data management must be worthwhile on life cycle basis.

Cost for implementing DBMS, data warehousing, hardware/software, database creation is once only and cost for licenses, control, support (also consultancy), verification, assessment, audit, conversion, personnel, training and operation are periodic.

The cost for implementing ERP within the Dutch Defence can rise to an unacceptable level, because it is a high-risk project. Because data quality depends on successful implementation, it is extremely important to prevent failure.
Getting into data warehousing may be costly on the set-up and expensive and complex to maintain (Chervinsky, 1997). The collection of data often takes an extensive amount of time and effort resulting in data collection cost having a significant impact on the analysis. In some cases, an existing database (legacy-data) can be referenced and analysed for a relatively minor cost. However, in most cases the existing database is only tangentially related to the specific problem at hand and will thus only provide tangentially related answers to the problem (Knight, 2007/1999).

Internal auditors can help their organizations achieve significant cost savings by evaluating the efficiency of their companies' information systems and making sure the systems are being managed economically (Clikeman, 1999). This means that activities must be concentrated rather on prevention than on detection and/or cleansing. Effective controls should be built into information systems, not added onto them later.

Most often, the cost associated with data collection, storage and retrieval are the most expensive aspects of a performance measurement program.

As with many other management decisions, cost plays a central role in the evaluation planning process. Data collection techniques vary widely with respect to cost. Program managers should balance the needs of the evaluation with the financial resources available for the evaluation. Some low-cost data collection techniques limit the type of information that can be collected or limit the data quality (DoE, 2001).

The more sophisticated the data collection and reporting system, the more expensive it will be to implement. Building cost effective data collection systems refers to many required attributes like experience, knowledge and simplicity (Chalfant, 1998)! Improved timeliness, depth of understanding, breadth of coverage, and ease-of-use come at a price.

Tight control of the reporting system development process is required, and close attention must be paid to collecting only what is needed, in stead of everything what is available. System developers need to resist the temptation to collect the easiest available data at the expense of data that may be harder to collect but is more valuable to decision-makers.

The relation between data quality and cost is comparable with cost for product quality. This relation will be certainly not linear (“100 % good” or “excellent” will never be
reached), but more exponential. Juran (1974) describes the relation of quality cost as asymptotically. At the other hand the asset cost (life cycle cost) will decrease when our decisions are better, i.e. when our data has more quality. There will be an optimum. Figure below is an example of the above mentioned cost factors (scaling is brought together):

![Data Quality Cost](image)

**Figure 19. Data quality cost**
Overview of the theoretical research

We have seen the methods of data collection, the different ways in which data can become dirty and the characteristics with which data quality can be determined.

A good data management system is well aware of the final goal and data need to reach this goal. The solution should be as simple as possible, not complex, but comprehensible, simple to understand and complete, clear and meaningful.

Data quantity is often too much and that can be counterproductive and very costly.

Data quality is often too low and data is more dirty than we realise, but data quality does not need to be 100% good. It should be as cost effective as can be.

High quality data is data which is fit for use. Two of the main reasons why data doesn’t fit the purpose are lack of motivation and complex, not user-friendly interface, caused by not gathering the right data in the right way.

Data can be assessed on many dimensions, characteristics or criteria in order to know if data meets with certain quality-requirements. Most commonly used are accessibility, accuracy, completeness, consistency, economy, reliability and timeliness. These seven dimensions should primarily be taken into account with data collection.

Data quality is a major problem and the assigned priority is too low. We should manage data as we manage products and be well aware of the problem by means of more responsibility and education. Especially courses in the field of “Asset -, Data –, Maintenance –, and Information Management” should be more focused on this subject.

We have seen an integrated database (ERP) as possible solution for the data quality problem, because ERP-systems require high data quality, they are efficient and data is integrated.

But there is also another risk: ERP-systems collect great floods of data, are complex, expensive and rigid. ERP is not an assurance for cost effective control of assets. The data can be cleansed with special provisions before importing them in the datawarehouse, but collecting data is expensive in relation to other data collection systems and data auditing or cleansing is necessary during life cycle.
4. Practical research

Preparations

The chosen research method has been interviewing persons by means of previous formulated questions via Internet. These persons are responsible for, involved with or well-informed about collection and registration of the data for producing management information.

In an early stage of the research it became clear that management information for LCM and cost effectiveness of an asset, defined and produced by RNLN/ME, probably nowhere else would be settled or found.

In advance of the assumption that there are not so many companies familiar with LCM and ILS the research questions are formulated in such a way that it handles management information in a general aspect, see Appendix A.

The questions are chosen sequentially, from the moments of data collection and information production including all possible effects on data quality and cost effectiveness. The number of the questions was totally sixteen (16).

The first two questions are to find out which company (1) and which function (2) the interviewee is representing.

Question 3-5 are intended to find out how (3), where (4) and who (5) the data is collecting.

Question 6 and 7 are to find out the subjective score of interviewees about quantity (6) and quality (7) of the collected data.

The (at first sight) most important, mentioned, recognizable dimensions from the pre-defined table “Dimensions of high-quality data” were the ingredients for question 8.
In this question is asked to give an indication (on scale 1 to 5; 1=not satisfying and 5=satisfying) for the following quality dimensions:

<table>
<thead>
<tr>
<th>Question number</th>
<th>Quality dimension</th>
</tr>
</thead>
<tbody>
<tr>
<td>8.1</td>
<td>Accuracy</td>
</tr>
<tr>
<td>8.2</td>
<td>Completeness</td>
</tr>
<tr>
<td>8.3</td>
<td>Availability</td>
</tr>
<tr>
<td>8.4</td>
<td>Reliability</td>
</tr>
<tr>
<td>8.5</td>
<td>Relevancy</td>
</tr>
<tr>
<td>8.6</td>
<td>Simplicity</td>
</tr>
<tr>
<td>8.7</td>
<td>Timeliness</td>
</tr>
<tr>
<td>8.8</td>
<td>Verifiability</td>
</tr>
<tr>
<td>8.9</td>
<td>Flexibility</td>
</tr>
<tr>
<td>8.10</td>
<td>Cost effectiveness</td>
</tr>
<tr>
<td>8.11</td>
<td>User friendliness</td>
</tr>
<tr>
<td>8.12</td>
<td>Credibility</td>
</tr>
</tbody>
</table>

Table 4. Quality dimensions question 8

Rather important was question (9) where is asked to indicate which factors are negatively influencing the collection of data.

Before producing information out of data some companies are cleansing their data. Question 10 is intended to find out how many companies follow that strategy.

Question 11-13 are intended to find out how information is produced (11), the satisfaction with this product (12) and what is doing with the results (13).

Question (14) is to find out the opinion about cost effectiveness of the whole data collection procedure, a very important part of the research information.

Finally, there were two questions with more textual answers. In question 15 was asked to indicate the possibilities for corrective actions and the last question (16) was trying to find out the impediments to perform these corrective actions.

There has been chosen to perform the research via Internet (Surveyworld.net). The research has been preceded by a test on Surveyworld.net with 7 persons. As result of the test the questions have been adjusted at some details.

A list of companies has been selected. The selected companies perform maintenance on assets and have a relation to sea, air and land systems, but also industrial assets.
The chosen number of companies has been based on the expectation that no more than 30-40% of the requested persons should respond on the research-request. While a number of 12 serious respondents from different companies as sample was set as a minimum for statistical analysis and reliable information, is chosen for a number of 30 companies. The selected companies have been approached by means of e-mail and/or telephone in order to raise the motivation of best possible participants for performing the objected research. The participants were promised to deal with the results of the research, but even this prospect did not bring a high participation result.

**Performance**

It became clear that collection of reliable and relevant data, not only for cost effective control of assets but also in this area, was very difficult and did cost a great deal of time. During the survey the results did not come as soon as expected. Surveyworld has the opportunity to collect data “anonymous” after the start of the survey. Not total anonymous, because company and function of the respondent were part of the questions and the possible respondent was familiar of course. The anonymous opportunity created another chance but also another problem, because it worked with “dynamic links” and that was for the most people an unknown phenomenon. As result of this, a certain number of possible responses have been lost. Anyway, in the first survey only 10 persons (in which 5 anonymous) from 8 different companies did react and this result was not satisfying.

So, because the first survey delivered too few data in the sample and could not represent the population as well, a second survey has been started. In the second survey also 10 persons (in which only 1 anonymous) from 8 different companies did react. Summarizing: 20 persons from 16 different companies were responding and that will satisfy. All the interviewed persons do have a relationship with data collection.
Alphabetical overview of interviewed companies and persons:

<table>
<thead>
<tr>
<th>Company</th>
<th>Function of interviewee</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. DMO (Operational management)</td>
<td>Senior employee</td>
</tr>
<tr>
<td>2. DMO/Air Systems (de Kooij)</td>
<td>Chief production</td>
</tr>
<tr>
<td>3. DMO/Land Systems (Dongen)</td>
<td>Project manager</td>
</tr>
<tr>
<td>4. DMO/Sea Systems/RNLN/ME (2 responses)</td>
<td>Data Manager; Systems Manager</td>
</tr>
<tr>
<td>5. ECN Petten</td>
<td>Chief dep. Quality, Safety &amp; Environment</td>
</tr>
<tr>
<td>6. Gasunie</td>
<td>Chief dep. Maintenance &amp; Data Centre</td>
</tr>
<tr>
<td>7. Gemini-Ziekenhuis Den Helder</td>
<td>Chief dep. Equipment</td>
</tr>
<tr>
<td>8. Havenbedrijf Rotterdam</td>
<td>Maintenance engineer &amp; project manager</td>
</tr>
<tr>
<td>9. NedTrain</td>
<td>Chief company development</td>
</tr>
<tr>
<td>10. NRG Petten</td>
<td>Instructor</td>
</tr>
<tr>
<td>11. Prov. Gelderland</td>
<td>Policy advisor &amp; program manager infrastructure</td>
</tr>
<tr>
<td>12. Rijkswaterstaat (2 responses)</td>
<td>Senior advisor/ specialist Maintenance co-ordinator</td>
</tr>
<tr>
<td>13. Schelde Shipbuilding</td>
<td>Senior engineer ILS-preparation</td>
</tr>
<tr>
<td>14. Shell</td>
<td>Rotating equipment engineer</td>
</tr>
<tr>
<td>15. TESO</td>
<td>Chief dep. Maintenance</td>
</tr>
<tr>
<td>16. Vitens (3 responses)</td>
<td>Specialist Asset Management; Manager Asset Management; Advisor Asset Management</td>
</tr>
</tbody>
</table>

Table 5. Overview of responses

Results

The survey results are represented in Appendix B.
The used methods to collect data are primarily manually (38%) and by means of maintenance software (28%).
Not integrated database management systems (isolated and flat) have been used in 60% of the cases of the researched companies. Integration of data is important for improvement of data quality.
In 62% of the cases collection of data is performed by user/operator or engineer. In the other cases (38%) is primary responsibility for data quality not incorporated in the right persons.
Similar to the general data quality situation the research showed the following results concerning data quantity and quality:
More than half the respondents have the opinion to collect too little data of a quality which has been judged as between moderate and reasonable. Nobody is judging data quality as good or excellent. Significant less than average are criteria like user-friendliness, cost effectiveness, flexibility, completeness and reliability of data.

In the figure below the average data quality is displayed for all respondents, sorted downwards and calculated from the mean qualifications for the seven dimensions mentioned in chapter data quality requirements (accessibility, accuracy, completeness, consistency, economy, reliability and timeliness).

![Average Data Quality](image)

**Figure 20. Average data quality per respondent**

Practically all findings are less than “average quality” or “reasonable”.

Also in line with the judgement of data quality satisfaction with produced information is not very good. 40% of the respondents are not satisfied and 45% think the quality could be better. Only 15% is satisfied. The result of data cleansing is hardly noticeable in this result.

In a substantial number of cases the data is collected manually (38%). Maintenance software is used in 28% of the cases.
Integrated databases are not much in use yet. Much more than half of the respondents (65%) are using not integrated databases (isolated and flat), but they think this can be one important step to solution of the problem.

All respondents think that there are one or more factors which influence data quality negatively in their company. Overall there are many causes (see question 9, appendix B). Most mentioned is factor “time” (19%). There simply is no time or opportunity to take corrective actions apart from the aspect if there are possibilities to improve. Related to “time” is “money”: With a low level of available resources the priorities of management can induce a negative impact for data collection procedures. Data collection procedures which are not properly settled, do contribute for 14 %, just like motivation.

The most important causes mentioned are:

<table>
<thead>
<tr>
<th>Nr.</th>
<th>Cause</th>
<th>Percentage (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Time</td>
<td>19</td>
</tr>
<tr>
<td>2</td>
<td>Ambiguous criteria</td>
<td>14</td>
</tr>
<tr>
<td>3</td>
<td>Motivation</td>
<td>14</td>
</tr>
<tr>
<td>4</td>
<td>Dysfunctional ICT</td>
<td>12</td>
</tr>
<tr>
<td>5</td>
<td>Knowledge</td>
<td>10</td>
</tr>
<tr>
<td>6</td>
<td>Money</td>
<td>9</td>
</tr>
<tr>
<td>7</td>
<td>Skill</td>
<td>9</td>
</tr>
<tr>
<td>8</td>
<td>Company conditions</td>
<td>7</td>
</tr>
</tbody>
</table>

Table 6. Overview of possible causes of poor quality data

Question (16) about impediments or corrective actions to solutions makes clear that complexity and investment (lack of resources) are two important causes of not performing the intended possible solutions.

Sixty percent (60%) of the respondents think that their data collection procedures are not cost effective during life cycle. This means that data collection costs more money than it delivers. It may be possible that data collection is performed with another purpose than reaching cost effective management. In 9% of the cases practically nothing was done with the collected data and in 27% the purpose was to produce reports, graphics and sheets. Our purpose should be “cost effectiveness” while more than half of the respondents think that this purpose is not achieved.
Overview of the practical research

An elaborate research at the field of data quality on a broad scale has been performed within companies which are comparable with RNLN/ME, i.e. supporting and maintaining capital assets.

Information has been gathered with this research about the ways and possibilities of collecting data.

All companies do need management information, the same or different did not matter for the research. Expectation was that there were no companies which are working like RNLN/ME with LCM and the research has been generalized at this point.

The research has indicated that data quality has an enormous impact on decision making and cost effectiveness of assets. While in practice management often is focused on decision making, primary attention should be paid in the production of good quality data. This is the starting point and the fundament of the production of reliable information. Many companies struggle with this problem, because this idea is initially not embedded in data management routine.

The data quality in the investigated companies is rather low. The average quality-index for the companies of all characteristics is not better than qualification “reasonable”.

Data cleansing is not a good option. When data quality is not good, the result of data cleansing is hardly noticeable. The challenge is to collect good quality data directly at the start.

Methods to measure and collect data are not unified. The used methods are not satisfying, because nobody qualifies data quality better than “reasonable”. This corresponds with theoretical research.

The data problem is exacerbated by using isolated and/or flat databases and giving persons responsibility for data collection who are not the most well-informed.

Integrated databases are not much in use yet. However, the general idea is that integration of data is one of the possible solutions of the data quality problem.

The data collection methods are, generally spoken, not appropriate and cost effective. There are several reasons why these methods are not appropriate and cost effective. Mostly it can be reduced to a lack of (combinations of) priority, time, resources, ambiguous criteria, knowledge and motivation.

Complexity of the problem and investment (lack of resources) are two important causes of not performing the intended possible solutions.
5. **Characteristics of a cost effective data collection**

For cost effective control of assets we must collect data. Data collection for asset management is certainly useful, but we should collect *only what is needed*, instead of everything what is available.

Data collection can be used for all sorts of analysis to produce the information we want. We require during life cycle recent, ongoing data at regular moments to get informed about trends for performance (failures, availability, reliability) and costs (materials, hours). Cost effective control of assets requires primarily the data collection of expensive, failure-sensitive and mission-critical parts. We must ensure that the data quality of these parts is good.

The data collection should be as simple as possible, preferably not sophisticated; the chance on high expenses which debit cost effectiveness is high.

The data collection system should be an integrated system. Alternative methods of data collection (cerebral, manual) are out of the question.

ERP-systems are not the most cost effective systems, they are complex and expensive, but all data is integrated and a preferable choice in relation to old, separated legacy systems.

Expert opinion is always an advisable addition to the data collection and information production, but especially when data quantity is less than planned, expert opinion should be part of our data collection to eliminate statistical uncertainty.

We need clean data which complies with a number of characteristics (accuracy, completeness, consistency, economy, reliability and timeliness). We should measure data quality and plan for improvement. It is ideal, but not realistic, nor needed, to demand that data quality is perfect. It should however be as cost effective as can be. It is useless, certainly not cost effective, to extract bad data out of the DBMS into the datawarehouse for the production of information. This information will not be reliable and will not contribute to cost effective control of assets.

Clean data means data which is assessed as good enough to analyse. Important is to collect clean data immediately at the source while making data verification and eventually cleansing not needed: it saves a lot of resources and elevates data credibility and data reliability. With more data credibility there will be more confidence in the
produced information and this upwards movement will even result in further improved data quality.

In practice data quality is often not as good as we want. Two of the main causes of bad data quality is lack of responsibility and motivation. Good notice should be taken if the person who collects the data has the appropriate skills, understands the system and the purpose for which the data is collected. By preference someone with relative more knowledge of the data and support of the asset like the user/operator or engineer, should be the “collector” of that data. Someone, by preference the engineer, should clearly be the “owner” of the data, i.e. a person who coordinates and manages data quality activities. Higher management must make this person aware of his/her responsibilities and making clear to handle data as they handle products or assets.

A good communication system should be settled between data collector(s) and data owner(s).

Care should also be taken to return information to those who collect data: motivation is rising and data quality also.
6. Summary

Chapter 2 gives an overview of the research methodology. In this chapter is indicated the way of performance of the research. To visualize the problem a conceptual framework was established. The problem has been explained, the way RNLN/ME is applying LCM and needs, but misses, appropriate data.

An overview of the data collection problem illustrated with a number of examples worldwide. From a description of the organization of RNLN/ME and data collection procedure the problem has been analyzed. There is an overview of reports indicating the bad data quality situation within RNLN/ME.

On the ERP-system is also focused which will be introduced in the RNLN the coming years with the chance for high data quality.

The expectation was that practically all companies collect data to use it for producing management information. The sort of information or KPI does not matter, because the way of data collection is similar.

Chapter 3 gives an overview of the theoretical review. All aspects of data are being reviewed. The conception of data quality is extensively researched. Analysed are the causes, problems, impossibilities and effects of bad data quality.

In this chapter is focused on the ERP-system and data warehouse where data has to be processed to produce necessary information.

Cost plays a significant role in data collection and data processing methods. We focused on the cost sorts and why data collection might be less cost effective.

Chapter 4 gives an overview of the practical research. In this chapter is indicated which preparations have been made to perform the research. Appendix A gives an overview of the research questions. After that a description has been given of the performance of the practical research with all encountered problems.

At last the results of the research have been analysed, mostly worked out in graphics (Appendix B).

Chapter 5 gives an overview of characteristics to approach a cost effective data collection.
7. Conclusions

In reference to the sub questions (ch.1) below the summarized answers as result of the research. For detailed description see theoretical and practical research:

a. The research has pointed out that there are no companies which are familiar with ILS and LCM like RNLN/ME.
b. Data quality can be measured by assessing a number of characteristics; most used are: accuracy, completeness, consistency, economy, reliability and timeliness.
c. The methods used in the researched companies to generate cost- and performance data are mainly manually and with help of maintenance software, assembled in flat, isolated and not integrated databases.
d. These methods are mainly not appropriate and cost effective.
e. Several factors are the cause why these methods are not appropriate and cost effective: "time" (including "money") is the most important factor. The problem is probably such great and complex that corrective actions have not been taken. Motivation and data collection procedures which are not properly settled are also important reasons.
f. The relation between cost of data collection and data quality is comparable with costs of product quality and will be exponentially or asymptotically.
g. An alternative method to generate reliable cost- and performance data is implementation of maintenance software. Several packages are available.
h. The main reasons that do influence reliability of data significantly at the point of registration are lack of knowledge and motivation and a complex, not user-friendly interface.
i. The opportunities that are available to optimise the generation of appropriate, reliable cost effectiveness data: keep it as simple as possible, not complex, but comprehensible and simple to understand, complete, clear and meaningful. Give more priority to data, treat it like a product or an asset. Beware of data corruption during collection, transfer, summary or storage of data.
Optimizing data collection for cost effective control of assets

- It is necessary to focus on data quality. We pay too little attention to data quality. Company members should be awakened on this subject by means of education, training, course or guidance to provide higher quality data for asset management control. Management courses (asset, data, information etc.) should include a module “Data quality”. Data should be treated as an asset, because it is an economic issue and should be cost effective.

- Legacy-data is often dirtier than we realize, but it should be prevented to collect dirty data. The principle should be: “Start with high-quality data or do not start at all”! Collecting data which does not meet certain pre-settled quality requirements is useless and should be avoided. It is a waste of time, energy and money. The old saying “garbage in, garbage out” is still true. Credibility is connected with this item and when this is missing, we have a problem: there is no confidence in the information processed from data, in spite of the real quality.

  The alternative is: assess (measure) data quality and plan for improvement.

- Bad data causes not only dissatisfaction and a waste of money, resource, sometimes huge amounts and much more than needed. Bad data may on the other hand also cause loss of performance. Both options must be prevented.

- Data quality is declining by not using data. The only way to truly improve data quality is to increase the use of that data. There is no alternative for data quality: use it or lose it!

- There can be many ways why data quality is bad. It is essential to coordinate and manage data quality activities. It is extremely important that someone is assigned to data quality who is responsible for data quality to assist in reaching and ensuring the required data quality.

- Management decisions require reliable information out of high-quality data. High-quality data is data that is fit for use by data consumers. Escaping from the requirement of high-quality data is not smart because it costs much more money than needed.

  But we can even make it worse. Cost effectiveness is declining by complexity (while data quality is not improving) and that is not the purpose of the LCM-idea. But the most fundamental conclusion is that LCM is hard to implement with bad data by which good quality information can not be produced.

- Information is necessary to manage LCM or arrange a cost effective control of assets. LCM needs information and requires data, more precise: high-quality data
during life cycle. System managers, engineers, technicians can not produce LCM-information which only is based on instant inquiry, experience and knowledge. This will end in window-dressing and that is not only aimless, it is also not cost effective and a waste of money.

- The most possible causes of poor data quality are time, ambiguous criteria, motivation and dysfunctional ICT. The complexity of the data quality problem and available resources are important causes of not performing the intended possible solutions.

- A company without competition has no drive to perform cost effectiveness and will not be focused on acting this way. If the customer only wants asset performance, the support cost of the asset will be neglected and is the basic idea of LCM hard to implement.

- RNLN/ME has a new challenge the coming years by implementing an ERP-system and by simultaneously cleansing reaching the level of high-quality data.

- Database systems that can integrate data and provide the analysis to support decisions are necessary. Implementation of ERP-systems is an opportunity for good quality data where the company data is integrated, but one of the requirements is that the legacy-data must be cleansed before data migration. This part of the process should not be underestimated. After implementation a well-organized system should keep the data clean, but data cleansing is not a good option.

ERP-systems are expensive database systems, not only for acquisition and periodical licenses, but also for personnel, training, maintenance and support (including expensive “insourced” consultancy). There are many full time equivalents needed for collecting data and that debits the cost effectiveness. At the other hand, the ERP-system is available, the only challenge is: extract the needed data and direct it to the datawarehouse.

ERP-systems are suitable for integrated processes in administration, financials, planning and so on, but it is very doubtful if they are really suitable for cost effective control of assets. They are an opportunity for good quality data, but they are also complex, expensive and not flexible.
8. Recommendations

- Data.
  Collect only what is needed, not all data which is available. Focus for instance primarily on the data collection of expensive, failure-sensitive and mission-critical parts.
  Data quality and cost effectiveness are positively influenced when unnecessary and optional data are omitted in the data set. Do not collect dirty, bad data: it costs a lot of time, energy and money and is not cost effective. Measure data quality and plan for improvement.

- Priority.
  Handle data management with the same priority as asset management. Assign the highest priority for ensuring to collect high-quality data. Without this data well-based decisions are not possible. Convince top-management of the importance for decision making and cost effectiveness of these decisions. Top-management should initiate and support data quality decisions.

- Responsibility.
  Make someone responsible for data quality. It is not cost effective to appoint a “Data Quality Manager”, but embed this task in an involved company member. Someone should clearly be the “owner” of the data, who coordinates and manages data quality activities. Make this person aware of his/her responsibilities.

- Hardware/software.
  Aim at a DBMS and data warehouse where company data is integrated, but do not make this system too complex. It will cost huge amounts of money, it will take much effort to maintain and it will not succeed.

- Education.
  Make the subject “data quality” an inherent and integral part of management courses, beginning with the AMC-course module “Information Management”.

9. **Suggestions for further work**

- Research ERP-systems (for instance PLM, LCM- and SCM-modules) if and how they can be suitable for cost effective control of assets.
- Investigate the cost effectiveness of data warehousing and the impact of data warehousing in the transformation process from data to information and knowledge.
- Perform an investigation of dirty data on the cost of asset support and impact on asset performance.
- Develop a data quality assessment method with a data quality index in order to assess data quality, watch trends and plan to improve data quality.
### 10. Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>AK</td>
<td>Availability Killer</td>
</tr>
<tr>
<td>AMC</td>
<td>Asset Management Control</td>
</tr>
<tr>
<td>AMICO</td>
<td>Asset Management Information &amp; Communication</td>
</tr>
<tr>
<td>BBS</td>
<td>Bedrijfs Beheer Systeem (in Dutch)</td>
</tr>
<tr>
<td>CD</td>
<td>Cost Driver</td>
</tr>
<tr>
<td>CE</td>
<td>Cost Effectiveness</td>
</tr>
<tr>
<td>CIPIER</td>
<td>Cost Information Performance Information Effectiveness Report</td>
</tr>
<tr>
<td>DBMS</td>
<td>Database management system</td>
</tr>
<tr>
<td>DDL</td>
<td>Data Definition Language</td>
</tr>
<tr>
<td>DM</td>
<td>Data Management</td>
</tr>
<tr>
<td>DMO</td>
<td>Defence Material Organization</td>
</tr>
<tr>
<td>DMS</td>
<td>Data Migration Street</td>
</tr>
<tr>
<td>DoD</td>
<td>Department of Defence</td>
</tr>
<tr>
<td>DoE</td>
<td>Department of Energy</td>
</tr>
<tr>
<td>DQI</td>
<td>Data Quality Index</td>
</tr>
<tr>
<td>DQM</td>
<td>Data Quality Manager</td>
</tr>
<tr>
<td>DTO</td>
<td>Defence Telematics Organization</td>
</tr>
<tr>
<td>ERP</td>
<td>Enterprise Resource Planning</td>
</tr>
<tr>
<td>ETL</td>
<td>Extract, Transform, Load</td>
</tr>
<tr>
<td>FMECA</td>
<td>Failure Mode Effects &amp; Criticality Analysis</td>
</tr>
<tr>
<td>FTE</td>
<td>Full Time Equivalent</td>
</tr>
<tr>
<td>ICT</td>
<td>Information &amp; Communication Technology</td>
</tr>
<tr>
<td>ILS</td>
<td>Integrated Logistic Support</td>
</tr>
<tr>
<td>IPM</td>
<td>Information Product Manager</td>
</tr>
<tr>
<td>KPI</td>
<td>Key Performance Indicator</td>
</tr>
<tr>
<td>LCA</td>
<td>Life Cycle Assessment</td>
</tr>
<tr>
<td>LCM</td>
<td>Life Cycle Management</td>
</tr>
<tr>
<td>LORA</td>
<td>Level Of Repair Analysis</td>
</tr>
<tr>
<td>LSAR</td>
<td>Logistic Support Analysis Record</td>
</tr>
<tr>
<td>MATRACS</td>
<td>Material Report and Configuration System</td>
</tr>
<tr>
<td>MBS</td>
<td>Materiaal Beheer Systeem (in Dutch)</td>
</tr>
<tr>
<td>Acronym</td>
<td>Description</td>
</tr>
<tr>
<td>---------</td>
<td>-------------</td>
</tr>
<tr>
<td>MCS</td>
<td>Materiaal Configuratie Systeem (in Dutch)</td>
</tr>
<tr>
<td>MDM</td>
<td>Master Data Management</td>
</tr>
<tr>
<td>MIS</td>
<td>Management Information System</td>
</tr>
<tr>
<td>MoD</td>
<td>Ministry of Defence</td>
</tr>
<tr>
<td>MTA</td>
<td>Maintenance Task Analysis</td>
</tr>
<tr>
<td>MTBF</td>
<td>Mean Time Between Failures</td>
</tr>
<tr>
<td>MTTF</td>
<td>Mean Time To Failures</td>
</tr>
<tr>
<td>NPDM</td>
<td>NATO Product Data Model</td>
</tr>
<tr>
<td>OLAP</td>
<td>OnLine Analytical Processing</td>
</tr>
<tr>
<td>OLTP</td>
<td>OnLine Transactional Processing</td>
</tr>
<tr>
<td>PBM</td>
<td>Performance Based Management</td>
</tr>
<tr>
<td>PDCA</td>
<td>Plan, Do, Check, Act</td>
</tr>
<tr>
<td>PDM</td>
<td>Product Data Management</td>
</tr>
<tr>
<td>PID</td>
<td>Project Initiation Document</td>
</tr>
<tr>
<td>PKCD</td>
<td>Performance Killers and Cost Drivers</td>
</tr>
<tr>
<td>PLM</td>
<td>Planned Maintenance</td>
</tr>
<tr>
<td>PRAWDA</td>
<td>Project Automatisering Wapentechnische Dienst Administratie (in Dutch)</td>
</tr>
<tr>
<td>RAPID</td>
<td>Rapportage Platform Incidenten en Defecten (in Dutch)</td>
</tr>
<tr>
<td>RCM</td>
<td>Reliability Centred Maintenance</td>
</tr>
<tr>
<td>RNLN/ME</td>
<td>Royal Netherlands Navy/ Maintenance Establishment (in Dutch: Koninklijke Marine/Marinebedrijf)</td>
</tr>
<tr>
<td>SCM</td>
<td>Supply Chain Management</td>
</tr>
<tr>
<td>SE</td>
<td>System Effectiveness</td>
</tr>
<tr>
<td>SM</td>
<td>System Management</td>
</tr>
<tr>
<td>STORM</td>
<td>Systems Technical Overview of Reliability and Maintainability</td>
</tr>
<tr>
<td>TCP</td>
<td>The Cleansing Project (UK/MoD)</td>
</tr>
<tr>
<td>TQM</td>
<td>Total Quality Management</td>
</tr>
<tr>
<td>VAS/LOGDOC</td>
<td>Voorraad Administratie Systeem/ Logistieke Documentatie (in Dutch)</td>
</tr>
<tr>
<td>WSM</td>
<td>Weapon System Management</td>
</tr>
</tbody>
</table>
11. Bibliography

2. Algemene Rekenkamer; 2006; Managementinformatie over inzetbaarheid Defensie (in Dutch)
4. Baarda D.B/ Goede M.P.M.; Teunissen J.;1997; Kwalitatief onderzoek; Stenfert Kroese (in Dutch)
12. Blankena F.; 2005; Bedrijfsgegevens zitten vol gebreken; Automatiseringsgids, 7 dec. 2005 (in Dutch)
16. Brink A.; 1997; Analysis of information services (in Dutch); RNLN
19. Calbo P.R./ Polle W.A.P.; 2005; Gereedsmaken waardevolle materieel-gerelateerde gegevens voor migratie naar een ERP-systeem; RNLN/ME (in Dutch)
21. Carlson R./ Pålsson A.C.; 1997; Adaptations of the Swedish national LCA database system; Chalmers University of Technology, dept. of IMI - Industrial Environmental Informatics, Göteborg, Sweden
26. CIHI; 2005; The CIHI data quality framework; Canadian Institute for Health Information, June 2005
27. Clikeman P.M.; 1999; Improving information quality; The Internal Auditor, June 1999, pg.32-33
29. Davis G.B./ Olson M.H. ; 1985; Management Information Systems; McGraw-Hill
31. Derksen P./ Brink A.; 1993; Eindrapport werkgroep PRAWDA-BWIS gerelateerde problemen, RNLN/ME (in Dutch)
33. Doorewaard/Verschuren; 2005; Het ontwerpen van een onderzoek
34. Dwornick R.; 1999; Factors influencing data quality; Direct Marketing, Apr 1999, pg.45-47
37. English L.P. ; 1999; Improving Data Warehouse and Business Information Quality: Methods for Reducing Costs and Increasing Profits; Wiley&Sons, New York, USA
38. English L.P. ; 2002; Mistakes to avoid if your data warehouse is to deliver quality information; DMreview, Jun 2002
39. EEA; 1997; Life Cycle Assessment (LCA), European Environment Agency
40. Fermont H.M.A.; 2002; Belangrijke vereisten voor goede prestatie-informatie ; Bedrijfswet, Jaargang 74, nr. 4, pg.38-43 (in Dutch)
41. Fest Glen; 2005; Bad data: filling in the blanks; Bank Technology News, New York, 1 Nov 2005, Vol.18, Iss.11, pg. 49
42. Fitzgerald K.; 2005; Weeding Out Bad Data; Credit Card Management, Mar 2005, pg.20-23
43. Gameren A.v.; 1998; Een geïntegreerd data-acquisitiesysteem; KIM (in Dutch)
44. Goff J.; 2003; Drowning in data; CFO Magazine 1 November 2003
45. Grady J.O.; 1998; System validation and verification; CRC-Press
47. Guess F.M.; 2000; Improving information quality for 21st century; University of Tennessee
48. Harjinder S./ Prakash C.; 1996; The official guide to data warehousing; QUE Corp.
49. Have W.v.d.; 2007; Stil verzet tegen het softwarepakket; Maintenance Management, nr.1/2,2007 (in Dutch)
50. Heina N.; 2005; Portals belangrijke trend in onderhoud; PT42, nr.11, nov.2005 (in Dutch)
51. Hernandez T.; 2006; Improving Factory Performance Via Web-Based Data Collection; Manufacturing Engineering, Dec 2006, pg.34-37
52. Hoogenboom; 2002; Betrouwbaarheidsonderzoek BBS; RNLN/ME (in Dutch)
54. INCOSE; 1998; Systems engineering handbook: a “how to” guide for all engineers; INCOSE, Seattle, USA
55. Informatica; 2007; Using Informatica Data Quality 3.1 level 1
56. Inmon W. et al; 1997; Managing the data warehouse; Wiley&Sons, New York
57. Inmon W.; 1998; Controlling data warehouse costs
58. Inmon W. et al; 1999; Data warehouse performance; Wiley&Sons, New York
59. Inmon W.; 2005; The data warehouse foundation
60. ISO/IEC 9126; 2001; Software engineering, Product Quality
61. ISO 14041; 1998; Environmental management, Life cycle assessment, Goal and scope definition and inventory analysis
63. Januschkowetz A.; 1997; Use of ERP-systems for LCA and PS; Dissertation Dresden University of Technology
69. Kaplan R.S.; 1990; The Four-Stage Model of Cost Systems Design; Management Accounting, Feb 1990, pg. 22-26
71. Kay E.; 1997; Dirty data challenges warehouses; Software Magazine, October 1997, pg. S5-S8
73. Kim W./ Choi/ Hong/ Kim/ Lee; 2003; A Taxonomy of Dirty Data; Data Mining and Knowledge Discovery, Jan.2003, Vol.7, pg.81–99
74. Kimball R.; 1996; The data warehouse toolkit; John Wiley & Sons
75. Kirkwood Hal P. Jr.; 2006; Data Processing and Data Management, pg.155-159
76. Knight B.; 1992; The Data Pollution problem; Computerworld, 28 Sep 1992, pg.81-85
77. Knight, J.E.; 1999; Data Collection and Data Analysis: Issues in Need of Managerial Attention; Proceedings of the American Society of Business and Behaviour Sciences, Las Vegas, NV.
79. Knowles A.; 1997; Dump your dirty data for added profits; Datamation, Vol.43 No.9, September 1997, pg. 80-82
81. Kozlowska B.; 2005; How to write an essay and research paper
82. Krill P.; 1998; Data warehouses have need for clean data; InfoWorld, 16 Mar 1998, pg.27
84. Kristen G.; 1998; Kennis is macht, 2e druk; Academic Service, ISBN 90 395 0700 7 (in Dutch)
85. Lammerse H.; 1993; Bedrijfszekerheidigegevens verzamelen, evalueren, presenteren; Kluwer (in Dutch)
86. Latino K.; 2004; Understanding event data collection: Part 1; Plant Engineering, Jul 2004, pg.31-33
89. Lobregt H.W.G./ Brink A./ Pronk M.A./ Goossens A.A.J; 2004; Systems Analysis makes the difference in acquiring Asset Management Control; AMC-course part 7, Systems Analysis
90. Ma C./ Chou D.C./ Yen D.C; 2003; Data warehousing, technology assessment and management
91. MIL-HDBK-502; 1997; Acquisition Logistics; DOD, Washington DC, USA
92. MIL-STD-1388-2B; 1993; Requirements for a Logistic Support Analysis Record; DOD, Washington DC, USA
93. Missier et al; 2003; Improving data quality in practice: A Case Study in the Italian Public Administration; pg.135-160
94. Myers M.; 1995; Do you really need a data warehouse?; Network world, Dec 1995, pg.26
96. NATO CALS Office; 2001; NATO Product Data Model (NPDM) version 4.10
97. Nelson W.B.; 1982; Applied life data analysis; Wiley&Sons, New York, USA
100. NIVE; 1999; Project management
101. Oakland John S; 1993; Total quality management; Butterworth-Heinemann
102. Olson J.E.; 2002a; How accurate is data?; Communication News, June 2002
103. Olson J.E.; 2002b; The data quality problem (from “Data quality, the accuracy dimension”); Morgan Kaufmann Publishers, ISBN 0585449953
104. Olson J.E.; 2007; Data quality, the accuracy dimension; Morgan Kaufman Publishers, ISBN 1558608915
Optimizing data collection for cost effective control of assets

106. Oorburg D.P.E.; 2003; ERP-tussen wal en schip?; KIM (in Dutch)
108. Oxbrow A.; 2006; Data quality is key to asset management; Utility Week, 3 Nov 2006, pg.16
109. Pålsson A.C.; 2002; An industry common methodology for environmental data management; Platform presentation, Stockholm, Sweden
110. Paridon K.; 2004; Product Data Management met legacysystemen; applicaties en prototypes (in Dutch)
111. Parsons R./ Brunton A.; 1993; Improving team data collection efforts; Journal for Quality and Participation, Cincinnati, Sep 1993, Vol.16, Iss.5; pg.56-57
113. Pedley S., 2000; Data management is the key to success; Petroleum Economist, May 2000, vol.67, iss.5, pg.42-43
114. Peelen E.; 2006; Data quality management; Nyenrode Business University
115. Perik K./ Weverink R.; 2007; Beste informatie staat op achterkant sigarendoos; Automatiseringsgids 07/2007 (in Dutch)
116. Petersen K.J.; 1999; Effect of information quality on supply chain performance; Michigan State University, dissertation
120. Redman, T.C.; 1997; Data Quality for the Information Age; Norwood, MA: Artech House
122. Redman, T.C.; 2001; Data Quality the field guide; Digital Press, ISBN 1555582516
123. Revelle J.B.; 2006; All about data; Quality Progress, Milwaukee, Jan 2006, Vol.39, Iss. 1; pg.96
124. RNLN; 2001; Audit configuratiemanagement; DEB/Afd. Beleidsevaluatie en Audit (in Dutch)

125. RNLN; 2004; ILS Handboek versie 1.5 (in Dutch)


127. Sammon D./ Finnegan P.; 2000; The ten commandments of data warehousing; Database for Advances in Information Systems; Fall 2000, vol.31, no.4, pg. 82-91

128. Schröder R.P.M.; 2003; Integratie van de materieelinformatiesystemen der KM; KIM (in Dutch)


130. Schulz G.; 2001; Information drives asset management; Manufacturing Engineering. Sep 2001, pg. 200


132. Smit K.; 1986; Onderhoudsmanagement; TU-Delft (in Dutch)


134. Stam A.J. et al; 1996; Instandhoudingsfunctie binnen de KM/SB (in Dutch)


137. Steenis H.v.; 2006; Asset management blijft steken bij de instrumentatie; PT-Automatisering, nr7/8-aug.2006


139. Tayi/ Ballou; 1998; Examining data quality; Communications of the ACM, February 1998, Vol. 41, No. 2, pg.54-57

140. Te’eni D.; 1993; Behavioral aspects of data production and their impact on data quality; Hershi, Spring 1993, Vol.4, Iss.2, Pg.30-39

141. Trembly A.C.; 2002; Poor data quality: A $600 billion issue?; National Underwriter, 18 Mar 2002, pg.21
144. Verschuren P.J.R./Wilms E.F.; 1994; Storingsregistratiesystemen; KIM (in Dutch)
146. Wallace B.; 1999; Data quality moves to the forefront; InformationWeek, 20 Sep 1999, pg.52-64
147. Wang R.Y./ Kon H./ Madnick S.; 1993; Data Quality Requirements Analysis and Modeling; Massachusetts Institute of Technology, pg.670-676
150. Wang R.Y./ Lee / Pipino / Strong; 1998; Manage Your Information as a Product; Sloan Management Review; Summer 1998; vol.39, no.4, pg.95-105
151. Wei N.S.; 2003; Drowning in data-starving for information; Pollution Engineering; Sep 2003, vol.35, no.8, pg.37-39
153. Whiting Rick; 2006; Hamstrung By Defective Data; Information Week, 8 may 2006
154. Whittle Sally; 2004; Clean data, clean profit; Computer Weekly, 21 Sep 2004, pg. 48-49
Internet sources

http://databases.imi.chalmers.se
www.babsonknowledge.org/analytics.pdf
www.billinmon.com
www.dataquality.com
www.dbazine.com/datawarehouse/dw-articles/dataquality
www.dmreview.com
www.dw-institute.com
www.eur.nl
www.inmoncif.com
www.key_performance_indicator.htm
www.maintenance_technology.com
www.oracle.com
www.orau.gov/pbm/pbmhandbook/pbmhandbook.html
www.stair.principles_of_information.index
www.storkmaintenancemanagement.com
www.surveyconsole.com
www.surveysystem.com.htm
www.surveyworld.net
www.tech-faq.com/erp.shtml
www.vnsi.nl
Appendix A. Research questions

Geachte

Ik ben op zoek naar ingangsbronnen voor mijn research op het gebied van het zo kosteneffectief mogelijk instandhouden van assets, de collectie van data daarvoor en vooral het optimaliseren daarvan. We moeten dan denken aan performance data (beschikbaarheid etc.) en kostedata (loon, materialen etc.). Maar het mag ook om gelijkssoortige data/informatie gaan.
Wanneer u zich hierbij betrokken voelt, wilt u dan s.v.p. onderstaande vragen beantwoorden en het resultaat verzenden.
Zo niet, wilt u de e-mail dan doorsturen naar een collega? Bij voorbaat dank!

Vraag 1
Bij welk bedrijf werkt u?

Vraag 2
Wat is uw functie in dit bedrijf?

Vraag 3
Uitgaande van de stelling, dat data wordt verzameld voor een kosteneffectieve instandhouding van assets: Welke methode(s) wordt(en) in uw bedrijf gebruikt om data te verzamelen?
- handmatig
- m.b.v. onderhoudssoftware
- vragenlijst, achteraf
- automatisch
- diagnostisch
- Is extern uitbesteed
- Anders:

Vraag 4
Wat voor soort Database Management Systeem (DBMS) gebruikt uw bedrijf om de data in op te slaan?
- Relationeel
- Hierarchisch
- Netwerk
- Diverse losstaande databases
- "Platte" database (bv. Excel)
- Anders:
Vraag 5
Wie voert de data in?

☐ Gebruiker/operator
☐ Engineer
☐ Database-/ICT-beheerder
☐ Administratief beheerder
☐ Manager
☐ Anders:

Vraag 6
Er is een bepaalde hoeveelheid data nodig om managementinformatie te produceren. Uw bedrijf verzamelt:

☐ veel te weinig data
☐ te weinig data
☐ voldoende data
☐ te veel data
☐ veel te veel data

Vraag 7
Wat is volgens uw inschatting de kwaliteit van de verzamelde data?

☐ slecht
☐ matig
☐ redelijk
☐ goed
☐ uitstekend

Vraag 8
Kunt u hieronder aanvinken in welke mate de data aan bepaalde kwaliteitseisen voldoet

<table>
<thead>
<tr>
<th></th>
<th>niet</th>
<th>wel</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accuraat</td>
<td>c</td>
<td>c</td>
</tr>
<tr>
<td>Compleet</td>
<td>c</td>
<td>c</td>
</tr>
<tr>
<td>Beschikbaar</td>
<td>c</td>
<td>c</td>
</tr>
<tr>
<td>Betrouwbaar</td>
<td>c</td>
<td>c</td>
</tr>
<tr>
<td>Relevant</td>
<td>c</td>
<td>c</td>
</tr>
<tr>
<td>Eenvoudig, simpel</td>
<td>c</td>
<td>c</td>
</tr>
<tr>
<td>Up-to-date</td>
<td>c</td>
<td>c</td>
</tr>
<tr>
<td>Verifieerbaar</td>
<td>c</td>
<td>c</td>
</tr>
<tr>
<td>Flexibel</td>
<td>c</td>
<td>c</td>
</tr>
<tr>
<td>Kosteneffectief, goedkoop</td>
<td>c</td>
<td>c</td>
</tr>
</tbody>
</table>
Vraag 9
Kunt u aangeven welke factoren het verzamelen van data negatief beïnvloeden?

- tijd
- geld
- kennis
- kunde
- bedrijfsomstandigheden
- motivatie
- apparatuur, meetmiddelen etc.
- gebreken in geautomatiseerd systeem
- geen eenduidige criteria
- Anders:

Vraag 10
Verifieert, controleert en schoont uw bedrijf de verzamelde data?

- Ja
- Nee

Evt. commentaar:

Vraag 11
Hoe wordt bij uw bedrijf de data-analyse uitgevoerd?

- Geïntegreerd in DBMS
- Download uit DBMS in analysetool (bv. SPSS)
- Oracle/SQL
- Excel / Access etc.
- nvt

Commentaar:

Vraag 12
Na het verzamelen van data wordt deze in het algemeen bewerkt tot (management) informatie. Vraag: Bent u tevreden met de informatie die uiteindelijk beschikbaar komt?

- Ja, zeer tevreden
- Ja, redelijk
Optimizing data collection for cost effective control of assets

Vraag 13
Wat wordt uiteindelijk gedaan met de verzamelde data en de informatie?
- Niets of te weinig
- Onderhoudssupport
- Modificatie-/ontwerpwijzigingen
- Logistieke aanpassing
- Proceswijzigingen
- Rapporten, overzichten, grafieken
- Anders:
  
Vraag 14
Is de hele procedure van datacollectie volgens u kosteneffectief?
- Ja
- Nee

Vraag 15
Kunt u aangeven welke maatregelen ter verbetering mogelijk zijn?

Vraag 16
Wat belemmert u om deze maatregelen uit te voeren?

Verzenden
Appendix B. Results of research

Question 3. Which method(s) is your company using to collect data?

- Automatic: 22%
- Manually: 38%
- Diagnostic: 3%
- Internally: 6%
- Maintenance software: 28%

Question 4. What sort of Database Management System (DBMS) is your company using to store the collected data?

- Flat database: 25%
- Relational database: 20%
- Isolated databases: 35%
- Network: 5%
- Access database: 5%
- SAP-PM: 10%

Question 5. Who is responsible for registration of the data?

- User/operator: 36%
- Engineer: 26%
- Database/ICT-personnel: 9%
- Administrative personnel: 17%
- Other: 3%
- Manager: 9%
**Question 6.** A certain quantity of data is needed to collect. Your company is collecting:

![Quantity of collected data chart](chart1.png)

**Question 7.** What is your opinion of the quality of the collected data?

![Quality of collected data chart](chart2.png)

**Question 8.** Can you indicate the quality criteria of the collected data below?

![Accuracy chart](chart3.png)
Optimizing data collection for cost effective control of assets

8.2 Completeness
mean: 2.6

8.3 Availability
mean: 2.85

8.4 Reliability
mean: 2.5
Optimizing data collection for cost effective control of assets

8.5

Relevancy
mean: 3.3

8.6

Simplicity
mean: 2.75

8.7

Timeliness
mean: 2.95
Optimizing data collection for cost effective control of assets

### Verifiability
- Mean: 2.8

<table>
<thead>
<tr>
<th>Percentage</th>
<th>1. NO</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5. YES</th>
</tr>
</thead>
<tbody>
<tr>
<td>0%</td>
<td>10%</td>
<td>25%</td>
<td>40%</td>
<td>25%</td>
<td></td>
</tr>
</tbody>
</table>

### Flexibility
- Mean: 2.3

<table>
<thead>
<tr>
<th>Percentage</th>
<th>1. NO</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5. YES</th>
</tr>
</thead>
<tbody>
<tr>
<td>0%</td>
<td>20%</td>
<td>45%</td>
<td>25%</td>
<td>5%</td>
<td>5%</td>
</tr>
</tbody>
</table>

### Cost effectiveness
- Mean: 2.25

<table>
<thead>
<tr>
<th>Percentage</th>
<th>1. NO</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5. YES</th>
</tr>
</thead>
<tbody>
<tr>
<td>0%</td>
<td>25%</td>
<td>40%</td>
<td>20%</td>
<td>15%</td>
<td>15%</td>
</tr>
</tbody>
</table>
Question 9. Can you indicate which factors are negatively influencing the collecting of data?
Question 10. Does your company audit, control and clean the collected data?

![Pie chart showing 55% No and 45% Yes]

Question 11. How is your company performing the data-analysis?

![Pie chart showing 75% Excel/Access etc. and 25% Integrated in DBMS]

Question 12. Are you satisfied with the ultimate available information?

![Bar chart showing satisfaction levels with mean 2.6]
Question 13. What has been done with the collected data and information?

![Pie chart showing the distribution of data usage]

- Reports, overviews, graphics etc.: 27%
- Plans: 6%
- Process changes: 11%
- Maintenance support: 17%
- Design changes: 21%
- Logistic changes: 9%
- Nothing or too little: 9%

Question 14. Is in your opinion the whole procedure of data collection cost effective?

![Pie chart showing the response to cost effectiveness]

- Yes: 40%
- No: 60%