THE NECESSITY OF INFORMATION QUALITY FOR EFFECTIVE BUSINESS INTELLIGENCE

P.J. Conradie[1] and P.S. Kruger[2]

Department of Industrial and Systems Engineering
University of Pretoria, South Africa
[1]pieter.conradie@eng.up.ac.za, [2]paul.kruger@eng.up.ac.za

ABSTRACT

Data quality, or rather the lack thereof, has been the downfall of many Data warehousing (DW) / Business Intelligence (BI) projects. The concept of Information Quality (IQ) has become increasingly important as the popularity of Business Intelligence (BI) increases. Industrial engineers have played a key role in the development and implementation of Total Quality Management (TQM) principles that are widely used in the manufacturing environment. In the information era, industrial engineers can apply the same principles to data. This article aims to identify strategies and processes to improve IQ that will eventually lead to data warehouses with complete, accurate and timely data that consistently meets the requirements of knowledge workers in their quest to make improved business decisions.

OPSOMMING

Die gebrek aan datakwaliteit het al menige Databergings- en Besigheidsintelligensie-projekte laat sneuwel. Die konsep van Inligtingskwaliteit word al meer prominent soos die gewildheid van Besigheidsintelligensie toeneem. Bedryfsingenieurs het reeds 'n sleutelrol gespeel in die ontwikkeling en implementering van Totale Kwaliteitsbestuurbeginsels wat algemeen in die vervaardigingsomgewing gebruik word. In die inligtingsera kan bedryfsingenieurs dieselfde beginsels ook op data toepas. Hierdie artikel identifiseer strategieë en prosesse om inligtingskwaliteit te verbeter wat uiteindelik behoort te lei tot datastore met volledige en akkurate data wat betyds en konsekwent aan kenniswerkers se behoeftes voldoen en hulle instaat stel om beter besigheidsbesluite te neem.

1. INTRODUCTION

Traditionally industrial engineers have been involved in the improvement of resource utilization and improved productivity through changes in processes, automation of machinery, layout of plants, scheduling of activities, balancing of resources, and so forth. However, since the 1980s the information resource has become the main driver of most competitive advantages. More and more industrial engineers are involved in the development, implementation and integration of information systems to streamline businesses that are in the manufacturing, retail or service industries.

Modern Industrial Engineering is defined by Sperotto [1] as being "concerned with the integration of resources and processes into cohesive strategies, structures and systems for the effective and efficient production of quality goods and services." Information systems are part of the resources that industrial engineers need to consider in their quest to improve enterprises. Business intelligence (BI), according to the definition by Kimball and Ross [2], is a generic term to describe leveraging the organisations' internal and external information assets for making better business decisions. BI related information systems in particular are therefore an important resource that should provide quality inputs for improved business decisions.

Industrial engineers have played a key role in the development and implementation of Total Quality Management (TQM) principles that are widely used in the manufacturing environment. The purpose of this article is to indicate that industrial engineers can apply the same principles in the creation of non-tangible "products" such as information from data warehouses, by analysing and improving the relevant business processes. Strategies and processes are identified to improve information quality (IQ) that will eventually lead to data warehouses with complete, accurate and timely data that will consistently meet the requirements of knowledge workers in their quest to make better business decisions.

The framework of the article consists of a brief overview on the development of information systems over the past fifty years, the growing need for analytical information systems, data warehousing as a foundation for BI, typical business measurements that need to be supplied by BI systems, the generic question of quality in information delivery and finally the discussion of a TQIM (Total Quality Information Management) approach as suggested by English [16].

2. THE DEVELOPMENT OF INFORMATION SYSTEMS

The use of electronic information processing has changed dramatically from mere automation of routine business data handling such as electronic accounting machines in the 1950s and throughout the 1960s, to very sophisticated and integrated systems linking organisations and individuals worldwide.

The types of information systems have grown through various phases from very structured transactional processing systems to less structured, business intelligence systems. Frenzel [3] distinguishes between:

- Transactional processing systems (TPS), such as order-entry programs and general ledger financial systems.
- Decision support systems (DSS).
- Office automation systems (OAS), such as e-mail and electronic diaries.
- Management information systems (MIS).
- Expert systems designed to augment human reasoning in complex analysis.

Since the mid 1990s the term Business Intelligence (BI) has become popular with various information systems supporting the concept of BI. According to Howard Dresner, (who has coined the term "Business Intelligence" in 1989 according to Connelly et al. [4]), this new generation of BI tools supersedes the earlier generation called decision support and management information systems.

3. A GROWING NEED FOR ANALYTICAL INFORMATION SYSTEMS

Although transactional information systems will always be necessary to capture data at various points in business processes, the more interesting advances in information systems have been on the analytical side. The growth in popularity of Customer Relationship Management (CRM) systems since the late 1990s, is a typical example of how businesses realise that by analysing their customer information in many different ways, they can serve and satisfy their customers in a much better way. They can, for example, identify cross selling opportunities, target and time their marketing efforts much better, provide personalized service and much more.

Connelly et al. [4] highlights the fact that most managers are multidimensional, meaning that they should look at information from different angles – grouping, filtering and sorting the same data in many different ways to uncover those issues that they must address to perform better. Transactional systems, which are running on relational databases, normalised to the third degree to enable ease of data capturing, were never designed to facilitate information retrieval in the same way. Business intelligence applications are therefore multidimensional by nature to primarily support analytical processing with ease. Some of the differences between OLTP (On-line transactional processing) and OLAP (On-line analytical processing) are given in Table 1.

It should be clear that both these two types of data processing are necessary to support the objectives of efficient data capturing and effective business analysis in terms of queries and reports. Even though the same relational database technology can be used for both OLTP and OLAP systems by using different data model structures, some OLAP systems run on multidimensional database technology.

Since BI systems are mostly analytical of nature, they are naturally based on an OLAP approach, but they rely very much on the data that is captured through transactional systems. Data quality problems in BI systems can very often be traced back to data quality problems in the transactional systems that are used as sources for the BI systems.

OLTP	OLAP
Optimized for data capturing	Optimized for queries
Normalised data models	De-normalised data models
Designed for efficiency	Designed for effectiveness
Automate processes	Support decision making
Atomised data	Summarized data
Record-at-a-time	Many-records-at-a-time
Current	Historical
Process oriented	Subject oriented
Structure the business	Adapt to the business strategy
React to events	Anticipate events

Table 1: Differences between OLTP and OLAP systems
(Adapted from [5] and [6])

4. DATA WAREHOUSING IN SUPPORT OF BI

Since the early 1990s different BI and data warehousing "gurus" have emerged, publishing various blueprints of BI methodologies. Without downplaying the role of many other contributors, the names of Bill Inmon (sometimes referred to as the "Father of data warehousing") and Ralph Kimball stand out. Inmon has established the "Corporate Information Factory" (CIF) architecture and has written many books and articles to explain and develop the concept [7]. His CIF covers the whole BI scope and includes:

- Activities, which are the events that are captured by the enterprise.
- Data acquisition area that includes the operational or transactional information systems and the integration/transformation area.
- Primary Storage Management area that includes the data warehouse (DW) and the operational data store (ODS).
- Data delivery area where subsets of data are delivered through statistical analysis, analytical applications and data marts.

Inmon et al. [7] also include meta data management and the links to external data as part of his bigger picture BI framework.

Kimball et al. [6] see a similar generic picture with extraction of data from the operational systems to a staging area (for data cleansing and transformation purposes as back office processes), which in turn feeds the data warehouse that consists of various data marts in star scheme format. From the data warehouse many different information delivery applications can provide information to the end-users. Kimball and Inmon differ mainly on the format of the data warehouse. Kimball proposes that the data warehouse should be built data mart by data mart using the concept of conformed dimensions (which are shared dimensions between data marts) to link the various star schemes together, while Inmon promotes a fully integrated, normalised, relational data model for the data warehouse from where star scheme data marts can be extracted to satisfy specific end-user needs.

The name that Inmon gives to his model, "Corporate Information Factory", sounds very attractive to the ear of an industrial engineer. The concept of a processing plant where raw material (data) is sourced from various suppliers (captured through transactional systems and other sources), transformed into usable products (information in various formats) by well defined ETL processes (Extraction, Transformation and Loading) and according to specifications (user requirements and functional specifications) that are under configuration management (meta data), using carefully selected resources (BI tools, servers, trained people) for production, quality assurance, packaging (e.g. robot logic, OLAP cubes, static intranet reports) and distribution (e.g. via client/server, web services, Personal Digital Assistant (PDA) devices, cell phones), are all too familiar concepts to the industrial engineering discipline.

However, although the authors fully support the concept of a corporate information factory where industrial engineering principles can be applied, the Kimball approach to the design of the data warehouse (simply data mart by data mart, driven by specific business needs and glued together by the Bus Architecture of conformed dimensions), leads the authors to favour the Kimball approach when developing a data warehouse to support BI requirements. The idea to accommodate the detailed transactional data requirements in a detailed data mart as part of the data warehouse, instead of a separate ODS as suggested by Inmon, is a further plus point for the Kimball approach.

Each star scheme, according to the Kimball approach, consists of a fact table and a number of dimensions. The fact table has foreign keys to all the dimensions with normally a number of additional facts for each record. Each star scheme addresses a specific subject/business process at a certain level of detail, also referred to as the grain of the fact table. A typical example of a star scheme that will support various multidimensional questions regarding the subject of "Sales" is given in Figure 1. Although not shown in the figure, each dimension normally has many attributes that can be used to filter, or group, query results. For example, the Customer dimension would probably have (among many others) attributes like gender and age band; the Sales person dimension could have attributes like gender and years experience; and the Period dimension normally has a record for each day with attributes such as day of week, weekend indicator, month end indicator, public holiday indicator and many more. With normal SQL (Structured Query Language) statements and the simplified star scheme design, many relevant queries can be executed for analysis purposes.

One can visualise that through the star scheme each record in the fact table is not only linked to the foreign keys in the dimension tables, but also to all other attributes (or columns) in each dimension. When the records in the dimensions are maintained in the correct manner (see slow changing dimension types in Kimball et al. [6]), it is possible to have more than one record in a dimension for the same entity, reflecting the history of changes to attributes of that entity. If, for example, a customer gets married at a certain stage, the marital status attribute of that customer should change. Instead of just updating the attribute of the existing record, one could implement a Type II change according to the Kimball rules and add an additional record to the Customer dimension to which all future fact records are linked (until another change

leads to a further record for that customer). By handling dimension changes in this manner, different groups of facts can exist for the same customer, linked to the different foreign keys – each reflecting the correct attributes for the customer at the stage when the facts were recorded.
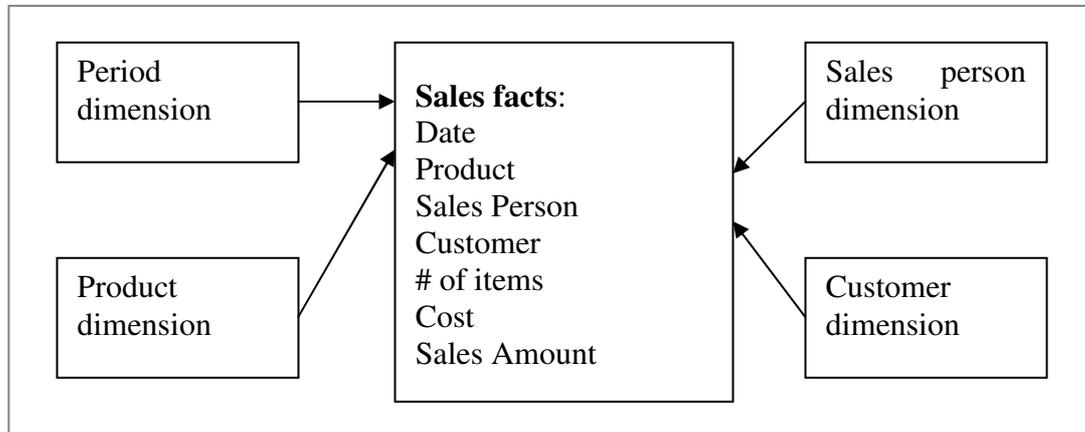
Figure 1: Example of a Star Scheme for dimensional reporting

The authors tend to support the Kimball methodology, since the incremental approach definitely delivers quicker and small successes are needed to fuel and fund the ongoing development of the data warehouse and relevant BI initiatives. The simplified structure of the star scheme data marts also provides a solid foundation for BI queries, avoiding complex joins in a fully normalised data model.

Given the above background, the question could be asked whether there can be different views on data warehousing. Naturally. However, if the idea of a full scale operational data store, replicating all possible data in the organisation crops up, it is not the definition that the authors would like to use in this case. By using the "Data Warehouse Bus Architecture" of building a succession of data marts that will, in time, create an overall data warehouse, as proposed by Ralph Kimball [6], "… it will free you from the requirement to build a centralized, all-encompassing data warehouse that must be completed before you release your first data mart." Seen in this context, the principle to run BI applications on a data warehouse, instead of directly on the transactional systems, begin to make sense. Add to that the way in which historical changes to the records in dimensions are handled (as discussed earlier) and the argument for a data warehouse, even in smaller enterprises, is normally won.

5. TYPICAL BUSINESS MEASUREMENTS TO BE SUPPLIED BY BI

In a business environment that is becoming more and more competitive, performance measurement is definitely not a luxury – it is a necessity for survival. Organisations need to measure themselves not only in a financial manner as reflected in their annual financial statements, but in many other aspects to anticipate events or circumstances and to steer the organisation in the right direction. These

measurements are the typical products that BI systems need to deliver accurately and consistently.

Over the years many performance measurement approaches have evolved. Management by objectives has become a popular management philosophy, but without regular, quantitative and consistent measurement of those objectives at various levels of the organisation, it cannot succeed. Benchmarking, as a technique, also implies that certain key performance indicators are measured and compared to industry standards where the same indicators are measured in the same, consistent way.

In an effort to align business strategy with personal behaviour, the performance of individuals in an organisation is also linked to the achievement of business indicators. Effective and objective performance measurement at an individual level therefore also depends on measurable indicators with targets that are agreed upon. The following paragraphs discuss Balanced Scorecard and 24 Ways to identify typical measurement frameworks that need to be supported by BI systems.

Balanced Scorecard

During the early 1990s Kaplan and Norton formalised their ideas on performance measurement in a number of Harvard Business Review articles and followed it up with a book on the balanced scorecard (BSC), sub titled "Translating strategy into action" [8]. For many businesses this new framework facilitated a transformation from numerous ad hoc measurements, to an integrated and structured view of carefully selected indicators that support a specific strategy at a certain point in time.

The innovative aspects of the BSC approach include the following:

- Four perspectives were identified to ensure that not only financial indicators are taken into account – the others being the Customer, Internal/business processes and the Learning and Growth perspectives.
- Cause-and-effect relationships should be established between objectives and measurements, including all perspectives. A strategy is a set of hypotheses about cause and effect. See Figure 2 for an example.
- A healthy mix between outcome measures and performance drivers should be part of the BSC. Performance drivers are seen as leading indicators and outcome measures as lagging indicators. "Outcome measures without performance drivers do not communicate how the outcomes are to be achieved. Conversely, performance drivers … without outcome measures may enable the business unit to achieve short-term operational improvements, but will fail to reveal whether the operational improvements have been translated into expanded business … and, eventually, to enhanced financial performance." (Kaplan and Norton [8])
- The link between strategy and operational measures is emphasised to ensure that all business units are aligned with the strategy and that unnecessary effort is spent on irrelevant measurements that may distract the focus away from the desirable behaviour.
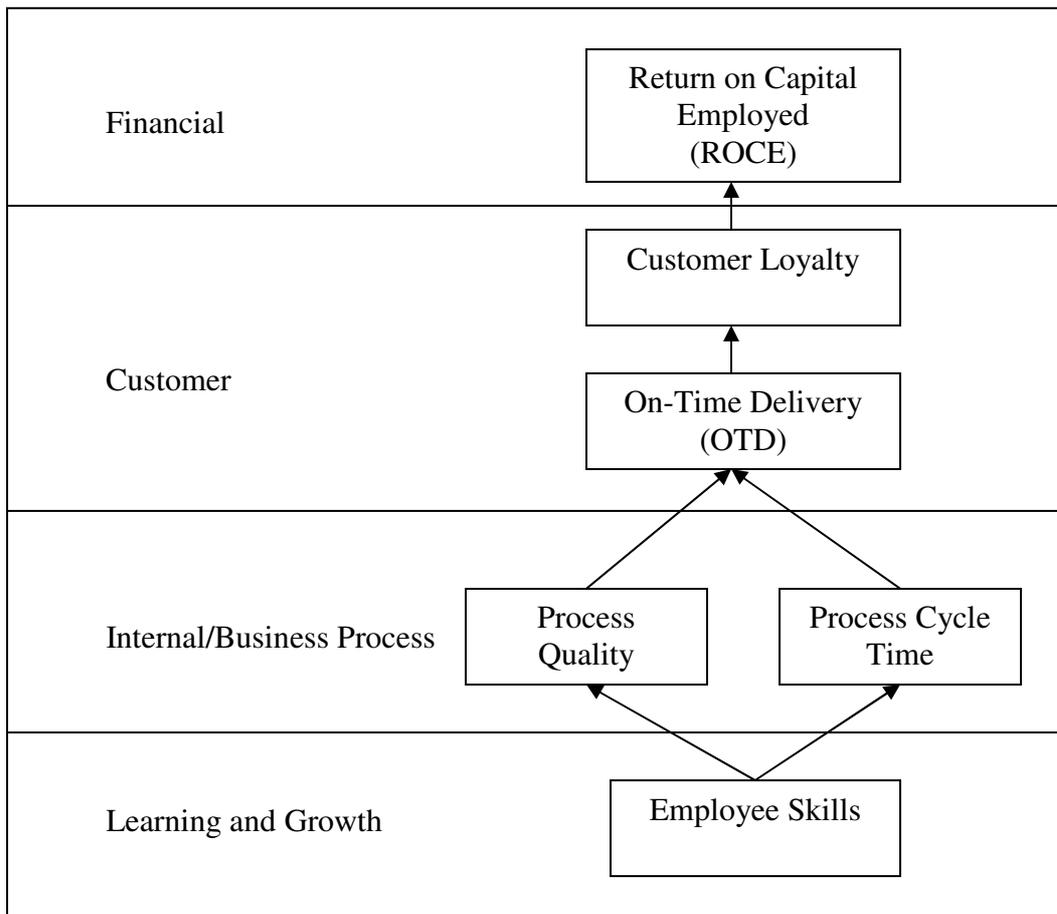
Figure 2: Cause-and-effect example (Kaplan and Norton [8])

The idea of cause-and-effect is obviously not new – Eli Goldratt [9] has based his whole theory of constraints (TOC) on this logical approach and industrial engineers often use TOC to identify bottlenecks and to identify strategies to eliminate the root cause of undesirable effects. However, what is often lacking in the future reality trees that are developed during the TOC process is the quantification of the effects of certain actions and the ultimate effect on the financial status of the organization.

With the BSC approach a framework is provided to guide the developers of a future reality tree in the identification of appropriate steps, taking all perspectives into account and showing the ultimate effect in financial terms. Connelly et al. [4] also stress the fact that in a business environment all actions taken should be traced back to the effect on the financial statement by referring each of the identified 24 Ways to the potential effect they might have on specific parts of the financial statements.

24 Ways approach

Whilst TOC provides a powerful means to identify root causes of undesirable effects and the BSC provides a useful framework to ensure that all perspectives in the business are taken into account when developing a strategy to improve the situation, or to get rid of the root cause of undesirable effects, exactly what to measure, or

which strategy to follow, is seldom specifically addressed in literature. For each business with its own circumstances and challenges at a specific point in time. Connelly et al. [4] suggest 24 Ways, which covers a variety of information "sweet spots" that should be addressed in a typical manufacturing organization that wishes to excel in using new business models where the emphasis moves from products and revenue to *customer-* and *profit-centric* organizations. Information "sweet spots" are defined as a relatively small number of positions in the information flow through an organization that contains the most valuable information for corporate decision-making. When these sweet spots are analysed by multidimensional managers with powerful BI tools, answering the relevant questions of WHEN, WHO, WHAT, WHERE and with WHICH RESULT (in all potential combinations) becomes easier, resulting in better decision-making.

Even though the 24 Ways are identified for the most general model of business, namely manufacturing, where it is strongly identified with the flow of products and services across the supply chain, the underlying business issues also apply to any corporation or governmental organization. The 24 Ways are grouped into eight areas that are normally organized into separate departments. They are briefly discussed in the following paragraphs and mapped to the BSC perspectives as identified by Kaplan and Norton. According to Connelly et al. [4], there is no specific sequence in which the 24 Ways must be addressed and all of them might not be applicable to all businesses.

Finance:

1) Multidimensional income statement
2) Profit drill-down analysis
3) Multidimensional balance sheet
4) Key financial ratios
5) Cash flow analysis

The first five Ways are mainly related to financial information and can strongly be identified with the **Financial** perspective of the BSC. Typical dimensions include Time period (Years, quarters and months), Organisational departments, Income statement lines and Balance sheet lines.

Sales:

6) Sales analysis
7) Customer and product profitability
8) Sales plan vs. forecast
9) Sales pipeline

The sales related Ways are shared between the BSC perspectives of **Finance** and **Internal business processes** (of marketing and sales) Additional dimensions include Product, Customer and Sale Type (e.g. New, Base, and Win back).

Marketing:

10)   Strategic marketing analysis
11)   Tactical marketing analysis

The two marketing Ways are associated with the BSC perspective of **Internal business processes** (of marketing). Additional dimensions could include Marketing channel, Marketing campaigns, Market segment and in certain cases, Product attributes (e.g. Colour or Material).

Purchasing:

12)   Inventory turnover
13)   Supplier scorecard

These Ways are associated with the BSC perspective of **Internal business processes** (of Purchasing). Additional dimensions include Supplier, Terms, Just-In-Time (JIT) performance and Inventory location.

Production:

14)   Capacity management
15)   Standard product cost and quality
16)   Cause of poor quality

The production related Ways are associated with the BSC perspective of **Internal business processes** (of Production scheduling and Quality Assurance), as well as **Finance**. Additional dimensions include Work stage (e.g. Set-up, Assembly, Inspection, Packaging), Production run and Reject reasons.

Distribution:

17)   Carrier scorecard

This logistic related Way is associated with the BSC perspective of **Internal business processes** (of distribution). Additional dimensions are Carrier, Destination, Distance category (e.g. Short, Medium, and Long) and Customer Type (e.g. JIT, Buy and Hold).

Customer service:

18)   On-Time delivery
19)   Complaints, returns & claims
20)   Cost of service relationship

The Ways are associated with the **Customer** perspective of the BSC. Additional dimensions include Lead time categories (e.g. greater than 30 days, 6-30 days, 1-5

days, same day), Percentage late categories (On Time/Early, 1-2 days late, 3-7 days late, more than 7 days late) and Reasons for complaints/returns/claims.

Human resources (HR)/Information Technology (IT):

21)   HR administration
22)   Core competence inventory
23)   BI deployment
24)   24 Ways return on investment (ROI)

These Ways are associated with the **Learning and Growth** perspective of the BSC. Additional dimensions include Job group, Salary grade, Status (e.g. Full time, Part time, Temporary), Length of service category, Performance, Core skill and Skill rating.

6. THE QUESTION OF QUALITY IN INFORMATION DELIVERY

With all the guidelines and methodologies at hand and various hints on what to measure, setting up an effective data warehouse with useful business intelligence should be straightforward. In theory it may be so, but not so in practice. Too often organisations go through the motions of establishing a data warehouse and then realise that the underlying data quality (and the resultant information quality) leaves much to be desired. An on-site survey conducted by the META group at the Data Warehousing Conference in Florida in 1996, put data quality issues in the two top places on a list of 16 challenging data warehousing issues – and the situation has not changed much since then. According to Larry English [10], the situation has worsened since then.

In a discussion of the maturing of the Information age, English [10] indicates that "the notion of data as a by-product of business processes is giving way to the eureka of information as a direct product that has value beyond its immediate processes." Seen in this context, information has processes that create and maintain it, it has suppliers and customers and, unfortunately, it is also subject to quality problems. In many instances the process of data warehousing exposes information quality problems that only become apparent when an organisation attempts to integrate disparate data.

English [11] defines information quality as "quality in all characteristics of information, such as completeness, accuracy, timeliness, clarity of presentation that consistently meets knowledge worker and end-customer expectations to meet their objectives". A distinction is also made between inherent characteristics (validity, accuracy, conforming to business rules) and pragmatic quality (usefulness, timeliness, presentation clarity). It is important to understand these characteristics, since ensuring total quality a whole range of processes and techniques are required. For example, a compulsory field with a lookup list in the transactional system to ensure that a value is valid within a fixed, acceptable range of values may enforce data validity. However, this does not prevent the operator from selecting a wrong value with an inaccurate result in the record whatever the reason may be.

In a facilities management environment the following situation was observed. Work orders have to be linked to a valid building within the portfolio before a maintenance job card can be generated. Since many maintenance jobs were delayed or linked to a wrong building, system administrators have also put a "NANA" (Not Available / Not Applicable) value in the lookup list and soon a large number of work orders were linked to the NANA building, since it was the quickest way to get through the process.

The above example also illustrates another principle. When judging the quality of data, it is not a question of whether it fits **a specific purpose** (in this case to generate a job card and get the maintenance process started), but whether it fits **all purposes** – also (among others) the analytic purpose of allocating cost per building. It is therefore important that the information producers – the people who capture data at various points in the business processes – know who all the information customers are, what their quality requirements are, how the data is used and what the cost of non-quality data is.

To explain what the characteristics of information quality means to the business or knowledge worker, English [10] describes it in the following manner:

QUALITY CHARACTERISTIC   **KNOWLEDGE WORKER BENEFIT**

The *right* data   The data I *need*
With the right *completeness*   *All* the data I need
In the right *context*   Whose *meaning* I know
With the right *accuracy*   I can *trust* and rely on it
In the right *format*   I can *use* it *easily*
At the right *time*   *When* I need it
At the right *place*   *Where* I need it
For the right *purpose*   *I can accomplish our objectives and*
                                        satisfy our customers

English [10] also puts the relationship between data, information, knowledge and wisdom into context by defining it as follows:

- Simply stated, **data** are the representation of facts about things. Data are only the raw material from which information may be produced.

- **Information** is data in context. Information quality requires quality of three components: clear definition or meaning of data, correct value(s), and understandable presentation (the format presented to a knowledge worker).
  Information = f(Data + Definition + Presentation)

- **Knowledge** is not just information known - it is information in context. Knowledge means understanding the significance of the information. Knowledge is applied information and may be represented as a formula:
  Knowledge = f(People + Information + Significance)

- **Wisdom** is applied knowledge and may be expressed by the formula:
Wisdom = f(People + Knowledge + Action)

The business value of quality information therefore lies in the actions that knowledge workers can take to improve the enterprise with information of significance. It is true, however, that the basic component in this value chain of information at different levels, will always remain data.

Wai et al. [12] have identified the following as some of the main causes for so-called dirty data:

- Misuse of abbreviations
- Data entry mistakes
- Duplicated records
- Spelling variations
- Unit differences
- Outdated codes

The 16th International Conference on Data Engineering [13] has also identified the absence of universal keys across different databases (also known as the object identity problem) and the presence of inconsistencies in data coming from multiple sources, as further reasons for data quality problems.

There are clearly two fronts on which the battle against poor information quality is waged:

- Cleaning dirty data from transactional systems as it enters the DW environment (using the staging area as described earlier).
- Building an information quality organisation – a longer term solution that is focused on prevention and addressing the root causes of dirty data, rather than merely fixing it. In an information quality organisation the whole enterprise is sensitive to the quality of information and the emphasis is on the improvement of operational processes to ensure quality. Processes are also defined to measure and improve data quality and to measure cost savings and increased customer satisfaction.

English [14] has identified four types of data quality tools that assist the data cleaning and data quality improvement processes, but points out that these tools can never fix the broken or incomplete processes that generated the dirty data in the first place. The four types of tools are:

- **Rule discovery** products which analyze data in legacy files and databases to discover data relationships and rules describing how the data is currently used.
- **Data quality analysis/audit** products that evaluate data against a set of defined business rules. These tools can check the validity of data, for example marital status or birth date, but it can never ensure accuracy.
- **Data cleansing/scrubbing** products that analyze and standardize data, identify potential duplicates and transform data to correct (or "probably" correct) values.

Typical applications are to standardize spelling of towns/cities, ensuring that the postal code is correct for the given street address, and transforming date formats from different operational systems to a common format. Once again, these tools cannot ensure accuracy of data.

- **Data defect prevention** products that have routines that can be invoked by data capturing applications, or that test input data in real time. These tools enforce integrity rules at the source of data creation – the ideal situation – and may include the generation of check digits to prevent typing errors.

It is clear that these tools can only play a limited role in ensuring information quality. A more holistic approach is needed – something similar to what Total Quality Management (TQM) has offered for manufacturing companies in the early 1980s. It also needs a good dose of KAIZAN (continuous improvement involving everyone in the organization). Imai [15] explained the process of continuous improvement to be: "Whenever problems occur, management must analyze them, identify the root cause, and revise the standards and improve the processes to prevent recurrence."

7. TOTAL INFORMATION QUALITY MANAGEMENT (TIQM)

In September 2003 English [16] published the TIQM methodology, which is a much more detailed and enhanced version of the TQdM (Total Quality data Management) process that was developed in 1996. The methodology is largely based on pure quality management principles that were developed by W Edwards Deming, Joseph Juran and Phillip Crosby and made it applicable to information quality.

Accordingly TIQM is not a program, it is rather a value system of caring for information customers; it is a mind-set of excellence in all products and services (including information products); it is a habit of continuous improvement – not only of application and data development processes, but also of the business processes by integrating quality management beliefs, principles and methods into the enterprise culture.

The TIQM methodology consists of six processes and 14 points of information quality. The 14 points are based on the 14 points of quality that Deming published in 1982 in his book "Out of the Crisis". [17] According to Cohen [18], a number of the points can be traced back to Deming's thoughts (as a statistician) about variation. "Variation was seen by Deming as the disease that threatened US manufacturing. The more variation - in the length of parts supposed to be uniform, in delivery times, in prices, in work practices - the more waste, he reasoned." Although Deming does not mention the concept of TQM (Total Quality Management) in his book he is widely accredited with those developments in the early 1980s.

Cohen [18] has summarized Deming's 14 points in the following manner:

1. Create constancy of purpose towards improvement. Replace short-term reaction with long-term planning.
2. Adopt the new philosophy. The implication is that management should actually adopt his philosophy, rather than merely expect the workforce to do so.

3. Cease dependence on inspection. If variation is reduced, there is no need to inspect manufactured items for defects, because there won't be any.
4. Move towards a single supplier for any one item. Multiple suppliers mean variation between feed stocks.
5. Improve constantly and forever. Constantly strive to reduce variation.
6. Institute training on the job. If people are inadequately trained, they will not all work the same way, and this will introduce variation.
7. Institute leadership. This makes a distinction between leadership and mere supervision. The latter is quota- and target based.
8. Drive out fear. Management by fear is counter productive in the long term, since it prevents workers from acting in the organisation's best interests.
9. Break down barriers between departments. Another idea central to TQM is the concept of the "'internal customer", that each department not only serves management, but the other departments that use its outputs.
10. Eliminate slogans. Another central TQM idea is that it's not people who make most mistakes - it's the process they are working within. Harassing the workforce without improving the processes they use is counter productive.
11. Eliminate management by objectives. Deming saw production targets as encouraging the delivery of poor quality goods.
12. Remove barriers to pride of workmanship. Many of the other problems outlined reduce worker satisfaction.
13. Institute education and self-improvement.
14. The transformation is everyone's job.

English [16] has taken these 14 points and adapted them to be applicable to Information Quality. The 14 points translate to the following:

1. Create constancy of purpose for improvement of information products and services. Must solve problems of today and problems of tomorrow; the obligation to the knowledge worker (information customer) never ceases.
2. Adopt the new philosophy of quality shared information as a tool for business excellence. It means transformation of information system management and business management.
3. Cease dependence on inspection to achieve information quality. Design quality in to data and application design and to business (information production) processes.
4. Develop common data create programs and invest and develop trust in information producers.
5. Improve constantly and forever the processes of application and data development and service, and information production and maintenance. This will result in continual reduction of costs of information scrap and rework and increase in opportunity gain.
6. Institute training on information quality for all employees, especially management and information producers.
7. Institute leadership for information quality. Management must assume accountability for information quality.

8. Drive out fear of data uncertainty or data correction. Create non-blame, non-judgmental environment where everybody is encouraged to improvement by eliminating root causes without blame.

9. Break down barriers between staff areas – work as partners in teams, including information management and application development, IT and business, business area and business area.

10. Eliminate slogans and … replace with actions for information quality improvement.

11. Eliminate quotas of "productivity" for information producers and management that increase errors and cost of information scrap and rework. Create a balanced scorecard that includes end-customer, internal information producer and knowledge worker satisfaction.

12. Empower information producers to fix the problems in the process for they know the problems of their processes and given the opportunity, will fix them. Develop a habit of information "defect prevention" as part of everyone's job.

13. Every knowledge worker must understand the paradigm shift and learn tomorrow's skills. Provide education in information-age principles, value chain management and process improvement.

14. Senior management must organise itself to make IQ happen, they must feel the pain of the status quo, must communicate to people why change is necessary and implement a Plan-Do-Check-Act process for IQ improvement.

The six processes identified by English [16] to help accomplish the 14 points are:

- Assess data definition and information architecture quality
- Assess information quality
- Measure non-quality costs and risks
- Data reengineering / Correction process
- Improve information process quality
- Establish an information quality environment

These processes are integrated as shown in Figure 3. All of the processes contain typical industrial engineering activities, such as:

- The extraction of statistically valid data from sources and to measure the information quality (supports P1 and P2).
- The identification of business performance measures and the calculation of information costs and risks of non quality data (supports P3).
- Standardize data and analyze defect data types (supports P4).
- Develop and implement Plan-Do-Check-Act cycles to improve processes of information producers (supports P5).
- Establish quality assurance procedures for the whole enterprise by building associations between products, processes, people and supporting systems and technology (supports P6).
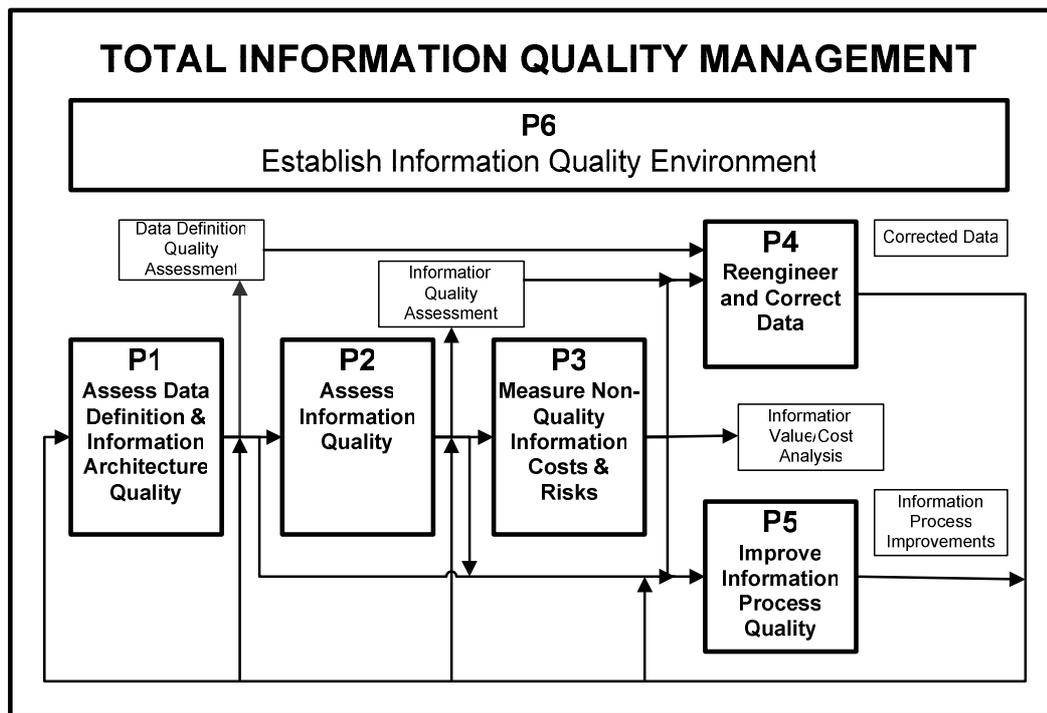
Figure 3: The six TIQM processes identified by English [16]

8. CONCLUSION

Optimising BI as a competitive resource and the way in which it is derived and presented in modern businesses is one of the most important tasks which industrial engineers are currently involved in. Just as TQM in the early 1980s helped manufacturing enterprises to produce quality products and to reduce scrap and rework by improving the operational processes, a similar approach is needed to ensure information quality that will eventually lead to better business decision making and ultimately to better financial results.

Philip Crosby [19] commented that: "Quality is free. It's not a gift, but it is free. What costs money are the "unquality" things – all the actions that involve not doing jobs right the first time. Every penny you don't spend on doing things wrong, over, or instead, becomes half a penny right on the bottom line."

The role of the industrial engineer is to understand the larger business picture and to play an active role in certain parts of the picture to ensure that the principles of TQM that have been successfully implemented in manufacturing environments since the 1980s, are also incorporated in the way in which information quality is handled. The scope of involvement in the bigger picture includes the following:

- Link business strategy to integrated business processes.
- Design business processes to enable information producers to provide quality data.

- Design and improve supporting transactional information systems to capture valid data and to design quality processes that will also ensure accuracy of the data, not only for a certain user groups, but also for all knowledge worker purposes.
- Identification and clear definition of performance measures that can effectively be used in a balanced scorecard approach that will give insight into more perspectives than just the financial view of an enterprise.
- Design and maintain data warehouses according to well-established methodologies that will evolve over time into the single version of the truth on which the whole enterprise relies for its strategic and tactical decisions. Not because of magical data cleaning during the staging process, but because of information quality that is built into the processes where the data is captured.
- Present the information to the right people, in the right format, at the right time, including involvement on the selection, or design, of BI tools that may vary from basic query tools to sophisticated analytical tools for different types of users. This function also includes evaluation of the cost effectiveness of the proposed solution in terms of licence fees (e.g. named user, CPU (central processing unit) licence, or concurrent users), in an era where many options exist from the IT vendors.
- Manage change processes in an ever changing business environment where a single decision by a business person may have far reaching effects on business processes, information systems, data warehousing processes, performance measurement and many more.

When industrial engineers apply their traditional quality assurance tools and thought processes to the subject of data quality throughout the enterprise, they are making a significant contribution to improved business decision making, which is becoming more dependent on business intelligence systems that are based on well designed and maintained data warehouses.

9. REFERENCES

[1]   **Sperotto, F.,** 1994, In the footsteps of Homo Industrialis, Picsie Press.

[2]   **Kimball, R., Ross**, M., 2002, The Data Warehouse toolkit, 2nd Edition, Wiley.

[3]   **Frenzel, Carroll W.,** 1999, Management of Information Technology, 3rd edition, Course Technology.

[4]   **Connelly, Richard, McNeill, Robin and Mosimann, Roland,** 1999, The Multidimensional Manager, Cognos Incorporated.

[5]   **Eckerson, Wayne,** 2003, Smart companies in the 21st century, The Data Warehouse Institute.

[6]   **Kimball, Ralph, Reeves, L., Ross M. and Thornthwaite, W**, 1998, The Data Warehouse Lifecycle Toolkit, Wiley.

[7]   **Inmon, W.H., Imhoff, C., Sousa, R.**, 2001, Corporate Information Factory, 2nd edition, Wiley.

[8]   **Kaplan, Robert S., Norton, David P.,** 1996, The Balanced Scorecard, Harvard Business School Press.

[9]   **Goldratt, E. M**., 1992, The Goal: A Process of Continuous Improvement, North River Press.

[10] **English, L.P.,** 1999, Improving data warehouse and business information quality: methods for reducing costs and increasing profits, Wiley.

[11] **English, L.P.,** 1999, Seven DEADLY Misconceptions about Information Quality, Information Impact International Inc.

[12] **Wai L.L., Mong L.L., Tok W.L.,** 2001, A Knowledge-based approach for duplicate elimination in data cleaning, Information Systems 2001 vol. 26, National University of Singapore.

[13] **International Conference on Data Engineering**, Proceedings of the 16th 2001, Edited by Galhardas, H., Florescuand, D., Simon, E. Shasha, D. New York University.

[14] **English, L.P**., 1996, Data Quality in the Data Warehouse: a Key Critical Success Factor, Information Impact International Inc.

[15] **Imai, M.,** 1997, Gemba Kaizan: a Commonsense, Low-Cost Approach to Management, McGraw-Hill.

[16] **English, L.P.,** 2003, Total Information Quality Management: A complete methodology for IQ Management, Information Impact International Inc.

[17] Deming, W.E., 1982, Out of the Crisis, SPC Press.

[18] **Cohen, P.,** 1998, Deming's 14 Points, HCi Consulting.

[19] **Crosby, P.B**., 1979, Quality Is Free, Penguin Group.