

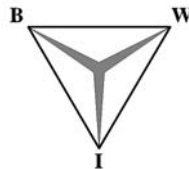


# ***Business Intelligence***



## **The Umbrella Term**

### **BWI-WERKSTUK**



***Deborah Quarles van Ufford***  
***November 2002***

*vrije* Universiteit Amsterdam  
Faculteit der Exacte Wetenschappen  
Studierichting Bedrijfskunde & Informatica  
De Boelelaan 1081a  
1081 HV Amsterdam



## **Preface**

The report that lies before you is the final report of a field exploration. This exploration is an important and compulsory element of Business Mathematics & Computer Science, an education that aims to combine the fields of Economics, Mathematics and Computer Science (or: Informatics). The goal is to take a topic related to at least two of these three fields, and investigate the existing literature on this topic. The student is free to choose the subject, but one can also choose to take on a subject presented by a member of the scientific staff.

In this case, the subject “Business Intelligence” was brought up by Prof. Dr. A.E. Eiben of the Faculty of Sciences, who is preparing to teach this as a new course. The objective of the assignment, therefore, was to review the current status of the field of Business Intelligence, to identify trends, key publications (papers and/or books) and commercial vendors of BI systems.

As mentioned above, at least two of the three fields of Economics, Mathematics and Computer Science have to be discussed in the literature study. As Business Intelligence is clearly a topic related to Computer Science and Economics, these two fields are dealt with here. The field of Mathematics could be used to go into detail about techniques used to analyze certain techniques used in Business Intelligence. However, seeing as the aim of this study is to investigate existing literature, I consider mathematical analyses to lie outside the scope of my research.

This report is meant as a reference work for the members of the Free University of Amsterdam, Faculty of Sciences, and for future students who are in the process of writing their literature study.

I would like to express my thanks to Professor Eiben who supervised me in my exploration and writing a report about it. Also I would like to thank him for introducing me to the world of Business Intelligence, a most interesting field, and for letting me use some of the pictures he designed for his course Business Intelligence. A word of thanks also goes out to Wojtek Kowalczyk, the second reader of this report.

Deborah Quarles van Ufford

Utrecht, november 2002



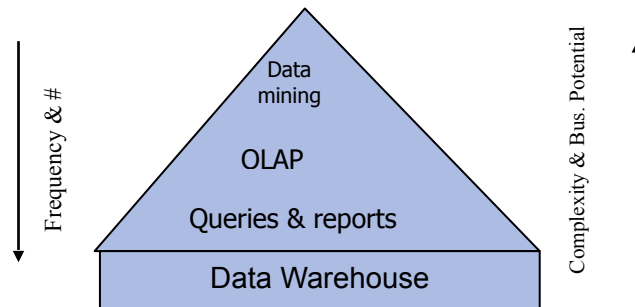


## Executive Summary

*Business Intelligence (BI) is a broad category of applications and technologies for gathering, storing, analyzing, and providing access to data to help enterprise users make better business decisions.*

[whatis.com, 2001]

Business Intelligence (BI) can be seen as an umbrella that covers a whole range of concepts. BI can be approached roughly as being a Data Warehouse, with three layers on top of it: Queries & Reports, OnLine Analytical Processing and Data Mining (see the pyramid below). Authors and companies adopt this ordering widely. However, other orderings exist as well, with the result that some contradict each other. This is simply because the boundaries between the different components are very vague.



A Data Warehouse consists of one or more copies of transaction and/or non-transaction data that have been transformed in such a way that they are suitable for querying, reporting and other data analysis. It forms the basis “on top of which” further analyses can be carried out.

The first level of analysis is Querying & Reporting. Querying means using a computer language to obtain immediate, online answers to user questions. Reporting refers to creating standard, point-in-time reports or generating reports by describing specific report components and features.

A level higher we have OnLine Analytical Processing (OLAP). This technology allows users to carry out complex data analyses with the help of a quick and interactive access to the information in data warehouses from different viewpoints. These different viewpoints are an important characteristic of OLAP, also called multidimensionality. The dimensions within the OLAP application usually reflect the different dimensions of an organization. A definition of OLAP that is adopted across the whole world is the one by Nigel Pendse: *Fast Analysis of Shared Multidimensional Information (FASMI)*. An advanced tool that uses the OLAP-methodology is the Balanced Scorecard.

The top layer is Data Mining. A simple definition is: analyzing and finding patterns in large amounts of data in order to support decision making and predict future behavior. Because Data Mining is such an advance technique, the process not only involves applying tools to a collection of data, but it starts with business understanding, data understanding and preparation, and selecting the right modeling techniques, and ends with evaluation and deployment.

The information and knowledge that is “dug up” by data mining can also be used to provide information about a web site and its visitors: Web Mining. When engaged in e-commerce activities it is the ‘invisible’ and ‘not-straightforward’ information that is most valuable, information hidden in the gigabytes of data generated each day that describe actions made by every visitor to the site.



With BI-tools it is possible to carry out analyses and reports on virtually all thinkable aspects of the underlying business, as long as the data about this business come in large amounts and are stored in a Data Warehouse. Departments that are known to benefit most from Business Intelligence are (Database) Marketing, Sales, Finance, ICT (especially the Web) and the higher Management.

Because the ICT-hype that we have been experiencing the last few years is decreasing, I do not expect much development to take place on the short term. On the longer term however, the World Wide Web will keep on growing, and with it the wish to keep storing data and information in structured ways, in order to gain as much benefit from the extracted knowledge as possible. In the area of Data Mining especially concepts like Customer Profiling will stay popular, because in the end it will always be rewarding to keep on knowing who your most profitable customers are.



Contents

BUSINESS INTELLIGENCE .....1  
*The Umbrella Term* ..... 1

**BWI-WERKSTUK**.....1

**PREFACE**.....2

**EXECUTIVE SUMMARY** .....4

**CONTENTS** .....6

**RESEARCH METHODOLOGY** .....9

**INTRODUCTION** .....11

    STRUCTURE OF THIS REPORT .....11

**1. BUSINESS INTELLIGENCE – THE UMBRELLA TERM** .....13

    1. THE “BIRTH” OF BI.....13

    2. WHAT IS BUSINESS INTELLIGENCE?.....13

    3. HISTORY OF BI.....14

    4. CUSTOMER RELATIONSHIP MANAGEMENT .....15

**2. DATA WAREHOUSING** .....17

    1. WHAT IS A DATA WAREHOUSE? .....17

    2. THE ‘INVENTION’ OF THE DATA WAREHOUSE.....18

    3. EXTRACTION, TRANSFORMATION AND LOADING .....18

    4. INFORMATION SOURCES .....19

        4.1. *Key publications* .....19

        4.2. *Commercial Vendors*.....20

**3. QUERIES & REPORTS**.....21

    1. WHAT ARE QUERIES AND REPORTS? .....21

    2. INFORMATION SOURCES .....22

        2.1. *Key publications* .....22

        2.2. *Commercial Vendors*.....22

**4. OLAP**.....23

    4.1.WHAT IS OLAP? .....23

    4.2.AN OLAP EXAMPLE.....24

    4.3.FASMI .....25

    4.4.OLAP APPLICATIONS .....26

    4.5.BALANCED SCORECARD.....27

    4.6.INFORMATION SOURCES.....28

        4.6.1. *Key publications* .....28

        4.6.2. *Commercial Vendors*.....29

**5. DATA MINING** .....31

    5.1.WHAT IS DATA MINING? .....31



5.2.THE DATA MINING PROCESS.....	32
5.3.DATA MINING TECHNIQUES .....	34
5.4.WEB MINING: THE INTERNET-VARIANT OF MINING.....	34
5.5.INFORMATION SOURCES.....	36
5.5.1. <i>Key publications</i> .....	36
5.5.2. <i>Commercial Vendors</i> .....	36
<b>6. BUSINESS INTELLIGENCE AGAIN .....</b>	<b>38</b>
6.1.WHAT IS BUSINESS INTELLIGENCE?.....	38
6.2.BUSINESS INTELLIGENCE VS. DECISION SUPPORT SYSTEMS.....	38
6.3.CURRENT STATUS .....	39
6.4.APPLICATION AREAS .....	40
6.5.RETURN ON INVESTMENT FOR BI PROJECTS .....	41
6.6.COMPETITIVE INTELLIGENCE .....	41
6.7.INFORMATION SOURCES.....	42
6.7.1. <i>Key publications</i> .....	42
6.7.2. <i>Commercial Vendors</i> .....	42
<b>CONCLUSIONS.....</b>	<b>44</b>
<b>APPENDIX A LINKS TO THE WORLD WIDE WEB.....</b>	<b>46</b>
PEOPLE’S HOME PAGES.....	46
COMPANY HOME PAGES.....	46
PAGES FOR FINDING JOURNALS AND MAGAZINES ON BI.....	46
CONFERENCE PROCEEDINGS.....	46
PAPERS AND ARTICLES .....	46
OTHER INFORMATION SITES .....	47
COMPETITIVE INTELLIGENCE .....	47
<b>APPENDIX B LIST OF ABBREVIATIONS .....</b>	<b>48</b>
<b>LITERATURE .....</b>	<b>51</b>
REFERENCE LIST .....	51
ADDITIONAL LITERATURE .....	53
FIGURE 1-1: THE PYRAMID OF BI .....	11
FIGURE 1-2: TRENDS AND INFLUENCES IN DATA WAREHOUSING, 1975-2000 .....	12
FIGURE 4-1 A 3-DIMENSIONAL OLAP CUBE .....	22
FIGURES 4-2: THE OLAP CUBE LOOKED AT FROM 3 DIFFERENT DIMENSIONS .....	23
FIGURE 4-3: THE 3 DIMENSIONS COMBINED IN THE OLAP CUBE .....	23
FIGURE 5-1: THE DATA MINING PROCESS .....	30
FIGURE 6-1: THE PYRAMID OF BI .....	35
FIGURE 6-2: KNOWLEDGE VALUE VERSUS USER EXPERTISE .....	37
TABLE 21: DATA WAREHOUSING BOOKS BY RALPH KIMBALL .....	17
TABLE 22: DATA WAREHOUSING BOOK BY BILL INMON .....	17
TABLE 23: VENDORS OF DATA WAREHOUSING .....	18
TABLE 31: VENDORS OF REPORT AND QUERY TOOLS .....	20



TABLE 3-2: BI VENDOR DIRECTORY (QUERIES AND REPORTS) .....20

TABLE 4-1: LESSONS TO BE LEARNT FROM A 35-YEAR HISTORY OF OLAP.....24

TABLE 4-2: OLAP APPLICATION AREAS .....25

TABLE 4-3: VENDORS OF OLAP AND MULTIDIMENSIONAL DATABASE TOOLS .....27

TABLE 4-4: BI VENDOR DIRECTORY (OLAP AND OLAP PACKAGES) .....27

TABLE 5-1: THE STEPS OF THE DATA MINING PROCESS .....31

TABLE 5-2: BI VENDOR DIRECTORY (DATA MINING) .....33

TABLE 5-3: VENDORS OF DATA MINING AND RELATED CONCEPTS .....34

TABLE 6-1: BI VS. DSS DEFINITION .....36

TABLE 6-2: SOME HYPERLINKS TO COMPANIES' WHITEPAPERS .....39

TABLE 6-3: MOST FREQUENTLY ENCOUNTERED BI-VENDORS, AS OF 2002 .....40

TABLE B-1: LIST OF ABBREVIATIONS .....45





## **Research Methodology**

On hearing the term “literature study” one usually thinks of reading books, journals and magazines, and conference proceedings as a main source of information. It is not surprising, however, that with the increasing number of sites on the World Wide Web, a large amount of information can nowadays also be found on the Internet.

What I found striking was that, even though the term Business Intelligence is about ten years old, not many books on Information Systems and such topics include the term Business Intelligence in the index or glossary.

There are more magazines than scientific journals, that have published articles about BI, but unfortunately they are hard to come by unless you are a subscriber or have access to libraries that possess the volumes you need. A magazine of valuable importance happened to be the Database Magazine (DB/M). The library of the Free University of Amsterdam does not possess these, but fortunately students are allowed to visit libraries of other universities to obtain and copy the necessary articles.

Very few magazine articles are published as a whole on the Internet. Nevertheless, the Internet proved to be the most important, and also the easiest, information source in this study.

The main search method used to search the Internet was the search engine Google. Because the author of this paper has a thorough command of only the Dutch and the English language, the scope of the results is restricted to Dutch- and English-language websites, articles, companies, etc.

In each chapter of this paper a section is included that covers useful information sources that can be used to investigate a particular topic. The sources that were used to investigate the topics in the report are listed in the Reference List; throughout the report references are made to this list where necessary.





## Introduction

### UnoVu Inc.<sup>1</sup> needs a System

One day, the CEO, the CIO and the CKO of the organization UnoVu Inc. got together for a meeting to discuss the status of their company. The OAS's that UnoVu had always worked with were out of date. It was clear UnoVu needed a System to help them in their CRM and PRM, but what System should they choose? The CEO suggested a DSS, but the CIO opted for an EIS. However, the CKO rather had a KMS or a KWS, with tools such as KAT and KDD. One thing they all agreed upon was that the System had to be some kind of MIS. An ASP to carry out some good ERP and MRP. The System would support the managing of UnoVu, therefore it could also be an MSS. In the line of support, the CEO's thoughts then drifted from a DSS to a GSS, and the CIO's view turned around from the EIS to an ESS. Or wasn't there also another type of EIS? The CKO's final opinion was that it had to be an ES, or more specifically an ESS.

The CEO, CIO and CKO came to the conclusion that they would not reach a decision in this way, so they decided to consult a book on Systems. To their utmost surprise, not only did they encounter all of the systems that had been passed in review during their meeting, but also a CSS, a DES, an EMS, a DMS, an ETS, a GIS, an ITS and so forth.....

<sup>1</sup> UnoVu Inc. is a fictive name.

Perhaps you have stopped reading this little fictive story after the first few lines. If so, I completely understand! What I am in fact trying to make clear is that we are in the middle of an era in which the management of our business is dominated by 3-letter abbreviations. With little effort the above story can be continued including 4-, 5- and even 6-letter abbreviations like GDSS's, DDBMS's and OOMBMS's. But let's not get carried away. No doubt these last few were just introduced to make reading and writing of documents on these subjects a lot easier...

But seriously, if one opens and leafs through the average book about Information or Support Systems, one is bound to be confronted with no less than all of the above abbreviations. In most cases the 'S' stands for System. What do these Systems do? Basically, most of them try to make sense of a big load of data and subsequently provide the user with structured information that can support his/her decision making. The challenge of companies today is to turn the growing amounts of data into meaningful information and knowledge, in order to formulate actions that could lead to increasing profits. This is exactly where Business Intelligence has come peeping around the corner the past couple of years.

### ***Structure of this report***

This report is built up according to the components that make up Business Intelligence (hereafter to be called BI): a Data Warehouse, Querying & Reporting, OnLine Analytical Processing and Data Mining. In Chapter 1 we will see why specifically this ordering is adopted; it is a short introduction to BI. Chapter 2 is about Data Warehousing, the basis for all BI activities. Chapter 3 describes the most basic BI activities: Querying and Reporting. In Chapter 4 we go a step further and introduce OLAP, OnLine Analytical Processing. Chapter 5 handles the most sophisticated BI activity: Data Mining. After treating these components separately, in Chapter 6 we return to Business Intelligence in general. At the end the Conclusions of this investigation are presented.

Because this report is about an investigation of literature, I was afraid it would turn out rather dry. Therefore I included a short, fictive story about a department store called UnovVu Inc. at the beginning of the chapters 2 to 5, to illustrate in a more "playful" way what we are actually talking about.

The objective of this assignment was not only to review the status of the field. Each section in the report includes a section 'Information Sources', in which key publications and names, and commercial vendors are discussed. It proved to be unnecessarily time-consuming to write down whole lists of vendors in this report. Instead, I chose to present a few direct links to Web pages that contain these



lists. It is up to the reader to visit these pages and find the necessary companies that provide BI-products.

The Appendix contains a list of interesting links to sites on the World Wide Web, mostly of companies' general page and some direct links to other pages (as far as they are not written down in the Reference List).



# 1. Business Intelligence – The Umbrella Term

## 1. The “birth” of BI

In search for the year in which Business Intelligence was first introduced, I encountered two conflicting statements. Naeem Hashmi (2000) says that BI is a term introduced by Howard Dresner of Gartner Group in 1989, whilst Hans Dekker (2002) claims that Howard Dresner invented the term in 1992! It is clear they speak of the same term BI and the same Howard Dresner, but the supposed “birth-years” of BI are somewhat puzzling! Who is right?

As we know, most companies include a “Contact Us” page in their website. Fortunately, Gartner Group is one of them. And seeing as “nothing ventured is nothing gained”, I contacted the company and confronted them with the above facts. After having been forwarded several times to the concerning department, my message was replied to by Kevin Cooper, Client Inquiry Analyst:

*The term “Business Intelligence” was created in 1989 and coined by Gartner in that year. Howard Dresner had a hand in the creation of that term, but did not join Gartner until YE 1992, when he drove it into the mainstream.*

## 2. What is Business Intelligence?

Having got the indistinctness of BI’s “birth-year” out of the way, we can proceed with questioning: what is Business Intelligence?

This paper has the title “Business Intelligence: The Umbrella Term”. The reason for this is that many authors speak of BI as being an “umbrella term”, with various components “hanging under” this umbrella. Another way to look at it is the first explanation of Business Intelligence given to me by my supervisor, which is the following pyramid:

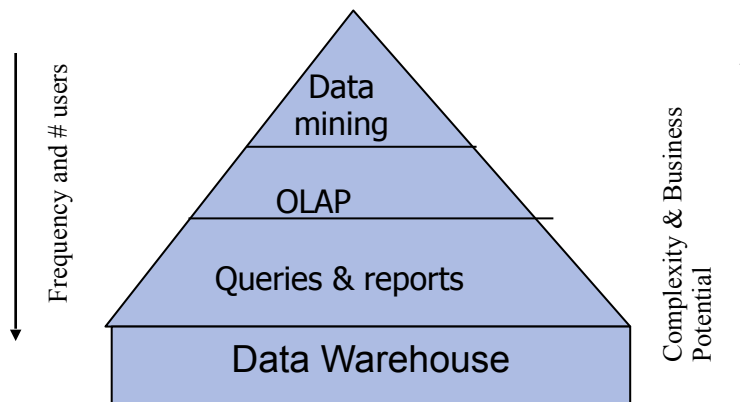


Figure 1-1: The pyramid of BI

What this simple picture tells us is that BI consists of various levels of analytical applications (and corresponding tools) that are carried out on top of a Data Warehouse. The lower you go in this hierarchy, the more frequently the tool is used and the more users it will have. Also, the more the extracted information is based on facts in figures. The higher you go in the hierarchy, the more complex the analyses taking place and the more business potential that lies in the resulting information and knowledge.



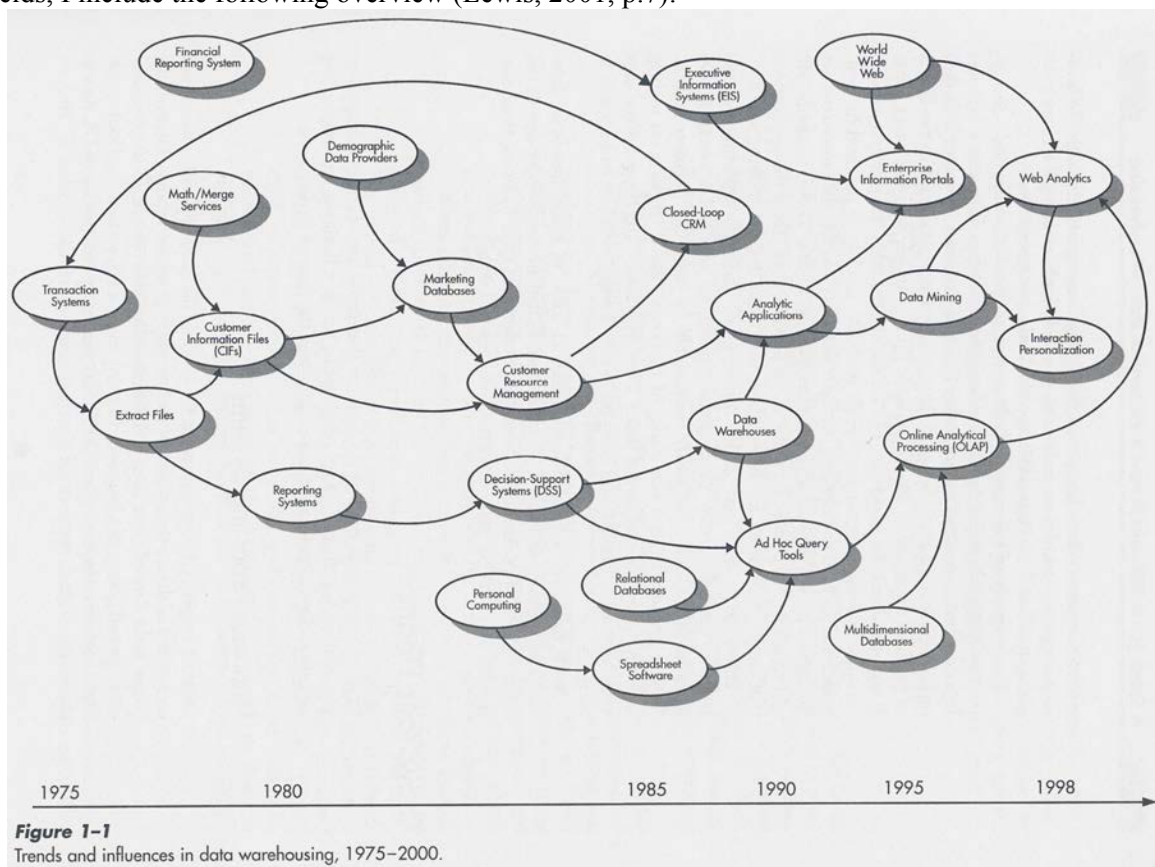
In researching what is written about the elements hanging under the umbrella, or contained in the pyramid of Business Intelligence, I came to the conclusion that the above ordering is one widely adopted. That is why I chose this ordering for my chapter layout. Also to meet the wishes of my supervisor in setting up this new course element, it seemed wise to follow this classification.

**3. History of BI**

Up to this point, we have agreed on Business Intelligence as being an umbrella that covers a whole range of concepts. It is clear that BI has somehow evolved from other concepts. Therefore, when exploring the history of Business Intelligence, it seems wise to take a look at what preceded Business Intelligence.

The problem with topics such as Business Intelligence, Decision Support Systems and many other acronyms with the ‘S’ standing for ‘System’ is that they are all part of a terribly volatile field. When I was halfway my education, around the year 2000, I was taught about Data Based Management Systems and I thought I was dealing with something hot, something new. Two years later, in 2002, I find out that a DBMS is out of date, compared to the systems, tools and techniques that have evolved over the past 10 to 15 years. But then why didn’t I learn of these newer tools and techniques two years ago?

Much has been written about Information and Support Systems, authors have filled tomes with describing the existing Systems: how do they work, how should they be built, what are the requirements, and so forth. Unfortunately, little to nothing is written on the history and development of the Systems. What you would have to do is take all these writings, lay them out next to each other and compare. I myself made an attempt to do this. However, with the one author saying that DSS were first seen in the 1980’s (Buytendijk, 2001) and the other defining the concept DSS as early as 1970 (Little 1970, according to Turban & Aronson, 2001), I decided this undertaking was too comprehensive for the purpose of this research. Nevertheless, to give the reader a few ideas on the development of these fields, I include the following overview (Lewis, 2001, p.7):



**Figure 1-1**  
Trends and influences in data warehousing, 1975-2000.

**Figure 1-2: Trends and influences in data warehousing, 1975-2000**



The information that is most volatile is that what we read on the Internet. Where up to about ten years ago authors wrote their findings down in books and journals, nowadays the easier, faster, cheaper and more accessible way of publishing is on the World Wide Web. The problem with this medium however, is that a web page has to be maintained and updated regularly to keep it and its topics alive. When this does not happen, pages get lost or wiped away or simply contain information that is out of date.

As mentioned in the section “Research Methodology”, the Database Magazine (also known as DB/M) proved a valuable source of information. DB/M is pinpointed as being a magazine you must not miss if you are interested in BI. Because it has been published since 1990, I figured it could at least give a slight idea of the history of Business Intelligence. However, it is only since 1997 that BI received the attention of the authors of DB/M.

#### **4. Customer Relationship Management**

Where companies used to be focused on delivering the right products to their customers, they are now focused on delivering their products to the right customers. The same goes for Business Intelligence applications. They used to be more of a ‘back-office’ tool, concentrated on reporting to the higher management of an organization. But with the shift from product to customer, we welcome Customer Relationship Management (commonly abbreviated as CRM). Within this framework of CRM, BI is no longer only used by management levels, but BI-tools and techniques are developed for all organizational levels.

At various points in the report we will see how Business Intelligence can influence CRM. To give a brief example up front, BI can be used to identify what is called ‘customer profitability’: which customer profiles are responsible for the highest profit? Based on the answer to this question, a company can choose to change their strategy and, for instance, make special offers to certain customer groups.







## 2. Data Warehousing

### UnoVu Inc. has to build a Data Warehouse

UnoVu is a gigantic chain of department stores with outlets in several countries. The UnoVu department stores sell products in many different categories, from food & beverages to clothing, from books & CD's to perfumes, from toiletries to garden products. There are various systems that keep record of the goings on in the warehouse. To name a few: each department has a Transaction Processing System that keeps track of all the transactions that take place. The sales data from the cash registers are used for managing the supplies. Next to this, each outlet has a Customer Database that holds all sorts of information about their customers: names, ages, addresses, other members of the family, profession, whether or not they have Frequent Buyer Pass.

What the Sales Department would like is to be able to analyze their business performance by creating sales reports, to compare results from different time spans, different countries, etc.

The Marketing Department would like to be able to make special offers to their customers and know before hand which customers could be more inclined to react than others.

But before all these wishes can be met, UnoVu must integrate the different systems that they have into one base: a Data Warehouse.

### 1. What is a Data Warehouse?

According to Simon & Schaffer (2001), there is no official definition of a data warehouse, that is, a standard definition supported by a standards committee such as the American National Standards Institute (ANSI).

Lewis (2001) writes that the most authoritative names in data warehousing define data warehouse as:

*A collection of integrated, subject-oriented databases designed to support the DSS function, where each unit of data is specific to some moment of time. The data warehouse contains atomic data and lightly summarized data.*

[Bill Inmon, Building the Data Warehouse]

*A copy of transaction data specifically structured for query and analysis.*

[Ralph Kimball, The Data Warehouse Toolkit]

Basically, a Data Warehouse consists of one or more copies of transaction and/or non-transaction data. These data have been transformed (a term further explained in section 2.3) in such a way that they are contained in a structure that is suitable for querying, reporting and other data analysis. One of the key features of a Data Warehouse is that it deals with very large volumes of data, in the range of terabytes. But this is not all. There are cases in which a Data Warehouse has to serve 100's to 1000's of users, process millions of daily records and carry out 1000's to 100.000's of daily queries and reports! Data rich industries have been the most typical users (consumer goods, retail, financial services and transport) for the obvious reason that they have large quantities of good quality internal and external data available (Pendse, August 2001).

In the opinion of most authors and companies, a data warehouse forms a base, on top of which tools like querying and reporting can be used to analyze business results. In particular, multi-dimensional warehouses allow more advanced techniques like OLAP and data mining to identify trends and make predictions. One definition that does not match this description and that I do not completely agree with is the following by Laudon & Laudon (2000):

*A data warehouse is a database, with reporting and query tools, that stores current and historical data extracted from various operational systems and consolidated for management reporting and analysis.*

The definition these authors give is incomplete. They fail to include the aspect of multi-dimensionality and with this the fields of OLAP and data mining. Also they include reporting and query tools in the concept of data warehouse, instead of placing them on top of the data warehouse. Finally, they ascribe the use of data warehousing to the management level, whereas most providers focus on business users at all levels when developing this kind of tools.



Of course a Data Warehouse does not come into existence out of nothing. A short description of this is given in the third section: Extraction, Information and Loading. In the last section of this chapter we will go into more detail on the two authoritative names in data warehousing, Inmon and Kimball.

When dealing with Data Warehousing, one could also come across the term “Data Mart”. Basically a Data Mart is a part of a Data Warehouse, specifically concentrated on a part of the business, like a single department. For instance, all the data needed by the Sales Department are copied out of the Data Warehouse into a Data Mart that will suit just the Sales Department.

Summarizing, the Data Warehouse has the following features:

- It forms the basis for analytical applications.
- It experiences enterprise wide usage.
- It is a replication of the data existing in the operational databases.
- The business data are cleaned, re-arranged, aggregated and combined.
- The warehouse is regularly updated with new data.
- It contains the “single truth” of the business.

## **2. The ‘invention’ of the Data Warehouse**

According to Brant (1999) many claims are going round on who actually invented the data warehouse. The right answer to this question, he says, is IBM. To find the roots of the data warehouse we need to go back to 1988, when Barry Devlin and Paul Murphy published their article *An architecture for a business and information system*. This article led to IBM developing an “information-warehouse”-strategy. These roots were quickly buried underneath the data warehouse rage that was created by Bill Inmon’s book *Building the Data Warehouse*.

## **3. Extraction, Transformation and Loading**

A data warehouse is the beginning of the business analysis. The most important process in creating a data warehouse is ETL, which stands for Extraction, Transformation and Loading. In the first step, Extraction, data from one or more data sources (databases or file systems) is extracted and copied into what is called the warehouse. Like in the example of UnoVu, this data source is often a Transaction Processing System. After the extraction, the data has to undergo the Transformation step. These transformations can range from simple data conversions, summarizing and unifying data codes to complex data scrubbing techniques. Especially when the data comes from many different sources, it has to be brought together so that all of the information from each source is brought into the transformation model cleanly. This is a crucial step in the chain from data sources to data warehouse, since it is here that the data quality is taken care of. After the transformation, the “cleansed” data is finally moved from flat files into a data warehouse. This last step is called Loading.



**4. Information sources**

**4.1. Key publications**

Simon & Schaffer (2001) say that most data warehousing professionals recognize Bill Inmon as having first coined the phrase “data warehouse”. Another name that is frequently encountered within the context of Data Warehousing is Ralph Kimball. The Web page <http://www.datawarehousingonline.com/people> calls Bill Inmon the “Father of Data Warehousing” and Ralph Kimball the “Dimensional Data Warehouse Guru”. We could go into detail on what different authors and web pages say about Mr. Inmon and Mr. Kimball, but I think it is enough to leave it at this. These two gentlemen have written books and articles about Data Warehousing, and so have numerous other authors, a lot of whom cite Inmon and Kimball. Therefore I would like to conclude that these two names are two of the most important, and stop the search for more key names in Data Warehousing.

Ralph Kimball is especially renowned for the following books on Data Warehousing:

<b><i>Data Warehousing books by Ralph Kimball (and co-authors)</i></b>
The Data Warehouse Toolkit: The Complete Guide to Dimensional Modeling
The Data Warehouse Toolkit: Practical Techniques for Building Dimensional Data Warehouses.
The Data Warehouse Lifecycle Toolkit: Tools and Techniques for Designing, Developing and Deploying Data Warehouses.
The Data Webhouse Toolkit: Building the Web-enabled Data Warehouse.

**Table 21: Data Warehousing books by Ralph Kimball**

For more information about these books, visit: <http://www.rkimball.com/html/books.html>  
 Kimball has also written numerous articles on this topic. A list of his articles (and the articles themselves) can be found at: <http://www.rkimball.com/html/articles.html>.

Also Bill Inmon has his own company and home page: <http://www.billinmon.com>.  
 In exchange for free access to Inmon’s library, one has to subscribe and provide contact information. In the line of books on Data Warehousing, Inmon has written the following:

<b><i>Data Warehousing book by Bill Inmon</i></b>
Building the Data Warehouse

**Table 22: Data Warehousing book by Bill Inmon**

An excellent site about Data Warehousing is The Data Warehousing Information Center on: <http://www.dwinforcenter.org>. This site was created by Larry Greenfield and according to Ralph Kimball (the guru himself!), it is recommended as the best site for both the overall picture and the detail. The site’s aim is to help readers learn about data warehousing and decision support (i.e. Business Intelligence) systems. It publishes Greenfield’s own essays about data warehousing and decision support, points the reader to external publications, provides links to sites of vendors of various tools, and lists service provider sites. The personal essays are limited to Data Warehousing topics. The external publications and vendor sites, however, include almost all other fields of BI (think about the pyramid). Information about one-time data warehousing and decision support conferences and seminars and organizers of recurring events is listed on: <http://www.dwinforcenter.org/confer.html>.



This last observation is an important one, as it indicates the strong relationships between the different fields that are contained in the concept of Business Intelligence. Hardly any Web page (or company or author, for that matter) can name one field without naming (all) other fields. This was already mentioned in section 1.2 and it is something we will see recurring in most sections of this report. In some situations it even leads to confusion, because it is not entirely clear in which category or field a tool or technique should be classified. But more about that later on.

One last Web site that can come in useful when learning about Data Warehousing is the *Oracle9i Data Warehousing Guide*. This online guide discusses the basic concepts of data warehousing, logical and physical design, managing the warehouse environment, warehouse performance. It is intended for database administrators, system administrators and database application developers who perform the following tasks: designing, maintaining and using Data Warehouses. It does not go into great detail, but it provides a good overview nevertheless. Unfortunately no easier hyperlink could be found than the following:

[http://download-west.oracle.com/otndoc/oracle9i/901\\_doc/server.901/a90237/title.htm](http://download-west.oracle.com/otndoc/oracle9i/901_doc/server.901/a90237/title.htm)

#### 4.2. Commercial Vendors

It is difficult to point out commercial vendors of Data Warehousing, because most BI-products concern tools and techniques that are used to perform activities on top of a data warehouse. A data warehouse forms the basis for advanced multidimensional decision support systems. It is not possible to have Business Intelligence without a data warehouse of some kind. The books by Kimball and Inmon that were named in the preceding subsection are meant to make readers understand and master techniques for creating, controlling and navigating (multi)dimensional databases.

The only web site I came across that lists vendors in the category Data Warehousing is the Business Intelligence Vendor Directory:

<b>Name of the page</b>	<b><i>Business Intelligence Vendor Directory</i></b>
<b>Hyperlink</b>	<a href="http://datawarehouse.ittoolbox.com/vnd.asp">http://datawarehouse.ittoolbox.com/vnd.asp</a>
<b>Description</b>	A site that lists vendors in the following categories: Software, Training, Consulting Firms, Recruiters and Organizations.
<b>Approximate number of links</b>	In the category Software: <ul style="list-style-type: none"><li>• Data Warehouses and Data Marts: 15</li><li>• ETL Packages: 29</li></ul>

**Table 23: Vendors of Data Warehousing**



### 3. Queries & Reports

#### UnoVu Inc. goes Querying & Reporting

Now that UnoVu has its warehouse the real fun can begin. If the Sales Department could acquire the right Reporting tools, they would be able to have standard reports rolling out of the system every day, like weekly summaries on sales by product group or geographical region, or information about new Frequent Buyer Pass owners.

The Marketing Department has other types of wishes. Just recently a new brand of diapers has been developed and UnoVu would like to make a special offer to all the customers that are known to buy diapers and who possess a Frequent Buyer Pass (because these are the customers that spend the most money on UnoVu). With stepwise Querying it is very easy to identify who these customers are. First, select all those customers owning a Frequent Buyer Pass. Then select those that have children under the age of three. After that it is even possible to make a selection of the families that have boys or girls or both (that is, provided the Data Warehouse contains all this information!).

#### 1. What are queries and reports?

The definitions of querying and reporting that I found most attractive are by Alter (1999):

*Query (language): Special-purpose computer language used to provide immediate, online answers to user questions.*

*Report (generator): Program that makes it comparatively easy for users or programmers to generate reports by describing specific report components and features.*

Comparatively little is written on querying and reporting (hereafter called Q&R). That is, compared to techniques like OLAP and data mining. This is probably due to the fact that queries and reports are the most basic forms of analysis on a data warehouse. They already existed back in the 1970's, in the form of hardcopy reports. As Lewis (2001) puts it, interactivity was limited to the visual and perhaps extended to writing notes or highlighting on the reports. Today users have available highly-interactive, online, analytic processing and visualization tools, where selected data can be formatted, graphed, drilled, sliced, diced, mined, annotated, enhanced, exported and distributed. Queries and reports fulfil the purpose of telling management and users “what has happened”, for example how high the sales were in the past month or how are the sales of this month compared to those of last month.

Nearly everywhere querying and reporting are lumped together in one tool. This is quite understandable. The way I see it is the following: There are two types of reporting. The first is the standard reporting. Examples of these are point-in-time reports on sales figures or other key business that appear each day, week, month, etc.

The second type of reporting is when a report is the output of an ad hoc query. Using a query tool, a user can ask questions about patterns or details in the data. Logically, the answer will be in some form of a report. Even though this type of reporting can also be standardized when necessary, the unique thing about queries is that they are built so that the user can ask extra questions about information that doesn't appear directly from the data. If you take this querying to a higher-dimensional level and shorter response times, you arrive at OLAP-tools. More on that in the next chapter.

The results in the reports form an important input element for the Customer Relationship Management. For instance, reports on sales and marketing analyses may result in readjusting the marketing strategies or promotions. Financial reports may indicate that the company is running risks in certain product areas. Analyzing customer profitability can lead to changes in the way certain customers are approached when buying their products. And there are many more examples where these came from.



**2. Information sources**

**2.1. Key publications**

Not many authors have put pen to paper when it comes to the topic Q&R. Probably this is because activities such as generating queries and requesting ad hoc reports are more often than not included in the portfolio of OLAP.

If the reader is especially interested in articles or white papers on Q&R, I suggest he or she try the Web pages that are recommended in Appendix A of this report.

**2.2. Commercial Vendors**

As for commercial vendors of Q&R tools and techniques, there are quite a few to be named. It is unnecessarily time-consuming to write down whole lists in this report. Instead, I chose to present a few direct links to Web pages that contain these lists. It is up to the reader to visit these pages and find the necessary links.

<b>Name of the page</b>	<b>Data Warehousing Information Center</b>
<b>Hyperlink</b>	<a href="http://www.dwinfocenter.org/query.html">http://www.dwinfocenter.org/query.html</a>
<b>Description</b>	The common thread in this list is that all these tools produce a tabular list of information stored in a relational database.
<b>Approximate number of links</b>	172

**Table 31: Vendors of Report and Query tools**

<b>Name of the page</b>	<b>Business Intelligence Vendor Directory</b>
<b>Hyperlink</b>	<a href="http://businessintelligence.ittoolbox.com/vnd.asp">http://businessintelligence.ittoolbox.com/vnd.asp</a>
<b>Description</b>	A site that lists vendors in the following categories: Training (Education), Software, Consulting Firms, Recruiters, Online Services (Data/Text/Web Mining).
<b>Approximate number of links</b>	In the category Software: <ul style="list-style-type: none"> <li>• Queries: 32</li> <li>• Reports and Report Publishing: 76</li> </ul>

**Table 32: BI Vendor Directory (Queries and Reports)**

Q&R is not as far away from our daily line of work as it may seem. One example that I do not want to withhold my readers is the SPSS Report Writer. This report generator is tightly integrated with SPSS to make it as easy as possible for users to write a report of SPSS data. It enables users to quickly create professional-looking, presentation-quality reports from SPSS data using intuitive, word processor-like page layout and formatting features. A demo of the SPSS Report Writer can be found on: [http://www.spss.com/spssbi/report\\_writer/](http://www.spss.com/spssbi/report_writer/)



## 4. OLAP

### UnoVu Inc. discovers OLAP

With the relatively simple types of analysis Querying and Reporting it is not necessary to have a separate staff specialized in making these analyses. Each employee of the Sales or Marketing Department or Management can carry out Querying & Reporting, which is mainly focused on just “telling what has happened”<sup>1</sup>. Now the results of queries and reports are reported back to the higher managers, who are not only interested in these results but also in the ‘how come’ and ‘why’. They would like a further analysis of the results by drilling into the underlying details and looking at results in different ways<sup>2</sup>. In other words, they want a tool that “tells them what happened, *and why*”<sup>3</sup>. But it would be much too time-consuming to teach advanced SQL to all the employees concerned!

Using OLAP UnoVu is able to create a user-friendly environment that even employees who are not so computer-literate, can handle. With relative ease, all staff members can search for answers to questions like: “what will the effects be on the sales of freshly baked bread if the prices of flour went down with €0,10 per kilogram and transportation costs went up by €0,05 per kilometer?” or: “Do the different product groups sell the same way in the different outlets, or are certain products more popular than others?”

<sup>1,2,3</sup> Simon & Schaffer, 2001

OLAP has proven to be the most extensive field in Business Intelligence, so this chapter has become the most extensive chapter in the report. OLAP is the concept that most authors have ventured to write about and most BI-companies claim to have in their portfolio of products and service.

In this chapter we will first look at some descriptions of the concept alone, and present a more general story about what OLAP consists of. In the section ‘Information Sources’, we will discuss what the various companies, authors and other authorities have to say about what OLAP amounts to.

### 4.1. What is OLAP?

A useful definition of On-Line Analytical Processing is the following:

*On-Line Analytical Processing (OLAP) is a category of software technology that enables analysts, managers and executives to gain insight into data through fast, consistent, interactive access to a wide variety of possible views of information that has been transformed from raw data to reflect the real dimensionality of the enterprise as understood by the users.*

[Olap Council, 1997]

OLAP is a technology that allows users to carry out complex data analyses with the help of a quick and interactive access to different viewpoints of the information in data warehouses. These different viewpoints are an important characteristic of OLAP, also called multidimensionality. Multidimensional means viewing the data in three or more dimensions. For a database of a Sales Department, these dimensions could be Product, Time, Store and Customer Age. Analyzing data in multiple dimensions is particularly helpful in discovering relationships that can not be directly deduced from the data itself. Managers must be able to analyze data across any dimension, at any level of aggregation, with equal functionality and ease. OLAP software should support these views of data in a natural and responsive fashion, insulating users of the information from complex query syntax (Forsman, 1997). The fact is that the multidimensionality of OLAP reflects the multidimensionality of an organization. The average business model cannot be represented in a two-dimensional spreadsheet, but needs many more dimensions. Equally, managers and analysts want to be able to look at data from these different dimensions. That is why all these dimensions should be contained in the OLAP database.

Next to this aspect of multidimensionality, Forsman reviews two other key features of OLAP: “calculation-intensive capabilities” and “time intelligence”. The first refers to the ability to perform complex calculations, in order to create information from very large and complex amounts of data. The second feature is the dimension “time”. Time is an integral component of almost any analytical



application. In an OLAP system comparisons of different time periods must be easily defined, as well as the concept of balances over time (totals, averages, etc.)

Turban & Aronson (2001, p.147) employ a much broader definition of OLAP:

*The term online analytical processing (OLAP) refers to a variety of activities usually performed by end users in online systems. There is no agreement on what activities are considered OLAP. Usually one includes activities such as generating queries, requesting ad hoc reports, conducting statistical analyses, and building DSS and multimedia applications. Some include executive information systems and data mining. To facilitate OLAP it is useful to work with the data warehouse (...) and with a set of OLAP tools. These tools can be query tools, spreadsheets, data mining tools, data visualization tools, and the like. (...)*

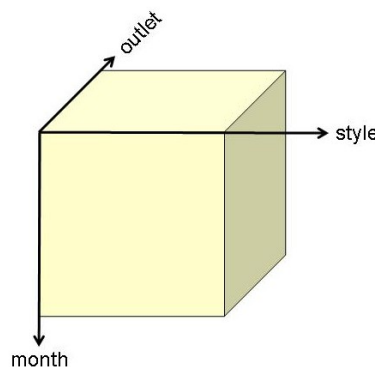
Funny enough, this definition describes exactly how I feel about OLAP after exploring the field. Not all organizations have the same idea about what products/tools/techniques are contained within the concept of OLAP. The only feature that all agree upon is that of Multidimensionality. For the rest, the borderlines between Q&R, OLAP and Data Mining (DM) are very vague. Some say OLAP is DM, some include OLAP in DM, and some include DM in OLAP. Recall what was written in section 1.2, that Turban and Aronson describe BI as ‘the new role of EIS’, so a replacement. Well, in the definition here above they tell us that ‘some include EIS and DM in OLAP’. But weren’t DM and OLAP part of BI?

A brief look at how BI-related organizations categorize their BI-products reveals that most of them offer products in the line of OLAP. OLAP is, I think, the component that is used most generally to describe the activities and services of an organization. As mentioned before, different BI-tools are then contained in this OLAP-element.

#### **4.2.An OLAP example**

To give the reader a feeling of how one should see OLAP, let us look at the following simple example (courtesy of Prof. A.E. Eiben):

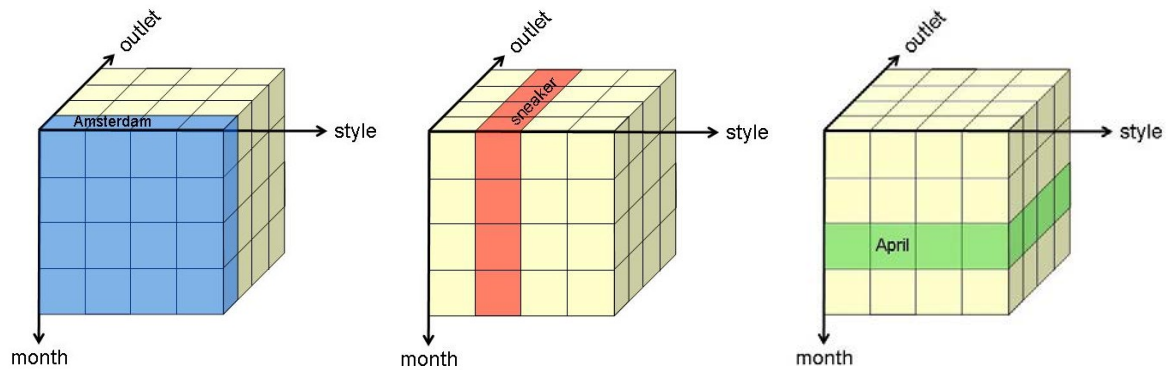
Consider a shoe retailer with many shops in different cities and many different styles of shoes, for example ski boot, gumboot, and sneaker. Each shop delivers data daily on quantities sold in numbers per style. These data are stored centrally. Now the business analyst wants to follow sales by month, outlet and style. These are called dimensions, for example month dimension. If we want to look at the data of these three dimensions and say something significant about them, what we are actually doing is looking at the data stored in a 3-dimensional cube:



**Figure 4-1: A 3-dimensional OLAP cube**

The following three cubes show us how we can look at, respectively: data on all shoe styles sold in all months in the outlet Amsterdam, data on shoe style sneaker sold in all months in all outlets, and data on all shoe styles sold in all outlets in the month April.





Figures 4-2: The OLAP cube looked at from 3 different dimensions

When we combine these three dimensions, we get data on the number of sneakers sold in the outlet Amsterdam in the month April:

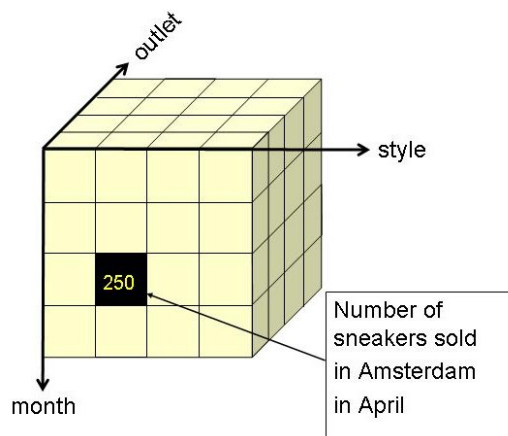


Figure 4-3: The 3 dimensions combined in the OLAP cube

Suppose we want information about the colors of the sneakers or the sizes sold, we would have to define new dimensions. This would mean a 4-, 5- or even more-dimensional cube. Of course cubes like this are no longer ‘visible’ to the eye, but in an OLAP-application they *are* possible!

### 4.3.FASMI

If we go back in time a few decades we come across Dr. E.F. Codd, a well-known database researcher during the 60’s, 70’s and 80’s. In 1993, Dr. Codd wrote a report titled: “*Providing OLAP (On-Line Analytical Processing) to User-Analysts: An IT Mandate*”, in which he defined OLAP in 12 rules. These rules make up the requirements that an OLAP application should satisfy. A year later, Nigel Pendse and his co-author Richard Creeth became increasingly occupied by the phenomenon OLAP. After a critical study of the rules of Dr. Codd, some were discarded and others lumped together in one feature, and a new definition of OLAP was born:

*Fast Analysis of Shared Multidimensional Information (FASMI).*

[Pendse, 2001]

In a later article they go on to describe what they mean exactly with the five separate words that make up this definition:

“Fast” means that the system is targeted to deliver most responses to users within about five seconds, with the simplest analyses taking no more than one second and very few taking more than 20 seconds.

“Analysis” means that the system can cope with any business logic and statistical analysis that is relevant for the application and the user, and keep it easy enough for the target user.



*“Shared” means that the system implements all the security requirements for confidentiality (possibly down to cell level) and, if multiple write access is needed, concurrent update locking at an appropriate level.*

*“Multidimensional” means that the system must provide a multidimensional conceptual view of the data, including full support for hierarchies and multiple hierarchies, as this is certainly the most logical way to analyze businesses and organizations.*

*“Information” is all of the data and derived information needed, wherever it is and however much is relevant for the application.*

[Pendse, 2002]

Nigel Pendse declares that this definition was first used by him and his company in early 1995, and that it has not needed revision in the years since. He states that the definition has now been widely adopted and is cited in over 120 Web sites in about 30 countries. Before a critical author like myself can say: “I agree with Mr. Pendse on this one”, this statement had to be subjected to a thorough verification. Research with the help of Google revealed there to be 34 countries with one or more Web site(s) containing the term “FASMI” (after ruling out the numerous sites that present articles by an Afghan journalist called Fasmi). A total of 21 countries host one or more Web site(s) that write about FASMI in combination with The OLAP Report. Based on these findings I can safely say that I agree with Nigel Pendse. The term is widely and globally used. Striking is, next to mostly English-language sites, the large number of German (university) sites that include the terms!

OLAP products and applications have been around for much longer than most people think, as Nigel Pendse says on <http://www.olapreport.com/origins>. On this page he describes the origins of today’s OLAP products, and provides us with his view on a few lessons to be learnt from a 35-year history of OLAP:

### ***Lessons to be learnt from a 35-year history of OLAP***

1. Multidimensionality is here to stay. Even hard to use, expensive, slow and elitist multidimensional products survive in limited niches; when these restrictions are removed, it booms. We are about to see the biggest-ever growth of multidimensional applications.
2. End-users will not give up their general-purpose spreadsheets. Even when accessing multidimensional databases, spreadsheets are the most popular client platform. Multidimensional spreadsheets are not successful unless they can provide full upwards compatibility with traditional spreadsheets, something that Improv and Compete failed to do.
3. Most people find it easy to use multidimensional applications, but building and maintaining them takes a particular aptitude — which has stopped them from becoming mass market products. But using a combination of simplicity, pricing and bundling, Microsoft now seems determined to prove that it can make OLAP servers almost as widely used as relational databases.
4. Multidimensional applications are often quite large and are usually suitable for workgroups, rather than individuals. Although there is a role for pure single-user multidimensional products, the most successful installations are multi-user, client/server applications, with the bulk of the data downloaded from feeder systems once rather than many times. There usually needs to be some IT support for this, even if the application is driven by end-users.
5. Simple, cheap OLAP products are much more successful than powerful, complex, expensive products. Buyers generally opt for the lowest cost, simplest product that will meet most of their needs; if necessary, they often compromise their requirements. Projects using complex products also have a higher failure rate, probably because there is more opportunity for things to go wrong.

**Table 4-1: Lessons to be learnt from a 35-year history of OLAP**

#### ***4.4.OLAP Applications***

OLAP technology can be used in a wide range of business applications and industries. The OLAP Report (Pendse, August 2001) lists the following application areas:



<i>Application Area</i>	<i>Description</i>
<b>Marketing and sales analysis</b>	Mostly found in consumer goods industries, retailers and the financial services industry.
<b>Clickstream analysis</b>	More on this can be read in 5.4.
<b>Database marketing</b>	Determine who are the best customers for targeted promotions for particular products or services.
<b>Financial reporting</b>	To address this specific market, certain vendors have developed specialist products.
<b>Management reporting</b>	Using OLAP based systems one is able to report faster and more flexible, with better analysis than the alternative solutions.
<b>Balanced Scorecard</b>	See the next section.
<b>Profitability analysis</b>	Important in setting prices and discounts, deciding on promotional activities, selecting areas for investment or divestment and anticipating competitive pressures.
<b>Quality analysis</b>	OLAP tools provide an excellent way of measuring quality over long periods of time and of spotting disturbing trends before they become too serious.

**Table 42: OLAP application areas**

In the next chapter we will see that many authors ascribe Clickstream Analysis and Profitability Analysis to the field of Data Mining, rather than OLAP.

According to the OLAP Council White Paper (Forsman, 1997), the following OLAP applications are typical:

- Financial modeling (budgeting, planning)
- Sales forecasting
- Customer and product profitability
- Exception reporting
- Resource allocation and capacity planning
- Variance analysis
- Promotion planning
- Market share analysis

**4.5. Balanced Scorecard**

A product that many BI-companies have to offer is the Balanced Scorecard. The Balanced Scorecard (hereafter abbreviated as BSC) was born in 1992, when Robert Kaplan and David Norton published an article about it in the Harvard Business Review. Applications based on the BSC methodology are often integrated with OLAP-environments. The Web page <http://www.balancedscorecard.com/how.htm> tells us that “The Balanced Scorecard application is the marriage of the balanced scorecard methodology and advanced OLAP technology.” Apparently the BSC is often related to OLAP, so therefore the section about BSC is included in this chapter.

The BSC is a simple control mechanism that helps managers monitor their business performance in the following four perspectives:

- |                         |                               |
|-------------------------|-------------------------------|
| • Customer knowledge    | • Internal business processes |
| • Financial performance | • Learning and growth         |



According to Hermelink & Van Bilsen (2000) the use of BSC revolves around three questions:

- ◆ What do we want to achieve with the organization? (Strategic goals)
- ◆ In order to achieve our goals, what should we be good at? (Critical Success Factors)
- ◆ How can we measure that we are achieving what we want? (Key Performance Indicators)

All BSC tools are able to lay down critical success factors (CSF's) in the four perspectives named above. These factors can be connected to the key performance indicators (KPI's). A good BSC contains all the KPI's that are critical for achieving the company's strategic goals. Based on the KPI's decisions for the appropriate actions can be made.

In comparing the Balanced Scorecard with Business Intelligence, I came to the following conclusion: The BSC is focused on the internal management of an organization, whereas most other BI-applications are focused on external business. The BSC is a tool for managers, and other BI-tools are suitable for all organizational levels.

According to The OLAP Report's author Nigel Pendse (August 2001) 'the balanced scorecard is a 1990s management methodology that in many respects attempts to deliver the benefits that the 1980s executive information systems promised, but rarely produced. (...) we have noticed that while an increasing number of OLAP vendors are launching balanced scorecard applications, some seem to be little more than rebadged EISs, with no serious attempt to reflect the business processes that Kaplan and Norton advocate.'

## **4.6. Information sources**

### **4.6.1. Key publications**

As said in section 4.2, Nigel Pendse is a key author when it comes to OLAP. From his hand comes The OLAP Report, which he and his co-authors call 'The Independent and Comprehensive Guide to OLAP Applications, Technologies and Products.' The first edition of The OLAP Report emerged in August of 1995. This initiative grew out of the lack of clear and unbiased information about OLAP tools and products. Accordingly, the report is aimed primarily at companies wishing to understand OLAP so they could better know whether it was suitable for their needs, and if so, how to select tools and deploy them successfully. It will not provide an immediate suggestion of which product to buy, but it will inform subscribers about the questions they should be asking, and which vendors to eliminate early. For an annual subscription of \$2900, The OLAP Report provides access to a large, regularly updated Web site, plus a copy of the latest printed edition. Because the authors also, as they say on their site, support the Web philosophy of providing a significant amount of free educational and newsworthy content, some information is accessible to non-subscribers.

There are quite a few organizations that provide us with their opinions and visions of OLAP. One of these is the OLAP Council: <http://www.olapcouncil.org>. This is a site that many other web sites refer to. However, be aware that the most recent date that appears on this site is the year 1999.

The fact is that every self-respecting company that offers products in the line of OLAP, has published articles and white papers about OLAP. Unfortunately, the fact is also that many references to Web pages with papers on OLAP either do not exist anymore or date back to a number of years ago. Therefore, I get a very strong feeling that the real hype to write about OLAP has already disappeared. But, this might also be a sign of maturity.



#### 4.6.2. Commercial Vendors

As with the commercial vendors of Queries & Reports, I chose to present some links to pages containing the names of and links to vendors, instead of writing down the whole lists here.

<b>Name</b>	<b><i>Data Warehousing Information Center</i></b>
<b>Hyperlink</b>	<a href="http://www.dwinfocenter.org/olap.html">http://www.dwinfocenter.org/olap.html</a>
<b>Description</b>	OLAP / Multidimensional Database Tools
<b>Approximate number of links</b>	150

**Table 43: Vendors of OLAP and Multidimensional database tools**

<b>Name</b>	<b><i>Business Intelligence Vendor Directory</i></b>
<b>Hyperlink</b>	<a href="http://businessintelligence.ittoolbox.com/vnd.asp">businessintelligence.ittoolbox.com/vnd.asp</a>
<b>Description</b>	A site that lists vendors. Click Software > Packaged BI Suites > OLAP Packages.
<b>Approximate number of links</b>	24
<b>Description</b>	Click Software > Reports/Queries > OLAP
<b>Approximate number of links</b>	8

**Table 44: BI Vendor Directory (OLAP and OLAP Packages)**





## 5. Data Mining

### UnoVu Inc. takes on Data Mining

Every time a UnoVo-customer qualifies for a Frequent Buyer Pass he or she gets a welcome-package. Now what would be a (strategically) better idea than to include those products in the welcome-box that are most likely to suit that customer? With Data Mining it is possible to discover patterns in the purchase behavior of customers. This way, profiles can be made of types of customers that are most likely to buy certain products. So if this new Frequent Buyer Pass customer happens to be 30-year old mother of twins aged 7 months who is a part-time lawyer, and Data Mining has revealed to the Marketing Department of UnoVu that people in the age group of 25 to 35 who are a lawyer mostly buy baby-food of the brand *BabySnack*, it would a good idea to include *BabySnack* in the welcome-box of this lady, seeing as there is a good chance that she will like giving this brand of baby-food to her twins. (Of course, it is actually the twins who will have to like *BabySnack*, but that is besides the point here...)

Also, Data Mining can be used to identify the most profitable customers for certain product groups. By "crossing" the data about most frequently bought products with the customer profiles, UnoVu will know which types of customers are most likely to buy which types of products.

As a last challenge, UnoVu would like to build a web site on which customers can buy products on-line and have them delivered to their home address. It will be very important to know how effective the web site is, and whether and how the site should be updated. Web Mining tools can assist in this, as we can read in the fourth paragraph of this chapter.

### 5.1. What is Data Mining?

In search for definitions of Data Mining, I encountered two ways in which the concept is defined by authors. An example of one way is:

*Data mining is the use of data analysis tools to try to find the patterns in large transaction databases.*

[Alter, 1999]

The extended versions are like the following:

*Data mining is analysis of large pools of data to find patterns and rules that can be used to guide decision making and predict future behavior.*

[Laudon & Laudon, 2000]

The first type of definitions talk about finding patterns in large databases; the second type also include *why* we want to find these patterns, namely to help decision making and predict the future. Based on this, in my opinion there are four key elements that make up a good definition of Data Mining:

- finding patterns
- large amounts of data
- help decision making
- predict the future

The idea of Data Mining (DM) is to discover patterns in large amounts of data. Whereas query and even OLAP functions require human interaction to follow relationships through a data source, data mining programs are able to derive many of these relationships automatically by analyzing and "learning" from the data values contained in files and databases (Lewis, 2001). The patterns that are found in the data could provide information that cannot directly be deduced from the data itself, patterns and connections that are not straightforward. These 'invisible' patterns might not always be logical and useful. For instance, for a supermarket chain that is based in several different countries, DM might show that the sales of yogurt in America might be strongly correlated with the sales of bicycles in the UK. Naturally this is a coincidental connection. But if DM reveals that customers who buy Product X most of the time also purchase Products Y and Z, it is a very valuable tool for the management to help them in their strategic decision making. Products X, Y and Z could be in shelves that are located close to each other, or the management could chose to make special offers for these

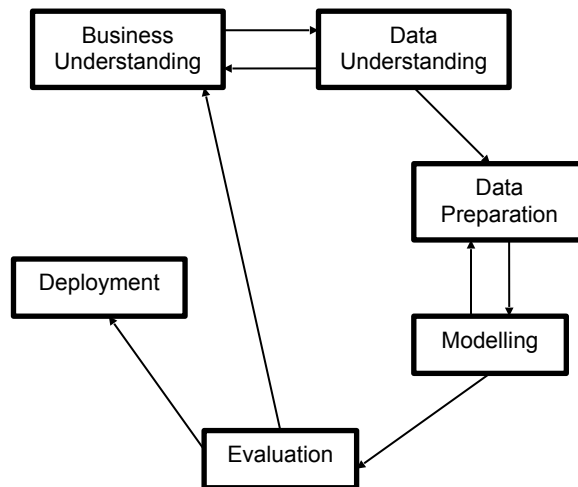


three products at the same time, to increase the sales in a short time. Also, the introductory example of UnoVu at the beginning of this chapter shows what is possible with DM.

Actually there is nothing new about looking for patterns in data. People have been seeking patterns in data ever since human life began. Hunters seek patterns in animal migration behavior, farmers seek patterns in crop growth, politicians seek patterns in voter opinion. A scientist’s job (like a baby’s) is to make sense of data, to discover the patterns that govern how the physical world works and encapsulate them in theories that can be used for predicting what will happen in new situations. The entrepreneur’s job is to identify opportunities, that is, patterns in behavior that can be turned into a profitable business, and exploit them. (Witten & Frank, 2000)

**5.2. The Data Mining Process**

A quite general view of the Data Mining process is the one offered by Van der Putten (1999):



**Figure 5-1: The Data Mining Process**

This model is also often referred to as CRISP, the Cross Industry Standard Process.

It is easy to read a book about DM-techniques and –modeling and think you understand DM and know what has to be done to ‘mine’ your data. It is also a big mistake. Very significant is the larger picture, or process, within which the data mining takes place. The whole business around it, the type of data, the preparations of the data and a thorough evaluation have to be taken into account. Each step of the process consists of a number of activities:







<i>Step in the process</i>	<i>Description</i>
<b>Business Understanding</b>	Determining the business objectives, situation assessment, determining the goal of the data mining, producing a project plan.
<b>Data Understanding</b>	Collecting the initial data, describing and exploring these data and verifying its quality.
<b>Data Preparation</b>	Selecting, cleaning, constructing, integrating and formatting the data.
<b>Modeling</b>	Selecting a modeling technique, generating test design, building and implementing the model.
<b>Evaluation</b>	Evaluating the results, reviewing the process and determining the next steps.
<b>Deployment</b>	Plan deployment, plan monitoring and maintenance, producing the final report and reviewing the project.

Table 51: The steps of the Data Mining process

(Van der Putten, 1999)

### 5.3. Data Mining Techniques

There are many techniques for carrying out Data Mining. A book by Witten & Frank (2000) presents a clear separation between (1) the desired output information and (2) the tools used to acquire this desired information. The type of output information is the way in which the newly gained knowledge is represented. For instance: classification of data, clustering of data, association rules, decision trees or tables or trees for numeric predictions. All these types of output can be the result of one or more of a wide range of techniques (also called algorithms). Examples of algorithms are: inferring rules, statistical modeling, constructing decision trees, constructing rules with covering algorithms, linear modeling. Other DM-tools are case based reasoning, neural computing, genetic algorithms and support vector machines.

All these techniques and concepts can also be found in the categories Machine Learning and Artificial Intelligence. In fact, they are *all* about Artificial Intelligence, because the information that is (artificially) gained provides the user with some form of intelligence. And the techniques used mostly involve a machine that ‘learns’ from the input examples it gets and is afterwards able to predict what will happen when other examples occur.

There is no real indication as to which techniques should be used in which cases. However, a choice of technique can be based on one or more of the following criteria:

- Solution quality
- Speed
- Solution comprehensibility
- Expertise required

In some cases it could be preferred to have a DM-tool that provides answers very quickly, no matter what the quality of the solution is. In other cases one might want a solution of very high quality, but if this means that the solution concerned becomes totally incomprehensible one will have no use for it.

### 5.4. Web Mining: the Internet-variant of Mining

An area of growing importance for companies trying to sell their products is e-commerce. To give an indication of the growth of this area: in a Data Mining book (Witten & Frank) written in 2000, the first before last sub-section is dedicated to mining the Web and the authors place it in an infancy stadium. Here we are, two years later, and a large part of the conversation on Data Mining is dedicated to Web Mining!



The idea behind Web Mining is that the information and knowledge that is “dug up” by data mining in every-day databases can also be used to provide information about a web site and its visitors. Web sites, and especially commercial ones, generate gigabytes of data a day that describe every action made by every visitor to the site. One should realize that there is much more information hidden in the pages of a web site than one would think there is. And it is exactly this ‘invisible’ and ‘not-straightforward’ information that is most valuable to have when engaged in e-commerce activities. Typical questions answered by Web Mining are:

- On which page of the web site do visitors enter / leave the site?
- How much time do visitors spend on which page of the site?
- How many visitors fill their shopping cart but leave the site without making a purchase?

An article by Carine Joosse (2000) gives a short but interesting description of the different ways of applying data mining to the Internet. The first is ‘Mining the Web’ itself. An example of this is collecting data from various sites and categorizing, analyzing and presenting them on new web pages for the benefit of the web visitor. An other example is a search engine on the Web: by searching for hits of a word, phrase or synonym, registering these hits, grouping them into categories and keeping up a history, the search engine could be made more powerful. The data mining element in this is making predictions, trend analysis, categorizing and data reduction.

A second type of Web mining is ‘Web usage mining’. The goal of web usage mining is analyzing the site navigation: how do visitors “click” through the site, how much time do they spend on which part (page) of the site, on which point do they enter or leave the site? This form of analysis is also referred to as Clickstream Analysis. Just as important is to keep records of which visitors finally make a purchase, which visitors start making a purchase (i.e. start filling their virtual shopping cart) and do not buy in the end, and which visitors leave the site without making a purchase. By combining all these data with the registered customer profiles it is possible to define those types of customers that are most likely to purchase using the internet. Also, these customer profiles in connection with their behavior on the Web site can be used to see if the site should be designed differently.

While most authors ascribe the Web Mining tool Clickstream Analysis to the Data Mining field, Nigel Pendse says in his “OLAP Report” that it is one of the latest OLAP applications (Pendse, 2001). As we saw in section 4.3, he also add Database Marketing to his list of OLAP applications. In his opinion, determining who the preferred customers are ‘can be done with brute force data mining techniques (which are slow and can be hard to interpret), or by experienced business users investigating hunches using OLAP cubes (which is quicker and easier)’. In other words, here we encounter once again the vague boundaries that exist between the concepts within Business Intelligence!

Web Mining applications of a more advanced level are personalization and multichannel-analysis. Personalization happens when rules are activated in order to offer personalized content to the visitor. A danger in this application is that the information is not always fully reliable, in the sense that the visitor cannot be categorized correctly. When individual visitors make use of a large company network, for example, they will not be recognized as separate visitors. What Multichannel-analysis comes down to is anticipating the behavior, wishes and possibilities of the customer in the use of different communication channels.



**5.5. Information sources**

**5.5.1. Key publications**

Information about the concepts of Q&R and OLAP is only for a small part contained in books that deal on more general topics. In contrast to this, whole books have been written on the concept of Data Mining alone. So in this section of Key Publications we can, in contrast to the previous two chapters, add quite a number of books. Unfortunately it was outside the time scope of this investigation to go into depth researching what books would be especially good books on Data Mining. The fact is that there are many books on this topic. I think this is because the field of Data Mining is quite an extensive one when it comes to the number of different tools and techniques. It seems like it is a topic of its own.

I can advise readers to also look for books on Machine Learning and Artificial Intelligence, when interested in Data Mining. As said in section 5.2, these two fields are closely related and even integrated with DM.

On the Internet I found the Web site of KDnuggets (where KD stands for Knowledge Discovery). This company claims to be ‘the leading source of information on Data Mining, Web Mining, Knowledge Discovery’. Publications can be found on: <http://www.kdnuggets.com/publications/index.html>

Also interesting links for Data Mining can be found on: <http://www.andypryke.com/university/TheDataMine.html>.

Another interesting one for publications is: <http://www.acm.org/sigkdd/>. This is the Association for Computing Machinery, Special Interest Group Knowledge Discovery and Data Mining. The primary focus of SIGKDD is to provide the premier forum for advancement and adoptions of the “science” of knowledge discovery and data mining. To do this, the SIGKDD encourages basic research in KDD (through annual research conferences, newsletter and other related activities), adoption of “standards” in the market in terms of terminology, evaluation, methodology, and interdisciplinary education among KDD researchers, practitioners and users. Each year they organize an international conference on Knowledge Discovery and Data Mining. Up to the year 1999 it seems these conferences have been sponsored by the American Association for Artificial Intelligence, links to information about the first five can be found on: <http://www.dcs.elf.stuba.sk/emg/kdd.htm#conf>. From the year 2000 onwards, information (and proceedings!) can be found on: <http://www.acm.org/sigkdd/2000/>, [/2001/](http://www.acm.org/sigkdd/2001/), [/2002/](http://www.acm.org/sigkdd/2002/), etc.

**5.5.2. Commercial Vendors**

The Business Intelligence Vendor Directory distinguishes four categories of products offered by DM-vendors:

- Fraud Detection
- Customer Profiling
- Forecasting
- Discovery / Trend Identification.

The following table contains the link to this site and the information needed to find vendors in the various categories:

<b>Name</b>	<b>Business Intelligence Vendor Directory</b>
<b>Hyperlink</b>	<a href="http://businessintelligence.ittoolbox.com/vnd.asp">businessintelligence.ittoolbox.com/vnd.asp</a>
<b>Description</b>	A site that lists vendors. Click Software > Data Mining
<b>Approximate number of links</b>	Fraud Detection: 7 Customer Profiling: 17 Forecasting: 28 Discovery/Trend Identification: 45



**Table 52: BI Vendor Directory (Data Mining)**

The Data Warehousing Information Center groups vendors in the following categories:

- Data Mining
- Web Analytics
- Text Mining
- Forecasting
- Data Visualization

<b>Name</b>	<b>Data Warehousing Information Center</b>
<b>Hyperlinks</b>	<a href="http://www.dwinfocenter.org/datamine.html">http://www.dwinfocenter.org/datamine.html</a> <a href="http://www.dwinfocenter.org/ecommerce.html">http://www.dwinfocenter.org/ecommerce.html</a> <a href="http://www.dwinfocenter.org/docum.html">http://www.dwinfocenter.org/docum.html</a> <a href="http://www.dwinfocenter.org/statisti.html">http://www.dwinfocenter.org/statisti.html</a> <a href="http://www.dwinfocenter.org/dataviz.html">http://www.dwinfocenter.org/dataviz.html</a>
<b>Description</b>	Listings of vendors providing tools for Data Mining, E-commerce, Text Mining, Forecasting and Data Visualization.

**Table 53: Vendors of Data Mining and related concepts**



## 6. Business Intelligence again

### 6.1. What is Business Intelligence?

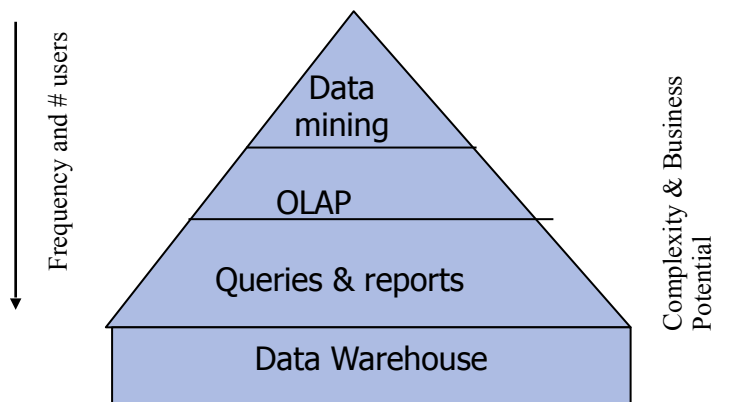


Figure 6-1: The pyramid of BI

As I mentioned in the first chapter, the above ordering of components hanging under the umbrella of Business Intelligence is widely adopted. However, it must be said that there are authors who do not adopt these four components, or who name only some of them and add other components. Simon & Shaffer (2001), for instance, include Executive information systems (EISs) as an easy-to-use ‘extension’ of OLAP. But then again, Turban & Aronson (2001) state that the term Business Intelligence is used to describe the new role of EIS. In other words, they see it as a replacement.

The dividing lines between these different fields are very vague. Like the classification was once invented but by now each organization is giving their own twist to it. This is proved by the fact that the one author contains for example OLAP in DM, while the other author turns it around and sees DM as a part of OLAP. It is not that they deliberately contradict each other; we are just dealing with different interpretations.

A search for the term Business Intelligence on <http://www.whatis.com> results in the following description:

*Business Intelligence (BI) is a broad category of applications and technologies for gathering, storing, analyzing, and providing access to data to help enterprise users make better business decisions. BI applications include the activities of decision support systems, query and reporting, online analytical processing (OLAP), statistical analysis, forecasting, and data mining.*

[whatis.com, 2001]

### 6.2. Business Intelligence vs. Decision Support Systems

Maybe the reader has noted that the term ‘decision support’ is used quite often throughout this report. Clearly, this is because the bottom line of Business Intelligence is supporting decision making. There are three types of Decision Support: model-driven, data-driven and user-driven. One of the first things my supervisor said when we discussed this topic, was that he wondered whether Business Intelligence is actually the “new term” for Decision Support Systems. And more specifically: is BI “replacing” Data-driven Decision Support?



The website <http://www.whatis.com> presents the following definitions of the two concepts:

<b>“Whatis.com” Definitions</b>	
<b>Business Intelligence</b>	<b>Decision Support System</b>
<i>Business Intelligence (BI) is a broad category of applications and technologies for gathering, storing, analyzing, and providing access to data to help enterprise users make better business decisions.</i>	<i>A decision support system (DSS) is a computer program application that analyzes business data and presents it so that users can make business decisions more easily.</i>

**Table 61: BI vs. DSS definition**

The key similarity in these two definitions is “making business decisions”, and in particular both concepts are focused on helping to make these decisions in a better and easier way. The other important similarity is they both involve decision making “based on data”.

The way Dekker (2002) looks at it is that Data Warehousing and Data Mining have two precursors: DSS and EIS. DSS is focused on the lower and middle management and makes it possible to look at and analyze data in different ways. EIS is the precursor focused on the higher management. Given the fact that Data Warehousing and Data Mining form a large part of Business Intelligence, you could indeed see DSS as the precursor of BI.

The following (Alter, 1999) fully enforces Eiben’s theory about BI replacing data-driven decision support: A number of approaches developed for supporting decision making include online analytical processing (OLAP) and data mining. The idea of OLAP grew out of difficulties analyzing the data in databases that were being updated continually by online transaction processing systems. When the analytical processes accessed large slices of the transaction database, they slowed down transaction processing critical to customer relationships. The solution was periodic downloads of data from the active transaction processing database into a separate database designed specifically to support analysis work. This separate database often resides on a different computer, which together with its specialized software is called a **data warehouse**.

What Alter points out here is that, because of the difficulties when analyzing the data to support decision making, the data are duplicated in a Data Warehouse on top of which OLAP and Data Mining can be applied without disturbing transaction processing. In other words, the components that make up Business Intelligence are replacing the old-fashioned way of performing data-driven decision support on the original transaction processing systems.

**6.3. Current status**

As we discussed earlier on, according to Howard Dresner (Buytendijk, no.8 1997) Business Intelligence is an umbrella-concept with a large number of techniques hanging underneath it. Several segments can be distinguished. On the underside of the market these are query & reporting tools and the so-called OLAP-viewers. On the upper side these are DSS- and EIS-packages. Business Intelligence is the covering concept of providing management information. If we add Dekker’s view from the previous section to this and replace the DSS and EIS with Data Warehousing and Data Mining (not respectively though), we have all the components of the pyramid above. Roughly speaking, that is.

Turban & Aronson (2001) write that the term Business Intelligence (BI) or Enterprise Systems is used to describe the new role of the Executive Information System, especially now that data warehouses can provide data in easy-to-use, graphics-intensive query systems capable of slicing and dicing data (Q&R) and providing active multi-dimensional analysis (OLAP).



Simon & Shaffer (2001) find the following classification of business intelligence applications to be useful:

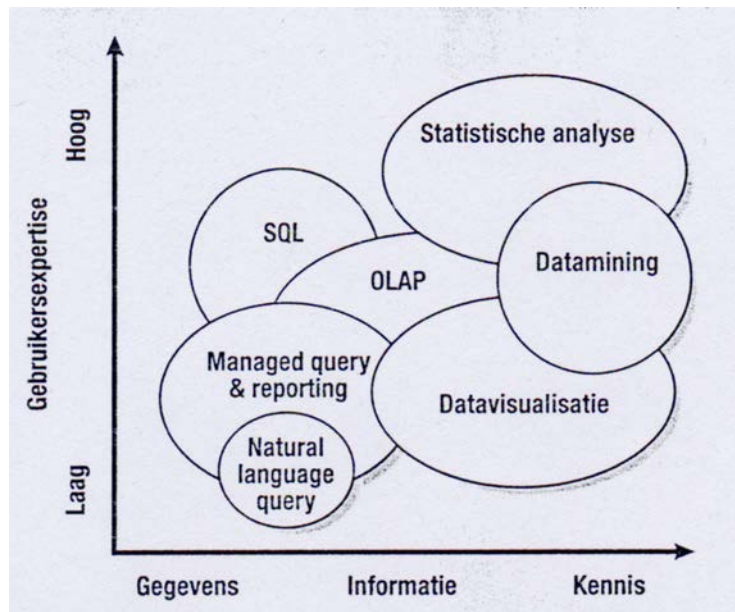
- Simple reporting and querying
- Online analytical processing (OLAP)
- Executive information systems (EISs)
- Data mining

Why do they include EISs as an application amongst Q&R, OLAP and DM? Don't EISs already have some form of Q&R and even OLAP-like activities in them?

One thing is certain, Turban & Aronson and Simon & Shaffer will not agree on a definition of BI. The first duo says that BI *replaces* EIS, and the second *includes* EIS in BI.

**6.4.Application Areas**

The following picture I found in an article by Pieter den Hamer (1998). It shows what applications are used by people with a certain level of Expertise (from Low to High) and concerning a certain “knowledge value” (Data, Information or Knowledge).



**Figure 6-2: Knowledge value versus user expertise**

What this figure makes clear is that end-users with different levels of expertise can apply Business Intelligence applications to different levels of knowledge. When we think of different types of users, we can picture a junior sales assistant or an accountant or an employee from the marketing department, but also their managers or the director’s personal secretary or the director himself. Some of these might not want to use BI, but the idea is that all types of end-users can use BI-tools. They will all use BI in a different way. After all, not everyone is equally computer-literate, not everyone has the same user expertise.

With BI-tools it is possible to carry out analyses and reports on virtually all thinkable aspects of the underlying business, as long as the data about this business come in large amounts and are stored in a Data Warehouse. Departments that are known to benefit most from Business Intelligence are (Database) Marketing, Sales, Finance, ICT (especially the Web) and the higher Management.





Recall in the chapter about Queries & Reports the remark about Q&R not being as far away from our daily line of work as it may seem. A very good example for this was the SPSS Report Writer that is tightly integrated with SPSS. Another BI-tool integrated with an application many of us use daily is *Business Intelligence for Excel* offered by Business Intelligence Technologies, Inc. This tool – also called BIXL – differs from other BI-tools in this respect: the product delivers to an end-user's Excel spreadsheet data that can be used for analytical and reporting purposes, from Microsoft's *Analysis Services* (and other OLE DB for OLAP cube providers), and adds all-important write-back capabilities for planning (and budgeting and forecasting) tasks (Business Intelligence Technologies, Inc., 2002).

### **6.5. Return On Investment for BI Projects**

*“Calculating ROI for Business Intelligence Projects”* is a white paper by Jonathan Wu. I found this paper to be relevant, because not only can it be found on the website of the company the author works for, but also numerous other websites and the magazine DM Review have published it. In this paper Wu addresses the calculation of ROI, including financial measures such as the net present value, internal rate of return and payback period, as well as other non-financial considerations for BI projects.

The paper's summary is given here. For the full contents see [Wu, 2000].

*The quantitative and qualitative benefits of a BI project need to be evaluated before the project is undertaken. Calculating the ROI on a BI project is one means of measuring the benefits to the organization. However, even more critical to the success of the BI project than the calculated costs, benefits, and ROI, is the extent of “buy-in” for the project from the executive leadership and the business operations of the organization. Having support for the BI project from senior management as well as user involvement in the configuration of the application(s) is essential to its success. Without the necessary support and involvement, the ROI calculation for the BI project is meaningless.*

[Wu, 2000, p.13]

### **6.6. Competitive Intelligence**

In the context of Business Intelligence I once stumbled upon the term Competitive Intelligence in an article written by Voorma (2001). In this article Voorma writes that Competitive Intelligence (hereafter to be called CI) is meant to transform information into action-focused knowledge with news value for the strategy of a company. CI can be used in many situations. Amongst others:

- Learning from the mistakes of competitors
- Anticipating new legislation
- Anticipating trends in environment and market
- Identifying partners and takeover candidates

The output of a CI-process is Actionable Intelligence, by Voorma abbreviated as AI, but please be sure not to get mixed up with the widely accepted abbreviation of Artificial Intelligence! Actionable Intelligence is the action-focused (actionable) knowledge, the intelligence that stimulates changes in an organization's strategy.

For the rest it is not entirely clear from Voorma's article what the added value is of CI. He names a few initial steps, like:

- Identifying clear goals
- Choosing between strategic or operational control
- Specifying information requirements
- Collecting data

Aren't these steps (or variations of these steps) essential in all types of Business Modeling and Systems Engineering? And thus part of the initial phase of creating Business Intelligence? Maybe Voorma's point is that these steps should be reconsidered well before collecting and integrating data from different systems into a data warehouse and start to develop or use BI-tools on top of it.



A search on Google reveals there to be many companies offering CI-solutions, often in the same line as BI, but for this research study outside the scope of BI. However, to give the interested reader a start I will present a few useful links. A pioneer in the field of Competitive Intelligence is Leonard Fuld. On the Web page <http://www.fuld.com/whatCI.html> he gives ten descriptions of what CI is and does for a corporation, and dispels ten common misconceptions of CI.

There is even a CI Academy, see: <http://www.academyci.com>. This Web page names three leading thinkers in the field of CI: Fuld (the dean of CI), Gilad (the CI guru) and Herring.

Another Web site on CI is <http://www.competitive-intelligence.co.uk>.

**6.7.Information sources**

**6.7.1. Key publications**

As we have seen throughout this report, various authors have committed themselves to writing about the different concepts under the umbrella of Business Intelligence. With respect to publications on BI as a total I would like to stick to just two names: Howard Dresner and Frank Buytendijk. As I explained in Chapter 1, Howard Dresner had a big hand in creating the concept Business Intelligence. During the years he has published articles and given interviews. He is quite universally seen as “the big man” when it comes to BI. Mr. Dresner is currently a vice president and research director in Gartner Research. Frank Buytendijk struck me as having published quite a few articles in which he considers the big picture of BI rather than just a part of it. That is why I include his name here. Mr. Buytendijk has worked as an advisor with Synergetics IT Consultants within the advice group Business Intelligence and is currently senior research analyst with the Gartner Group.

Many companies that provide BI-solutions publish white papers and articles. Here are a few links:

<i>Company</i>	<i>Hyperlink</i>
<b>Hyperion</b>	<a href="http://www.hyperion.com/products/whitepapers/">http://www.hyperion.com/products/whitepapers/</a>
<b>IBM</b>	<a href="http://www-3.ibm.com/software/data/pubs/papers/">http://www-3.ibm.com/software/data/pubs/papers/</a>
<b>MicroStrategy</b>	<a href="http://www.microstrategy.com/Publications/Whitepapers/">http://www.microstrategy.com/Publications/Whitepapers/</a>

**Table 62: Some hyperlinks to companies' whitepapers**

**6.7.2. Commercial Vendors**

Somewhere in the beginning of my investigation of the field Business Intelligence, I read the statement: “this-and-that company (I don’t recall exactly which one) is leading in BI.” Bingo, I thought, I have found the number one BI-provider! But I was wrong. Gradually I came across one company after the other claiming to be either global leader, leading provider or best in BI. There is no number one. Neither would I dare make a ranking of a top three or top ten of BI-vendors. After all, what would be the criteria based on which the vendors could be ranked, when these vendors do not even agree with each other on the components that make up BI? All I can do, and here we see the advantage of being somewhat a layman (or, laywoman), is make a list of BI-vendors that I most frequently encountered during my quest on the World Wide Web and in journals and articles. Two comments have to be made. First: here the volatility of the field proves itself yet again. Names of BI-providers that appear as being ‘promising’ in articles dating from 1997 have in 2002 disappeared altogether as being promising BI-vendors. Secondly: please keep in mind that this list is no doubt incomplete.

<i>Most frequently encountered BI-vendors, as of 2002</i>	
<i>Company</i>	<i>Hyperlink</i>
Business Objects	<a href="http://www.businessobjects.com">http://www.businessobjects.com</a>



Brio Technology	<a href="http://www.brio.com">http://www.brio.com</a>
Cognos	<a href="http://www.cognos.com">http://www.cognos.com</a>
Crystal Decision	<a href="http://www.crystaldecisions.com">http://www.crystaldecisions.com</a>
Hummingbird	<a href="http://www.hummingbird.com">http://www.hummingbird.com</a>
Hyperion	<a href="http://www.hyperion.com">http://www.hyperion.com</a>
IBM	<a href="http://www.ibm.com">http://www.ibm.com</a>
Informatica	<a href="http://www.informatica.com">http://www.informatica.com</a>
Microsoft	<a href="http://www.microsoft.com">http://www.microsoft.com</a>
MicroStrategy	<a href="http://www.microstrategy.com">http://www.microstrategy.com</a>
Oracle	<a href="http://www.oracle.com">http://www.oracle.com</a>

**Table 63: Most frequently encountered BI-vendors, as of 2002**



## Conclusions

In the economy of today, companies are much too careful to go investing tons of money in applications like the BI-tools discussed in this paper. For the moment of today and tomorrow on the short term, I do not expect much development to take place. Many organizations have put the plug in IT-vacancies. Budgets have to be cut, and the first to go are all those pricey IT-consultants. The hype around the emergence of Information Technology has grown so out of control that companies are recoiling one after the other. And what BI-products can be sold if there are no consultants to sell them? And what BI-products can be bought if there is no money to buy them?

Many companies are still working hard on the road to OnLine Analytical Processing, which will probably keep on being redefined every time technologies are altered or added.

On the long term the World Wide Web will keep on growing, and with it the need, or should I say the wish, to keep storing data and information in structured ways, in order to gain as much benefit from the extracted knowledge as possible. In the area of Data Mining especially concepts like Customer Profiling will stay popular, because in the end it will always be rewarding to keep on knowing who your most profitable customers are.

The critical reader will have noticed that in this report a lot terms, concepts, etc. have appeared in more than one section. This might seem confusing, but the borderlines are so thin and vague that in some cases it was just unclear where exactly to place certain facts. That is why I found this quite a complex, but also a challenging topic to do research on. Most organizations selling BI-related products have definitions and descriptions ready to explain to their customers what their products involve. However, hardly any two definitions or descriptions are the same. The different technologies of data analysis are often all named in the same breath. If an organization is specialized in Reporting, it can be expected that they are also specialized in Querying and Balanced Scorecards. Often Reporting is part of OLAP-activities. In fact, it is almost out-of-date to provide Reporting for one or two dimensions; the real challenge lies in the multidimensional.

The moment you have to round off an investigation like this one, there will always be more questions for which there is no time to find the answers. This is quite healthy, but also somewhat a pity. Further research I would like to have done is find out *how* the OLAP-applications can be used (i.e. take a look at demos), and not only list them as I have done in Chapter 4.

What I would really have liked to do is supplement this investigation of literature with a practical investigation, in order to find out to what extent all these products have penetrated the Dutch market. It is striking how many English/American companies there are in this sector, but how far are we in The Netherlands? The conclusions I can give here are merely my own ideas based on what I have read about Business Intelligence. I would like to know whether these ideas are correct, or whether I have a wrong picture or have missed vital information.

In any case, this research study is not complete. As time progressed I kept coming across more and more interesting sites and companies, often just by chance whilst searching for something else. So many of them had interesting things to say about BI. When I thought I had finished a chapter I kept bumping into information that could or maybe even should still be added to that chapter. So please forgive me if I have left out a company, name, publication or vendor that turns out to be a very important one. I couldn't possibly catch them all!

One thing is for sure. If you ever forget what exactly Business Intelligence is, just remember that 'it covers a whole range of concepts': it is The Umbrella Term!





---

## **Appendix A Links to the World Wide Web**

### ***People's home pages***

Ralph Kimball: <http://www.rkimball.com>  
Bill Inmon: <http://www.billinmon.com>

### ***Company home pages***

Brio Technology: <http://www.brio.com>  
Business Intelligence Technologies: <http://www.businessintelligencetechnologies.com>  
Business Objects: <http://www.businessobjects.com>  
Cognos: <http://www.cognos.com>  
CorVu: <http://www.corvu.com>  
Crystal Decisions: <http://www.crystaldecisions.com>  
Gartner Group: <http://www.gartner.com>  
Hummingbird: <http://www.hummingbird.com>  
Hyperion: <http://www.hyperion.com>  
IBM: <http://www.ibm.com>  
Microsoft: <http://www.microsoft.com>  
MicroStrategy: <http://www.microstrategy.com>  
OLAP Council: <http://www.olapcouncil.org>  
OLAP Solutions: <http://www.olapsolutions.co.uk>  
Oracle: <http://www.oracle.com>  
SAS Institute: <http://www.sas.com>

### ***Pages for finding journals and magazines on BI***

BI Quarterly: <http://www.biq.nl???>  
Database Magazine: <http://www.array.nl/dbm>  
Datawarehouse Infocenter: <http://www.dwinfocenter.org/periodi>  
Intelligent Enterprise: <http://www.intelligententerprise.com>  
KDnuggets: <http://www.kdnuggets.com/publications/index.html>

### ***Conference proceedings***

KDD Conference Proceedings: <http://www.kdnuggets.com/publications/sigkdd.html>

### ***Papers and articles***

The OLAP Report: <http://www.olapreport.com>  
Datawarehouse Infocenter: <http://www.dwinfocenter.org/whitepap>  
Hyperion White Papers: <http://www.hyperion.com/products/whitepapers/>  
IBM White Papers: <http://www.microstrategy.com/Publications/Whitepapers/>  
MicroStrategy White Papers: <http://www-3.ibm.com/software/data/pubs/papers/>



***Other information sites***

Business Intelligence for Excel            <http://www.bixl.com>  
Decision Support Systems Resources:   <http://www.dssresources.com>  
Oracle9i Data Warehousing Guide:      [http://download-west.oracle.com/otndoc/oracle9i/901\\_doc/server.901/a90237/title.htm](http://download-west.oracle.com/otndoc/oracle9i/901_doc/server.901/a90237/title.htm)

***Competitive Intelligence***

Leonard Fuld:                                <http://www.fuld.com/whatCI.html>  
CI Academy:                                  <http://www.academyci.com>  
CI Solutions:                                 <http://www.competitive-intelligence.co.uk>



## **Appendix B      List of abbreviations**

<b>AI</b>	Actionable Intelligence
<b>ASP</b>	Application Service Provider
<b>BI</b>	Business Intelligence
<b>BSC</b>	Balanced Score Card
<b>CEO</b>	Chief Executive Officer
<b>CI</b>	Competitive Intelligence
<b>CIO</b>	Chief Information Officer
<b>CKO</b>	Chief Knowledge Officer
<b>CRISP</b>	Cross Industry Standard Process
<b>CRM</b>	Customer Relationship Management
<b>CSS</b>	Cascading Style Sheet
<b>DDBMS</b>	Distributed Data Based Management System
<b>DES</b>	Data Encryption Standard
<b>DM</b>	Data Mining
<b>DMS</b>	Document Management System
<b>DSS</b>	Decision Support System
<b>DW</b>	Data Warehousing
<b>EIS</b>	Enterprise Information System
<b>EIS</b>	Executive Information System
<b>EMS</b>	Electronic Meeting System
<b>ERP</b>	Enterprise Resource Planning
<b>ES</b>	Expert System
<b>ESS</b>	Executive Support System
<b>ESS</b>	Expert Support System
<b>ETL</b>	Extraction, Transformation & Loading
<b>ETS</b>	Expertise Transfer System
<b>FASMI</b>	Fast Analysis of Shared Multidimensional Information
<b>GDSS</b>	Group Decision Support System
<b>GIS</b>	Geographic Information System
<b>GSS</b>	Group Support System
<b>ICT</b>	Information & Communication Technology
<b>ITS</b>	Intelligent Tutoring System





<b>KAT</b>	Knowledge Analysis Tool
<b>KD</b>	Knowledge Discovery
<b>KDD</b>	Knowledge Discovery in Databases
<b>KMS</b>	Knowledge Management System
<b>KWS</b>	Knowledge Work System
<b>MIS</b>	Management Information System
<b>MRP</b>	Materials Requirements Planning
<b>MSS</b>	Management Support System
<b>OAS</b>	Office Automation System
<b>OLAP</b>	OnLine Analytical Processing
<b>OLE DB</b>	Object Link Embedding Data Base
<b>OOMBMS</b>	Object-Oriented Model Based Management System
<b>PRM</b>	Partner Relationship Management
<b>Q&amp;R</b>	Queries & Reports

**Table B1: List of abbreviations**





**Literature**

**Reference List**

Legend:  
(B) = available as a book  
(I) = article or white paper found on the internet  
(P) = on paper, copied from DB/M or got from prof. Eiben

Alter, Steven  
*Information Systems: a management perspective*, Addison Wesley Longman, 1999 (B)

Brant, Klaas  
*De reus is wakker*  
Database Magazine (DB/M), no. 6, 1999, p. 35 (P)

Business Intelligence Technologies, Inc.  
*Business Intelligence for Excel White Paper*, 2002  
<http://www.bixl.com/BIXLwhitepaper.pdf> (I)

Buytendijk, Frank A.  
*Het walhalla van de informatie-democratie* (een interview met Howard Dresner),  
Database Magazine (DB/M), no. 8, 1997, p. 35 <http://www.array.nl/dbm/art9708/howarddresner.htm> (I)

Buytendijk, Frank A.  
*Business Intelligence sterk in beweging*  
Database Magazine (DB/M), Database Systems, maart 2001, p. 61 (P)

Dekker, Hans  
*Business Intelligence*, June 2002, Emerce  
<http://www.emerce.nl/archives/magazine/juni2002/Dossier/13997.html> (I)

Forsman, Sarah  
*OLAP Council White Paper*, 1997, OLAP Council  
<http://www.olapcouncil.org/research/whtpapy.htm> (I)

Hamer, Pieter den  
*Van datawarehouse naar knowledge warehouse?*  
Database Magazine (DB/M), no. 6, 1998, p. 12 (P)

Hashmi, Naeem  
*BI for sale*, 2000, Information Frameworks  
<http://www.intelligenterp.com/feature/2000/10/hashmiOct20.shtml> (I)

Hermelink, Judith & Bilsen, Joost van  
*IT bepaalt de score*  
Database Magazine (DB/M), no. 8, 2000, p. 13 (P)

Joosse, Carine  
*Datamining in het kielzog van Internet*  
Database Magazine (DB/M), no. 6, 2000, p. 13 (P)

Laudon, Kenneth C. & Laudon, Jane P.  
*Management Information Systems: organization and technology in the networked enterprise*, Prentice Hall, 2000 (B)



- Lewis, William J.  
*Data Warehousing and E-Commerce*, Prentice Hall, 2001 (B)
- Pendse, Nigel  
The OLAP Report: OLAP Applications, August 2001, Business Intelligence Ltd.  
<http://www.olapreport.com/Applications.htm> (I)
- Pendse, Nigel  
*The origins of today's OLAP products*, July 20 2002, The OLAP Report  
<http://www.olapreport.com/origins.htm> (I)
- Putten, Peter van der  
Sentient Machine Research, CMG Academy 1999  
(slides gekregen van Prof. Eiben) (P)
- Simon, Alan R. & Shaffer, Steven L.  
*Data Warehousing and Business Intelligence for e-Commerce*,  
Morgan Kaufmann, 2001 (B)
- Turban, Ebrahim & Aronson, Jay E.  
*Decision Support Systems and Intelligent Systems*, Prentice Hall, 2001 (B)
- Voorma, Onno  
*Op zoek naar 'actieve' kennis voor de bedrijfsstrategie*  
Database Magazine (DB/M), no. 5, 2001, p. 28 (P)
- Whatis.com  
Whatis.com, June 2001  
[http://searchcrm.techtargt.com/sDefinition/0,,sid11\\_gci213571,00.html](http://searchcrm.techtargt.com/sDefinition/0,,sid11_gci213571,00.html) (I)
- Witten, Ian H. & Frank, Eibe  
*Data Mining: Practical Machine Learning Tools and Techniques with  
Java Implementations*, Morgan Kaufmann, 1999 (B)
- Wu, Jonathan  
*Calculating ROI for Business Intelligence Projects*, 2000, BASE Consulting Group  
[http://www.baseconsulting.com/Assets/applets/Calculating\\_ROI.pdf](http://www.baseconsulting.com/Assets/applets/Calculating_ROI.pdf) (I)



**Additional literature**

- Buytendijk, Frank A.  
*Een glazen bol voor de kubus* (een interview met Nigel Pendse),  
Database Magazine (DB/M), no. 3, 1997, p. 15  
<http://www.array.nl/dbm/art9703/pendse.htm> (I)
- Buytendijk, Frank A.  
*De BI schokwave*  
Database Magazine (DB/M), no. 4, 1999, p. 23 (P)
- Dresner, Howard  
*Business Intelligence in 2002: A Coming of Age*, 2001, Gartner Group  
<http://www.gartner.com/resources/103200/103282/103282.pdf> (I)
- Eiben, A.E.  
*Introduction to Business Intelligence* (lecture slides), Course on Business Intelligence,  
Free University Amsterdam (P)
- Eiben, A.E.  
*Three approaches to DSS: Operations Research, Artificial Intelligence, and Business Intelligence* (lecture slides), Course on BI, Free University Amsterdam (P)
- Gallas, Susan  
*Kimball vs. Inmon*, September 1999, dmDirect  
[http://www.dmreview.com/editorial/dmdirect/dmdirect\\_article.cfm?EdID=1400&issue=090199&record=3](http://www.dmreview.com/editorial/dmdirect/dmdirect_article.cfm?EdID=1400&issue=090199&record=3) (I)
- Habers, Frank & Dieleman, Peter  
*Hoeveel sterren krijgt uw rapportagetool?*  
Database Magazine (DB/M), no. 1, 2001, p. 20 (P)
- Information Technology Toolbox, Inc.  
*Business Intelligence Vendor Directory*, 1998-2002, Information Technology Toolbox, Inc.  
<http://businessintelligence.ittoolbox.com/vnd.asp> (I)
- Inmon, W.H.  
*Building the Data Warehouse*, 3<sup>rd</sup> edition  
John Wiley & Sons, 1996 (B)
- IntelliBusiness Inc.  
*DataWarehousing Online.com*, 2001, IntelliBusiness Inc.  
<http://www.datawarehousingonline.com> (I)
- Kimball, Ralph  
*The Data Warehouse Toolkit: The Complete Guide to Dimensional Modeling*, 2<sup>nd</sup> edition  
John Wiley & Sons, 2002 (B)
- Kimball, Ralph  
*The Data Warehouse Toolkit: Practical Techniques for Building Dimensional Data Warehouses*  
John Wiley & Sons, 1996 (B)



- Kimball, Ralph  
*The Data Warehouse Lifecycle Toolkit: Tools and Techniques for Designing, Developing and Deploying Data Warehouses*  
John Wiley & Sons, 1997 (B)
- Kimball, Ralph  
*The Data Webhouse Toolkit: Building the Web-enabled Data Warehouse*  
John Wiley & Sons, 2000 (B)
- Koenderman, Mark & Kiewiet, Henk-Jan & Gulick, Katja van  
*Kiezen uit een stortvloed aan scorecard-tools*  
Database Magazine (DB/M), no. 1, 2001, p. 41 (P)
- Koning, Annelies & Koekkoek, Marcel  
*Goedkoop of duurkoop?*  
Database Magazine (DB/M), no. 6, 2000, p. 37 (P)
- Lek, Harm van der  
*Business Intelligence: wat doen bedrijven nu echt?*  
Database Magazine (DB/M), no. 7, 1999, p. 34 (P)
- Linden, Paul van der  
*Business Objects loopt voorop met ‘e-BI’*  
Database Magazine (DB/M), no. 5, 2001, p. 43 (P)
- Linden, Paul van der  
*‘Eén beeld van de waarheid’ met MicroStrategy 7i*  
Database Magazine (DB/M), no. 4, 2002, p. 29 (P)
- OLAP Council  
OLAP and OLAP Server Definitions, 1997  
<http://www.olapcouncil.org/research/glossaryly.htm> (I)
- Pendse, Nigel  
The OLAP Report: What is OLAP?, July 27 2002, Business Intelligence Ltd.  
<http://www.olapreport.com/FASMI.HTM> (I)
- Power, D.J.  
*A Brief History of Decision Support Systems*, DSSResources.COM, World Wide Web,  
version 2.0, 2002  
<http://dssresources.com/history.dsshhistory.html> (I)
- SDG Computing, Inc.  
*The Business Intelligence and Data Warehousing Glossary*, 1995-2002  
<http://www.sdgcomputing.com/glossary.htm> (I)
- Vos, Albert H.  
*OLAP – Online Analytical Processing*, Rapport, March 25 1997, TUDelft  
Reference: <http://www.io.tudelft.nl/education/ide441/archief/9697/905308.html> (I)
- Winter, Richard  
*Field Experience with Large Scale Data Warehousing on Oracle*, March 2002, Winter  
Corporation, Waltham, MA.  
Reference: [http://www.oracle.com/ip/collateral/vldw\\_winter.pdf](http://www.oracle.com/ip/collateral/vldw_winter.pdf) (I)