# Chapter 17
# Evolution of Business Intelligence

**W.H. Inmon**

**Abstract** From the first simple report to data warehousing to BI tools to today's evolved state of intelligence, BI continues the evolution. Once BI was for structured data only. Now BI can operate on textual data, and in doing so BI can operate on the full spectrum of data found in the corporation.

## 17.1 Introduction

There are three constants in life—death, taxes, and change. Change occurs in the form of evolution. Some evolutions occur at glacial speeds. The formation of the continents—the grating of tectonic plates that cause earthquakes and volcanic eruptions—are recognizable today, although at very slow speeds. Other evolutions occur much more quickly. The evolution of the automobile began at the turn of the 20th century and continues today. From the black, boxy model T to today's Hybrid or Porsche, the automobile has evolved in front of our very eyes, in memorable history.

## 17.2 Evolution of the Cell Phone

Evolving even faster has been the evolution of the personal telephone. In the 1950's there was no cell phone. In those early days, people had black dial up phones with wall mounted cords for the most part. There were party lines, exchanges and expensive long distance phone calls. Then in the 1980 timeframe, there first appeared the cell phone. The initial purpose of the cell phone was to call home. Businesspeople on the road now had a means of staying connected to home and headquarters.

In the very early days of cell phones, the cost was high, the coverage was spotty, and the quality of the connection was questionable. But there was a very real marketplace and the evolution of the cell phone began. Over time the cost and size of

W.H. Inmon
Inmon Consulting Services, Castle Rock, CO, USA
url: http://www.inmoncif.com

P. Rausch et al. (eds.), *Business Intelligence and Performance Management*,
Advanced Information and Knowledge Processing,
DOI 10.1007/978-1-4471-4866-1_17, © Springer-Verlag London 2013

cell phones dropped. Over time the coverage of cell phones increased. Over time the quality of the service and the reliability of the connection improved. Today, the cellular industry looks vastly different than the cell phone industry of just a few short years ago. The evolution of the cell phone has come at a whirlwind pace.

But there was another most interesting aspect to the evolution of cell phones and cell phone technology. That interesting aspect was the increase in functionality of the cell phone device. Today calling home is only one rather limited aspect of what cell phones do. Today cell phones:

- can be used as a camera
- can be used as a recording device
- can be used for Instant messaging
- can be used for amusement, with all sorts of games
- can be used for storing and retrieving business information
- can be used for managing a calendar
- can be used for handling alerts
- can be used for many other purposes.

And amazingly all this diverse functionality is found in a single device, in a single place. In truth, handling phone calls is only one small function among many that has evolved over time in the cell phone. And all of this functionality is passed through one sophisticated, electronic device.

## 17.3 Evolution of Business Intelligence

Now consider the evolution of Business Intelligence (BI). The origins of BI go back to the humble application report, generated in the days of COBOL and assembler processing. The application report purported to tell the organization what was happening in the environment. While the humble COBOL or assembler report served a very real purpose, there were many problems with application reports. Application reports chewed up a lot of paper. Application reports took a long time to run. Application reports were usually out of date by the time they were printed. But application reports were a start.

Soon there were online transactions which also told the organization what was occurring, but in real time. Online transactions took no paper. Online transactions showed data that was up to date as of the moment of access. And online transactions were fast. But online transactions had their limitations, too. Online transactions were not good for showing large amounts of data. Online transactions did not leave an auditable trail. And the data behind online transactions often changed by the second. A person could do an online transaction only to have the information invalidated in the next second. And online transaction processing systems were expensive and fragile.

But the real problem behind online transactions was that online transactions showed only a limited type of data and showed data that was unique to an application. Given enough online systems, it was possible to find the same piece of data with multiple different values coming from multiple different applications. User A looks for some information and finds it. User B looks for the same information in another system, and finds it. However the values shown to user A are not the values shown to user B.

The value of integrating data into a corporate format began to be obvious. Decision making in a world where there was no definitive source of data became a challenging and dicey exercise. For large corporations, it became obvious that a historical, integrated, subject oriented source of data was needed for the corporation. Having multiple overlapping applications simply was not a sound basis for business decisions. Thus born was the data warehouse. With the data warehouse it was now possible to do a whole new style of reporting and analysis. Once there was definitive corporate data in a data warehouse, the world of BI began, at least as we know BI today.

## 17.4 Enter Business Intelligence

In the early days of data warehouse and BI, doing simple analysis was a breath of fresh air compared to the world that existed before data warehouse and BI. Simple reporting from data warehouses was the first infant step of BI. But people quickly discovered that there were many other possibilities for BI. Many other functions belonged to the purview of BI.

Some of the newly discovered functions included:

- graphical visualization of results
- looking at information over time
- looking at very large volumes of data
- doing statistical analysis of data
- looking at small subsets of data in a personalized fashion
- using spreadsheets to analyze data
- people building customized collections of information to suit their individual needs called data marts
- collecting information about what analysis has already been done
- collecting metadata so that an analyst knows where to begin in doing an analysis
- transforming data so that different sources of data can become integrated, and so forth.

Indeed once BI became a reality, all sorts of activities occurred. Unfortunately there was no coordination of these BI activities. One department would run a report. Another department would do a projection. One individual would create a spreadsheet. Another analyst would create metadata by building a data set. In many ways, the world of BI resembled the California gold rush. One individual or one organization worked in their own self interest with no concern for the work being done by others. The California gold rush and BI resembled a colony of bees in the springtime. There was little or no apparent coordination of people working in tandem with each other.

The activities of BI soon encompassed lots of different types of technology. There certainly was reporting. There was graphical software. There was ETL. There were spreadsheets. There was statistical processing, and so forth. And sitting on top of this beehive of activity was no organization or coordination of the different activities of BI.

## 17.5 The Evolution to Textual ETL

But the evolution of BI is ever evolving. There is another important aspect of the evolution of BI, and that aspect is the evolution to unstructured, textual data. Consider this—most organizations build data bases based entirely on the basis of what is termed structured data. Structured data refers to data that occurs repetitively. Consider banking transactions. For the most part all banking transactions are the same, insofar as the processing that occurs is concerned. The only real difference between one banking transaction is the date, the account number, the amount of the transaction, and the parties involved in the transaction. Other than those differences, there really isn't any difference between one banking transaction and the next. And transaction processing occurs everywhere. Airlines do transactions. Retailing does transactions. Insurance does transactions. Manufacturing does transactions, and so forth. In one way or the other, all businesses do transactions as a part of their day to day processing.

Modern data base management systems (dbms) are geared to handle repetitive transactions. Dbms lay out data and the repeated occurrences of data that are generated by a transaction are recorded by a record or a row created by the dbms. Dbms are designed to handle efficiently many, many transactions and many transaction types.

It comes as a surprise to many people that in most corporations the majority of data in the corporation is not repetitive transaction based data. It is estimated that approximately 80 % of the data in the corporation is unstructured textual data, not transaction based, repetitive data. (This is a surprise to the IT professional who has spent his/her entire life working with transaction based data.) So where is this unstructured textual data found? It is found in many places. Some of those typical places are:

- in email
- in contracts
- in human resource files
- in warranties
- in chat log/help log sessions
- in medical records
- in loan applications
- in customer responses, and so forth.

In short unstructured, textual data is found everywhere. It is—in a word—pervasive.

## 17.6 Textual Data Has Great Business Value

And there is much textual data that is very important to the corporation. Corporate contracts contain a wealth of valuable information to the corporation. Corporate contracts represent legal obligations either to the corporation or by the corporation. Chat log sessions represent the direct interface with the customer. In addition, chat log sessions contain invaluable information about products, services and defects or customer complaints. Medical information contains huge amounts of information about disease, treatments, therapies, medications, and so forth. Warranty claims contain important information about the quality of parts and products. Insurance claims contain important information about fraud. And the list goes on. In fact there is probably more important information about the corporation found in text than there is in transaction based information. Yet, because textual information is not repetitive, it does not fit well with dbms, and as a consequence is not used in the decision making of the corporation.

BI continues its evolution and an important part of the evolution is the ability to start to read textual data, transform it, and move the text into a standard relational data base. Once in a standard relational data base, text is able to be analyzed like any other source of data.

## 17.7 Integrating Text

The process of reading and analyzing text is called the process of “integrating text”. Unlike a search engine, textual integration starts with the assumption that the raw text needs to be changed. Search engines and data mining processes make the basic assumption that raw text should either be changed not at all or should be changed only very lightly. Textual integration on the other hand starts with the assumption that raw text needs to be heavily changed before it is fit to be placed into a data base.

The process of integration is done by (patented) technology called “textual ETL” such as that sold and supported by Forest Rim Technology. Textual ETL reads raw text, does the integration, and produces the output into a standard relational data base such as Oracle, Teradata, DB2/UDB, SQL Server, or other dbms. Once the data has been placed in other BI technology, standard analytical processing can be done against the text.

## 17.8 Some Differences

There are some fundamental differences between processing raw text and processing repetitive transaction data. One of those differences is the ongoing nature of processing. Repetitive transactions require constant and ongoing processing. As long as the bank is doing transactions, those activities generate data that must be placed in a data warehouse. But much of textual ETL processing against text is of a one time nature. Once a contract is completely and thoroughly processed, there is no need to go back and reprocess it. Of course if the contract is updated, reprocessing is necessary. But if the contract is updated, it is in fact a new contract. As a rule people don't update or change documents. People create new documents, but it is not normal for a document to be updated. If an email says something incorrectly, a new email is written. If a contract specifies something improperly, a new contract is written. If a chat log session discusses information incorrectly, a new conversation is held.

Because updates of text are not the norm, there is no need to constantly process and reprocess the same document.

Another difference in transaction processing and document processing is the volume of data that must be processed. While there certainly are environments that must process large amounts of transactions, the transactions enter the system on a finite basis. There may be lots of transactions, but as a rule the transactions are relatively small.

When it comes to processing documents, there may be MASSIVE volumes of data. Take loan application portfolios, for example. A given loan may be up to 250 pages in length. The portfolio may contain 2,000 or more loans, and there may be many, many portfolios. In all, there may be multiple petabytes of text associated with loan applications to be processed.

Another factor when dealing with the textual ETL processing of text is the fact that text comes in many different forms. Classical text is in the form of proper English grammar. This paper (hopefully!) is written in proper English grammar. There are verbs, adverbs, nouns, adjectives, pronouns and so forth. Words are spelled properly. There are periods, question marks, and exclamation points. There is a proper structuring of the words in the sentence. Some parts of the text are capitalized. Those are all the earmarks of proper English sentences. And proper English sentences should certainly be able to be processed by textual ETL.

But text comes in other forms as well. There are doctor's notes. Doctor's notes have their own shorthand—both in terms of spelling and in terms of structure. "HA" may mean "headache". Bp 120/67 may mean "blood pressure" of 120 systolic over 67 diastolic. Doctors just don't have the time or take the time to write proper grammar. And as long as the information is for the doctor's personal usage, shorthand and comments are just fine.

Teenagers write in "IM" or instant messaging. "THX" may mean "thanks". "2" may mean "to". "U" may mean "you". And textual ETL needs to be able to handle ALL forms of text.

In addition, occasionally textual ETL comes in the form of different languages. An international banking organization may do business in Spanish, English, and Russian. But the data base analyst must have all of the data base written in the same language for the purpose of doing analytical processing.

There are then some major challenges in the handling of text. Text is not text. Text takes many different forms and textual ETL needs to be able to handle all of those forms.

One of the real values of textual ETL and the integration and assimilation of text into a data base is that—once committed to a data base—the textual data can be queried and analyzed along side of classical structured data. The ability to read and analyze both structured and unstructured data together is powerful. Entirely new kinds of analysis can be done that simply are not possible when there is only structured data that can be analyzed.

## 17.9 Summary

From the first simple report to data warehousing to BI tools to today’s evolved state of intelligence, BI continues the evolution. Once BI was for structured data only. Now BI can operate on textual data, and in doing so BI can operate on the full spectrum of data found in the corporation.