

Chapter 6

Future Developments and Trends

1. Introduction

Data warehousing is a form of technology identified as one of the desired applications needed by organisations today and in the future (Du Plessis, 1998: 1). It is important that management and the internal auditor are aware of expected developments in the data warehouse environment. This will enable them to manage internal control risks.

Another significant future trend is the increase in utilisation of data warehousing technology by internal auditors. Any tool that allows the internal auditor to identify risks more effectively can result in the internal audit department being more efficient (Geiger, 1997: 31). Internal auditors use data warehousing as a means of assessing internal controls within other audit cycles. Routine audit cycles may include: accounts payable and receivable reviews, marketing and advertising assessments, fraud investigations, etc.

2. Aim

This chapter identifies what future data warehouse developments can be expected and what affect these changes could have on internal control risks.

In addition to this, the study also provides reasons why the data warehouse environment can improve the efficiency of internal auditing when evaluating other routine cycles. This portion of the study focuses specifically on the data mining technique which allows the internal auditor to effectively interrogate small or large volumes of data when reviewing other audit cycles.

The chapter concludes by referring to the results of the empirical survey.

3. Future developments and the effects on internal control risk

The following areas have been identified as future developments which could significantly change existing data warehouse environments (DCI, 1998: 1-3). In addition to describing the development, the study also identifies what affect these changes could have on internal control risks:

3.1 Closed-loop business performance management

3.1.1 Definition

This development will result in the data warehouse providing management with information on a real-time basis. Rather than the user requesting data trends, information systems will smartly identify those data elements of defined interest and provide results timeously to the user without any prompting.

3.1.2 Internal control risks

- Although real-time data provision allows for timeous decision making, it can result in the user being inundated with unnecessary data. This will occur in instances where the criteria used in identifying reported trends are not established according to stringent management standards.
- If not properly managed and monitored, real-time data provision can result in the over utilisation of system resources.

3.2 Increased access to data warehouse information

3.2.1 Definition

Increased access to the data warehouse by a larger percentage of the organisation's personnel is another expected development. In addition to personnel access, it is expected that the data warehouse will also be accessible by suppliers and customers in the future.

3.2.2 Internal control risks

- The risk of confidential information being disclosed to unauthorised users may be due to incomplete or inconsistent application of access rights procedures and policies.
- Increasing the number of users who have access rights to the data warehouse environment can negatively impact the overall performance of environment. This risk is increased if ongoing monitoring and the necessary upgrading of software and hardware is not conducted.

3.3 Removal of source data quality problems automatically

3.3.1 Definition

Data quality will improve as methods are developed to move from unclean data detection to the automatic cleansing of source quality problems up-front by the data warehouse interfaces and applications.

3.3.2 Internal control risk

The fact that any procedure which will be able to correct source data automatically within source applications without any form of stringent management control and audit trail is concerning. Such automatic procedures could result in unauthorised changes being made to source data.

3.4 Re-engineering the development methodology

3.4.1 Definition

Data warehouse requirements have been defined after the development of the associated source systems. It is also expected that the future source systems will be developed in such a manner that limited interfaces and data cleansing will be required.

This can be achieved by ensuring that data warehouse requirements are considered as part of the development methodology for traditional applications.

3.4.2 Internal control risk

There is a risk that the development process may become so inwardly focused that system development methodologies will not consider end user needs as the primary input in the development process.

3.5 Transferring of report and query functionality

3.5.1 Definition

Reporting and other query functions included in the data warehouse's source applications will be removed and will become functionality provided by the data warehouse environment.

3.5.2 Internal control risks

- The data warehouse will need to be accessed by a far larger audience. Therefore the risk of unauthorised access to privileged information is increased.
- User training and data warehouse package licenses are necessary and will increase costs.

4. Utilisation of data warehouse technology by the internal auditor

4.1 Background

There has been an increase in the use of data warehousing by internal auditors when evaluating internal controls within other audit cycles (Geiger, 1997: 31-32). In this section we will consider the use of data warehousing by the internal auditor, define the concept of data mining and finally provide a framework on how data warehousing can be used by the internal auditor in extracting needed audit evidence.

The major reasons why this technology has in recent years received so much more attention is based on the following developments (Moxon, 1996: 2):

- There has been a movement towards identifying information as the key corporate asset.
- Data mining provides a tool which can navigate through data and provide the exact level of detail without the need for technical assistance.
- The dramatic increase in hardware power and cost reductions therefore has provided the internal auditor with the ability to analyse and evaluate large volumes of transactions.

It may seem initially that the data warehouse cannot be relied upon by the internal auditor because the data has been manipulated. To ensure that the integrity of information is in its most objective form, the data should be obtained as close to the source system as possible (Geiger, 1997: 31).

There are however situations where data warehouse technology can be utilised effectively by the internal auditor in evaluating internal control risks within other routine audit cycles (Geiger, 1997: 31-32):

- Gathering data from multiple sources and then combining the data to obtain sensible evidence can be time-consuming and frustrating for the internal auditor. In the case of a data warehouse, the combination of all strategic information has already been performed and is a source which can be relied upon. The internal auditor should however ensure that he/she verifies the integrity of data relied upon. Steps which are performed to verify the data include evaluating the process adopted in migrating the data from its source to the data warehouse and identifying what steps management have implemented to ensure data integrity.
- The data warehouse environment includes metadata, therefore it can assist the internal auditor in simplifying interrogation and analysis tasks by providing information on how similar data from various outside sources compare.
- Data is quality checked before being permitted access to the data warehouse environment. Therefore information on data rejected provides the internal auditor

with a good indication of possible system and operational processes which require future auditing.

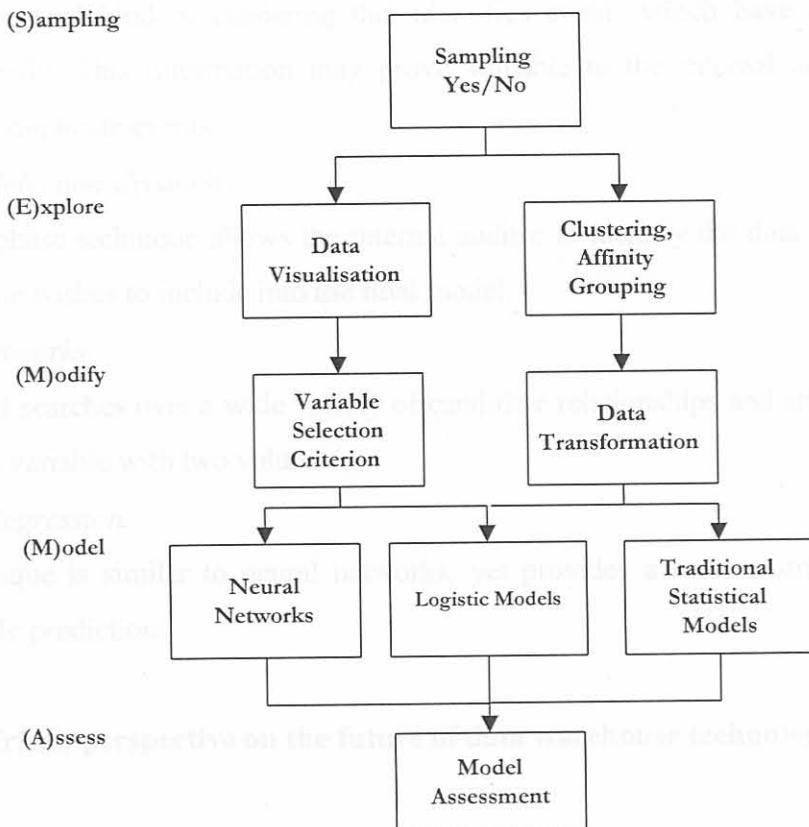
4.2 Defining data mining

Data mining is the automated analysis of customer transaction data for the purposes of discovering underlying trends. Data mining as with all other components of the data warehouse environment is incorporated into the Knowledge Discovery in Databases process (Rosen, 1997: 51).

4.3 Process followed in utilising data warehouse technology

Data mining is effectively applied when a consistent methodology of evaluation and interrogation is adopted. To ensure the reliability of results and provide the internal

Figure 6.1 - SEMMA data mining process



Source: Casarin, 1997: 44

auditor with credible audit evidence, it is vital that a suitable methodology is adopted. The SEMMA methodology developed by the SAS Institute has proved to be the most frequently used of these models. The SEMMA acronym is based on the SAS's five stages in the data mining process, viz. Sampling, Exploration, Manipulation, Modeling and Assessment (Casarin: 1997: 44).

Figure 6.1 depicts the SEMMA process. Details on the various terms used in the SEMMA process include (Casarin, 1997: 44-46):

- *Data Visualisation*

This involves representing data graphically to simplify evaluations.

- *Clustering*

This second phase function allows the user to examine large volumes of transactions and identify whether they can be grouped based on common criteria. This is also often termed “undirected data mining”, because the user has no predetermined objective and is hoping that the data mining tool will reveal significant trends.

- *Affinity Grouping*

This is a special kind of clustering that identifies events which have occurred simultaneously. This information may prove valuable to the internal auditor in identifying duplicate events.

- *Variable Selection Creation*

This third phase technique allows the internal auditor to identify the data elements which he/she wishes to include into the final model.

- *Neural Networks*

This model searches over a wide variety of candidate relationships and attempts to highlight a variable with two values.

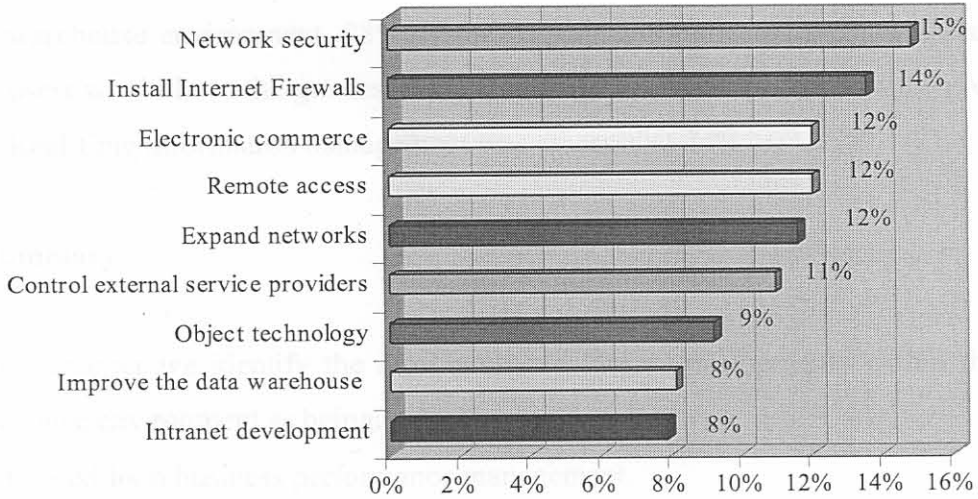
- *Logistic Regression*

This technique is similar to neural networks, yet provides a more restricted and interpretable prediction.

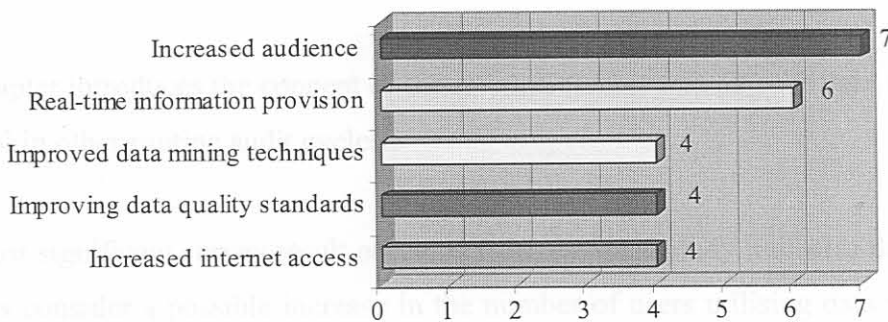
5. A South African perspective on the future of data warehouse technology

The results of the local survey are featured below. The results relate specifically to the future trends within data warehouse technology:

1. Which of the following areas have been identified as the major risk applications and/or IT related systems within the next two years?



2. Which of the following data warehouse developments do you think will significantly affect your future assessment of the control environment?



The most significant findings include:

- 15% of the respondents indicated that they expected network security to be the greatest area of concern within the next two years. Electronic commerce, remote access and expanding networks were rated at 12% individually. 8% of the

respondents felt that the data warehouse environment would be a significant area of concern in the foreseeable future. The most probable explanation for this low rating could be that management have not identified the data warehouse environment as a critical area as yet.

- Internal audit managers were requested to indicate which specific data warehouse elements they felt would significantly impact their assessment of the data warehouse environment. 28% of the respondents indicated that the increase in users would have the greatest impact on their assessment of internal control risk. Real-time information management was assessed at 24%.

6. Summary

In this chapter we identify the most probable future developments within the data warehouse environment as being:

- Closed-loop business performance management.
- Increased access.
- Improved data standards.
- Re-engineering the development methodology.
- Transferring of report and query functionality.

An explanation of the possible affect these developments could have on internal control risk was also provided.

The chapter introduces the concept of the internal auditor utilising the data warehouse as a tool in other routine audit cycles.

The most significant survey result obtained from the local study indicates that internal auditors consider a possible increase in the number of users utilising data warehouse technology as the factor which would most affect their assessment of internal control risk within the data warehouse environment.

7. Conclusion

The results obtained from the empirical survey indicate that the improvement of the data warehouse environment is considered to be the eighth most significant high risk area out of a possible nine selections. It is therefore imperative that internal audit teams ensure that they are equipped with suitable knowledge to evaluate the impact these future developments might have on the data warehouse environment. Only through close liaison with the management team and end user, will the internal auditor be able to effectively assess the changing data warehouse environment and provide value adding input to the organisation's operations.