Annexure 1
Data Warehousing Survey Questionnaire

Name of person completing the questionnaire		
Designation		
Organisation's name		
Your office telephone number		
Organisation's staff size (*)	1-500	
	500-1500	
	1500-3000	
	3000-5000	
	5000-10000	
	> 10000	

(*) - Please mark with an "X" in the appropriate box alongside the correct value.

Completion instructions:

1. *Please note that no individual's name or title will be disclosed in the final results presented as part of the thesis submitted to the University of Pretoria, South Africa.*
2. *Please mark the appropriate answer with an X in the last column of the tables below except for the last question.*
3. *If your organisation does have a data warehouse/data mart structure in place, please complete sections A, B, C of the questionnaire. If however you do not have a data warehouse structure in place, but are planning on implementing such a structure, please complete Section B and C of the questionnaire.*
4. *Data warehouses and data marts are considered synonymous for the purposes of this questionnaire.*

Section A: An Already Implemented Data Warehouse Environment

Question	*Possible Answers*	*(X)*
1. For how long has your organisation had a data warehouse structure in operation?	Less than 1 year	
	1-2 years	
	2-5 years	
	5-10 years	
Comments?		
2. Which levels of staff within your organisation utilise the data warehouse structure (more than one reply is acceptable)?	Director level	
	Senior manager	
	Supervisory staff	
	Line personnel	
	Not sure	
Comments?		
3. Is the data warehouse environment identified as an application reviewed by internal audit on a periodic basis?	Yes	
	No	
Comments?		

Data Warehousing Survey Questionnaire
(continued)

Question	*Possible Answers*	*(X)*
4. If yes to question 3: How often will the data warehouse environment be audited by internal audit or appointed consultants?	Every year	
	Every 2 years	
	Every 3-5 years	
	Management request	
Comments?		
5. If yes to question 3: Which aspects of the data warehouse environment will be reviewed as part of the internal control review (definitions of the criteria used are included at the end of the questionnaire - more than one reply is acceptable)?	Confidentiality of information	
	Integrity of information	
	Reliability of Information	
	Availability of Information	
	Efficiency with which data is introduced into the data warehouse	
	Effectiveness of information in attaining management's initial goals	
	Compliance with data standards	
	Other (please include under comments)	
Comments?		
6. After completion of the post implementation review, did management realise the expected benefits of the data warehouse (if partially, express as a percentage of total under comments)?	Yes, fully	
	Partially	
	No	
Comments?		
7. Following on from question 6: If management did not realise/partially realise the expected benefits of the data warehouse, which of the following was considered the major cause for the failure?	Unrealistic expectations	
	Lack of sufficiently qualified personnel	
	Lack of funds	
	Unrecoverable/incomplete data from feeder systems	
	Other (please include under comments)	
Comments?		

Data Warehousing Survey Questionnaire
(continued)

Question	*Possible Answers*	*(X)*
8. Was the data warehouse implementation completed on time?	Yes	
	No	
Comments?		
9. If no to question 8: What was the major cause for the implementation not meeting the expected deadline?	Additional functionality requested by users after scope approval	
	Lack of qualified personnel	
	Lack of funds	
	Insufficient buy in from users	
	Unexpected system complexity	
	Other (please include under comments)	
Comments?		

Data Warehousing Survey Questionnaire
(continued)

Section B: The Data Warehouse Development

This section applies to both organisations that have completed their data warehouse development and those who are in the progress of developing such an environment.

Question	*Possible Answers*	*(X)*
1. Did the MIS/IT department develop a system development methodology specific for the data warehouse environment?	Yes	
	No	
Comments?		
2. On what basis were all possible "feeder" systems which could impact of the comprehensiveness of the data warehouse determined?	A source analysis prepared jointly by heads of department	
	Management knowledge	
	Data warehouse intended only for a single category of data	
	Other (please include under comments)	
Comments?		
3. What involvement will/does internal auditor play in the system development life cycle of the data warehouse?	Full project involvement	
	Integration testing only	
	Security aspects	
	No involvement	
	Other (please include under comments)	
Comments?		
4. What methodology was applied in ensuring that uniform data was introduced into the data warehouse environment?	A data quality standard was developed applied to each feeder system	
	Automated data extraction and conversion tools were utilised	
	Other (please include under comments)	
Comments?		
5. Was capacity management identified as an area which needed to be addressed as part of the development cycle?	Yes	
	No	
Comments?		

Data Warehousing Survey Questionnaire
(continued)

Question	*Possible Answers*	*(X)*
6. If yes to question 5: Which one of the following reasons was the major cause for including capacity management as part of the development cycle?	Scalability to handle future enhancements	
	Extensibility ensuring that the data warehouse can adapt to new hardware	
	Storage costs are monitored to verify return on investment	
	Other (please include under comments)	
Comments?		
7. What was management's major intention in implementing the data warehouse structure	Improved long-term focus	
	Empowering lower level management	
	Improved marketing strategy	
	Determining core competencies	
	Other (please include under comments)	
Comments?		
8. Will/has the data warehouse environment been included in the organisation's continuity plans/efforts?	Yes	
	No	
Comments?		
9. Will the data warehouse environment be distributed in nature?	Yes	
	No	
Comments?		

Section C: Audit Considerations

This section should be answered by all respondents.

Question	*Possible Answers*	*(X)*
1. What audit source materials were/will be used in formulating a suitable audit approach and program?	Data warehouse product manual	
	COBIT/COSO	
	Other ISACA/IIA source material	
	Internet sources	
	Other (please include under comments)	
Comments?		

Data Warehousing Survey Questionnaire
(continued)

Question	*Possible Answers*	*(X)*
2. Which of the following data warehouse developments do you consider will significantly affect your future assessment of the control environment (more than one reply is acceptable)?	Increased access via the internet and open communication channels	
	Real-time information provision	
	Increasing the audience of the data warehouse environment	
	Improving data quality standards	
	Improved data mining techniques for management use	
	Other (please include under comments)	
Comments?		
3. Which of the following areas have been identified as the major risk applications and/or IT related systems within the next two years (please use the scale 1 to 9. 1 being the most important and 9 being the least important).	Expanding network systems	
	Controls over external service providers	
	Deploying object technology	
	Implementing a data warehouse	
	Controls over electronic commerce	
	Deploying an intranet	
	Installing firewalls	
	Controls over remote access	
	Improve network security	
	Other (please include under comments)	
Comments?		

Data Warehousing Survey Questionnaire
(continued)

Criteria	*Definition*
Effectiveness	Deals with the information being relevant and pertinent to business process as well as being delivered in a timely, correct, consistent and usable manner.
Efficiency	Concerns the provision of information through optimal (most productive and economical) use of resources.
Confidentiality	Concerns the protection of sensitive information from unauthorised disclosure.
Integrity	Relates to the accuracy and completeness of information as well as to its validity in accordance with business values and expectations
Availability	Relates to information being available when required by the business process now and in the future. It also concerns the safeguarding of necessary resources and associated capabilities
Compliance	Deals with complying with those laws, regulations and internal quality standards for data.
Reliability of information	Relates to the provision of appropriate information for management to operate the entity and for management to exercise its financial and compliance reporting responsibilities.

Source: ISACA:1998.

Any other comments:

_____

_____

_____

_____

Would you like an e-mail copy of the thesis once completed? _____

If yes, please provide your e-mail address: _____

Thank you for taking the time in completing the survey. Please contact me if you require any other information.

Contact Person: *Sean de la Rosa*
Phone: *+27 11 490 0675*
Fax: *+27 11 493 1580*
Mobile: *+27 82 82 13471*
e-mail *sean.delarosa@afrox.boc.com*

Annexure 2

Strategic Data Warehouse Checklist

Criteria	*Guidelines*	*Response*
i. What are the number of internal source systems and databases?	Each source system along with their databases and files will take additional research, including meetings with those who have knowledge of the data. The time to document the results of the research and meetings should also be included in the time estimates.	
ii. How many business processes are expected for the data warehouse project (Examples: analyse sales, markets and financial accounts).	A project should be limited to just one business process. If management insist on more than one, the time and effort will be proportionally higher.	
iii. How many subject areas are expected for the project (Examples: customer, supplier / vendor, store/ location).	If possible, a project should be limited to just one subject area. If management insist on more than one, the time and effort will be proportionally higher.	
iv. Will a high level enterprise model identifying all possible systems be developed during the project?	Ideally, an enterprise model should have been developed prior to the start of the data warehouse project. If the model has not been finished and the project requires its completion, the time scheduled must be adjusted.	

Strategic Data Warehouse Checklist (continued)

Criteria	*Guidelines*	*Response*
v. How many attributes (i.e. fields and columns) will be selected for the project?	The more attributes to research, understand, clean, integrate and document, the more resources that will be needed.	
vi. Are the source files well modeled and documented?	Documentation is critical to the success of the project. Extra time and effort must be provided for if the source files and databases have not been well documented.	
vii. Will there be any external data in the project, and if so, is it well documented?	External data is often not well documented and usually does not comply with organisation standards. Integrating data is often difficult and time consuming.	
viii. Is the external data modeled (i.e. accurate, actively being used and comprehensive, etc.)?	Without a model, the effort to understand the source external data is significantly greater. It is often unlikely that the external data has been modeled.	
ix. How much cleaning will the source data require?	Data cleaning both with and without software tools to aid the process is tedious and time consuming. Organisations usually overestimate the quality of their data and may underestimate the effort to clean the data.	

Strategic Data Warehouse Checklist (continued)

Criteria	*Guidelines*	*Response*
x. How much integration will be required?	The need to integrate various source systems from various data stores can require significant resources (especially when complex external data elements must be introduced into the data warehouse environment).	
xi. What is the estimated size of the data warehouse database?	Databases in excess of 500 gigabytes may generate performance problems. These concerns should be kept in mind when considering costs and service level requirements.	
xii. What are the service level requirements?	In instances where the data warehouse will be required five days a week, eight hours a day, etc. significant running and maintenance costs will be incurred.	
xiii. How frequently will data elements need to be loaded and updated?	In instances where uploads are required more frequently, an increased effect on performance will be noted.	
xiv. Will a new hardware platform, user PC's or network infrastructure be required? If so, will it be different than the existing hardware?	The installation of new hardware will require planning, operations training and familiarisation.	

Strategic Data Warehouse Checklist (continued)

Criteria	*Guidelines*	*Response*
xv. How many query tools will be chosen?	The acquisition of new query tools also involves significant training and needed support time.	
xvi. Is user management sold on and committed to the data warehouse project and what is the organisational level at which the commitment was made?	If management are not sold on the project, the risk is significantly greater. Risks could include increased difficulty in getting resources and timely responses from affected parties.	
xvii. Who do the data warehouse project manager's report to?	The higher up the project manager's report, the greater the commitment and the more visibility the project will receive.	
xviii. Will the appropriate users be committed and available for the data warehouse project?	The more users which will be needed to provide input into the project, the more difficult coordination efforts will be. Also, if people important to the project are not committed and available, it is unlikely that the project will be completed on time.	
xix. Will knowledge application developers be available for the migration process and system testing?	Consideration should be given to ensuring that a sufficient number of application developers can be obtained to assist in the data warehouse development.	

Strategic Data Warehouse Checklist (continued)

Criteria	*Guidelines*	*Response*
xx. Will the database administrators be familiar with the chosen database management system and be available for the project duration?	As with all new software, a new database management system will require recruitment. It is also vital that the database administrator is available for the full duration of the data warehouse project.	
xxi. Will technical support people be available for capacity planning, performance monitoring and trouble shooting?	In instances where the organisation is considering an organisational-wide data warehouse, the need for support people will increase. Support in issues such as capacity planning and performance monitoring will become more critical as the data warehouse increases in size and complexity.	
xxii. How many queries are expected and what is the anticipated level of complexity of such queries?	The higher the volume of queries and the greater their complexity, the more user training that will be required.	
xxiii. Are there any significant security and/or audit issues which must be considered before the data warehouse is implemented?	Audit involvement in the data warehouse project is recommended and will require audit's commitment.	

Source: Adapted from Adelman, 1998: 1-4

Annexure 3

Roles and Responsibilities of the Data Administrator

The roles and responsibilities of the data administrator are sub-divided into four areas of specialisation. These areas of specialisation are detailed below as well as the associated objective of each. The roles and responsibilities are also included under each of the functional areas identified.

1. Data Administration Infrastructure

A. Objective of function

To provide the support framework for definitions, use and maintenance of data resources.

B. Roles and responsibilities

i. Establish and maintain the appropriate data administrator infrastructure for each customer type

This includes establishment of data administrator organisation, roles, responsibilities, external liaison interfaces and data administrator/repository administration internal procedures. A mission statement or charter needs to be written to align the four functions and new services with customer data needs. Job descriptions must reflect the role(s) staff members have in the new four functions and the skill sets required for such roles.

ii. Coordinate centralised/decentralised data administrator functions

Revised data administrator functions may remain centralised at the enterprise level, yet decentralised at the application project level to meet customer needs. Application data administrators perform data modeling roles and limited project repository administration roles to leverage data administrator activities within the development project. Decentralised application data administrator roles are to be coordinated by the centralised data administrator function.

iii. *Establish and maintain data administrator tools and tool roles*

Evaluation and recommendation of tools are conducted with other appropriate organisations. Once the tool has been determined, the appropriate infrastructure surrounding that tool must be established (i.e., repository/tools architecture, training, tool roles, tool security, etc.). The appropriate data/metadata policies, standards, procedures for each customer type needs to be enhanced and/or developed.

iv. *Plan strategically for future data/information needs*

This involves coordinating with other enterprise planning efforts to develop and publish a strategic long-term (usually five years ahead) data/information plan. This long-term plan provides target strategies, policies, standards, models and software tools. A migration plan of steps to take for reaching the target goals should be included. This strategic information plan feeds the data administrator organisational short-term plan.

v. *Establish and maintain a data administrator organisational short-term plan*

The initial data administrator short-term plan (two years or less) coordinates customer types, upcoming projects and data administrator roles.

vi. *Communicate data administrator topics*

Communicate data administrator topics to each customer type by giving presentations, training, publishing on the Web, etc. Data and metadata concepts as well as the appropriate policies, standards and procedures need to be disseminated to customer groups.

vii. *Establish and maintain standard practices*

Establish and maintain standard practices with regard to data analysis approach, data element identification, deliverables and tools and techniques for all customer types. Standard practices include the policies, standards and procedures that are enforced to minimise redundancy. It also includes data standardisation and data quality techniques that analyse mapping of data elements and constructs across databases.

2. Data Model Administration

A. Objective of function

To support creation and maintenance of the data architecture.

B. Roles and responsibilities

i. Create, publish and maintain the data architecture

This includes subject areas identification as well as creation and maintenance of the enterprise or corporate data model that contains high-level entities and their relationships derived from business rules.

ii. Project support

Provide project support for each customer type through the development and/or the review of data models.

3. Repository Administration

A. Objective of function

To develop and maintain the repository environment that each customer type should be encouraged to use. Data administrators should work toward the goal of having the repository become the centralised location of metadata for both the development and production environments as well as for end-user data access.

B. Roles and responsibilities

i. Establish and maintain the repository model architecture and repository objects (meta model)

The repository/tools architecture is a diagrammatic view of all tools and applications that fall within the scope of the initial repository interface requirements. It provides an essential road map of the way the repository environment should operate in a multitiered environment. Also, every

repository comes with a meta model with related entity types, policies, methods, templates, security, migration, etc. The data administrator should establish a meta model customised to its customer environment for use in the repository.

ii. *Enforce naming standards, keywords, abbreviations*

Manual enforcement of established repository standards is difficult and time-consuming. What is preferable is automated tool enforcement through proper implementation of edits that represent the standards. This step of validating names, keywords, abbreviations, etc., must be built into the systems development life cycle as part of the data administrator review activities.

iii. *Establish and maintain ongoing repository administration for each data administrator customer type*

This involves communicating and advising users on what exists in the repository and how to access it. Repository controls and security profiles must be established to protect, access and change repository components. Templates and standard reports need to be developed for customers to view repository contents and obtain reports.

4. End-User Data Access Administration

A. *Objective of function*

To create and maintain the data access categories and support the data warehouse and other data access projects.

B. *Roles and responsibilities*

i. *Create and maintain subject areas and relationships*

Create and maintain subject areas and relationships to entities for each customer by producing data for end-user access. Data warehouse or data mart projects are designed for end-user data retrieval and analysis. The enterprise may provide external data files for public use. Each data access project needs a

meta access model for the selected data access browser. This model contains the identified major and minor categories of entities for ease of use and data retrieval.

ii. *Provide data warehouse support*

Provide data warehouse support in the areas of modeling, source data quality, business rules, end-user tools, etc. Models need to be created, reviewed and maintained for each data warehouse. Data administrator support can include identification of source data to feed the warehouse as well as creation and maintenance of source-to-target data mappings, transitional data source-to-target rules and data access rules. The data administrator should establish the appropriate customer data warehouse/data access metadata directories that interface and integrate with the established repository environment.

Source: Adapted from Cupoli, 1999: 1-5

Annexure 4

Vendor Prescreening and Application Selection Questionnaire

Criteria	*Guidelines*	*Response*
A. Vendor Information Business Profile		
• Where does the vendor see their company in the future? • What is the vendor's guiding philosophy? • Which sector of the industry does the vendor fall under (examples: data warehousing, middleware, etc.)? • Which industry need is addressed by the product (examples: extraction process, testing and verification of data quality, etc.)?	Vendor information will help the organisation predict the vendor's long-term competitive strength in the market. The vendor's vision, or lack thereof, is a reflection of their view of the future and will provide insight into the viability of a long-term relationship.	
Indicate the number of years that the vendor has been in business by providing the aforementioned product.		
How many clients does the vendor have where the evaluated product is being evaluated and/or is already installed?	The completed questionnaire should be returned with full details of 5 (or more) clients and their references with contact numbers. Reference sites should refer to the evaluated product.	

Vendor Prescreening and Application Selection Questionnaire
(continued)

Criteria	*Guidelines*	*Response*
B. Financial Profile		
• What has been the annual increase in turnover over the past 5 years? • Is the vendor organisation funded by personal capital or a publicly traded company? • Provide a recent copy of the vendor's financial statements (if allowable).		
C. Product Information Version Information		
• What is the current version number of the product? • When last was a new release of the product being evaluated, issued? • How often does the company come out with new releases of the product being evaluated?		
D. Environment Supported		
• Specify the hardware platforms on which the product will function as per the specification. • Specify the software products and their versions that the product has been tested for.		

Vendor Prescreening and Application Selection Questionnaire
(continued)

Criteria	*Guidelines*	*Response*
• Specify the network protocol and any other information regarding the network compatibility of the product (Examples: TCP/IP, Remote Procedure Call, etc.).		
For data warehouses: • Is the evaluated product metadata aware? • Can the evaluated product read and use external metadata? • Does the product provide an interface to the metadata interchange? • Does the product use metadata internally?	Metadata is retained within the data base management system and includes information about the structure, content, keys, and indexes of the migrated data.	
D. Dependency on Other Vendor Products		
List the products that are critical for functioning of the product or on which the evaluated product must be used.	Ensure that the vendor's product keeps up with the changing versions of software it may be dependent on. This is done to ensure that future application upgrades in a multiple-product dependent environment may not be adversely affected.	

Vendor Prescreening and Application Selection Questionnaire
(continued)

Criteria	*Guidelines*	*Response*
E. Quality Information		
• Describe the testing environment existing in the vendor's company for development and support of the product.	Vendors that subscribe to software process models such as CMM, ISO or Cleanroom provide an indication of their commitment to quality.	
• Briefly list the average defect rate over the life of the product.		
F. Product Pricing Information		
• Obtain a copy of the vendor's terms of conditions. • Identify the pricing per copy of software purchased. • Detail any licensing fees in addition to the previous costs. • Detail the annual support and maintenance fee charged. • Obtain details of the product's warranty period and terms. • Obtain details of any other miscellaneous charges (examples: call-out costs, transportation and delivery, etc.).		

Vendor Prescreening and Application Selection Questionnaire
(continued)

Criteria	*Guidelines*	*Response*
G. Technical Support and Service Commitment		
• Identify the different technical support schemes on offer by the vendor. • Define what services will be included in the service commitment agreements offered by the vendor.	The service commitment agreement should bind the vendor to provide a guaranteed level of service within a specified period of time. Failure to comply with sanctioned terms should result in some form of penalty being incurred.	
H. Consulting/Mentoring Training and Consulting		
Provide full details of the training and consulting services provide by the vendor and what the associated costs are.		
I. Year 2000 Compliancy		
• Verify that the product is Year 2000 compliant. • Request third party verifications and authentic test results verifying Year 2000 compliance.		
J. Scalability		
• Can the product allow additional clients and servers to be added and removed in the future to handle increases and decreases in load?		

Vendor Prescreening and Application Selection Questionnaire
(continued)

Criteria	*Guidelines*	*Response*
• Provide independent results on the product performance when increasing and decreasing data loads.		

Source: Adapted from Tiwary S., Tewary A., 1998: 1-5