# Risk Management of Financial Information Systems Using Bayesian Networks

**Ann Muthoni Kibe**

**A thesis submitted in fulfilment for the degree of doctor of philosophy in Information Technology in Jomo Kenyatta University of Agriculture And Technology.**

**2015**

## DECLARATION

This thesis is my original work and has not been presented for a degree in any other university.

Signature: …………………………… Date: …………………………

**Ann Muthoni Kibe**

This thesis has been submitted for examination with our approval as the University supervisors:

Signature: …………………………… Date: …………………………

**Prof. Ronald Waweru Mwangi**

**JKUAT, Kenya**

Signature: …………………………… Date: …………………………

**Dr. Stephen Kimani**

**JKUAT, Kenya**

# DEDICATION

I dedicate this research project to my two children Joan Wambui and Cax Njagah, my parents, brothers and sister for their continuous support and encouragement.

# ACKNOWLEDGEMENT

I would like to thank the Almighty God for availing me an opportunity and strength to pursue my education. It is through his abundant grace that has brought this research project to a success. My utmost sincere gratitude also goes to my supervisors, Prof. Ronald Waweru Mwangi and Dr. Stephen Kimani, for their continuous support during my research study, for their patience, motivation, and immense knowledge.

I am also thankful to my colleagues at JKUAT especially the members of the postgraduate committee, whose encouragement, guidance and support from the initial level to the final level enabled me to develop an understanding of the research study.

My special thanks also go to my research assistant Lenny Habby, for the immeasurable effort, understanding and commitment throughout the research period. I also appreciate anyone else who provided input and positive criticism hence the accomplishment of this research project.

## TABLE OF CONTENTS

# LIST OF TABLES

# LIST OF FIGURES

# LIST OF APPENDICES

## ABBREVIATIONS AND ACRONYMS

**BN** – Bayesian Networks

**BBN –** Bayesian Belief Networks

**DBMS** - Database Management System

**FIS** - Financial Information System

**FISRM** - Financial Information System Risk Management

**ICT -** Information and Communication Technology

**IS** - Information System

**ISO** - International Standards Organization

**IT -** Information Technology

**KPI** - Key Performance Indicator

**RM** - Risk Management

**RAD** - Rapid Application Development

**ROI -** Return on Investment

**SPSS -** Statistical Package for social sciences

**SDLC** - System Development Life Cycle

# ABSTRACT

Managing financial information system threats is achieved through a process of risk management that enables an organization to have relatively safe and stable operating conditions. There however, exists a gap in the risk management programs of financial information system in regards to the use of all available information sources including from professionals, prior knowledge, and the use of multiple disparate frameworks. This research seeks to bridge that gap by managing financial information systems risk throughout the entire systems development life cycle, and incorporating Bayesian Networks in Financial Information System Risk Management. To achieve this the following specific objectives were formulated: (1) To assess the current state of financial information system risk management by financial institutions, (2) to analyse financial information system risks and management programs used by financial institutions (3) to analyse the application of Bayesian networks in risk management in financial information systems and (4) to propose and evaluate a generic approach that can be deployed to manage risks using Bayesian networks throughout the system development life cycle. The researcher presents detailed analysis in the domain of system development, risk management and their relationships. Purposive sampling is used to select 40 respondents from the population of major approved financial institutions in Nairobi County. The findings show that the current financial information risk management programs are not mature and are facing some challenges posed by the dynamic technological environment. Here the limitations and desired improvements to the current approach are brought out. The proposed framework which presents immense potential to address the current challenges faced by financial information system risk management programs is discussed in detail. Its success is however subject to elaborate planning and careful implementation Summary and conclusions are used to recoup the main areas of the research after which the researcher presents the recommendations.

# CHAPTER ONE

## 1.0 INTRODUCTION

### 1.1 Background of the study

Operating and maintaining technology is inherently risky and since financial institutions rely heavily on technology to support complex business processes and handle large volumes of critical information. As such, information risk management is crucial. With technology increasingly tied to critical business processes, technology failure can have crippling impact on an organization. Historically, technology risk management functions have struggled to keep pace with this trajectory. This has left some financial institutions exposed to potentially catastrophic technology risks that could impact their brand and reputation (PwC, 2013).

The reason why information technology has its significant implications in financial sector, is because, the financial sector is an information intensive industry, and information production technology is an important source of the competitiveness. These organizations depend on IT and the IS to successfully carry out their missions and business functions. The risk that comes with the wide application of information technologies in business, grows together with the increase of enterprise’s correlation from its customers, business partners and outsourced operations. Information and information systems are subject to serious threats that can have adverse impacts on organizational operations, mission, functions, reputation, organizational assets, individuals, and other organizations (NIST, 2010; Kim et al., 2012).

Siayor (2010) noted that the advancement of Information Technology (IT) has brought about rapid changes to the way businesses and operations are being conducted in the financial industry. IT is no longer a support function within a financial institution but a key enabler for business strategies including reaching out to and meeting customer needs. Financial systems and networks supporting financial institution’s business operations have also grown in scope and complexity over the years. MAS (2013), observed that institutions offering a diversity of products and services could have their financial systems operating in multiple locations and supported by different service providers. They are also faced with the challenge of keeping pace with the needs and preferences of consumers who are getting more IT-savvy and switching to internet and mobile devices for financial services, given their speed, convenience and ease of use. In response are deploying more advanced technology and online systems, including internet banking systems, mobile banking and payment systems, online trading platforms and insurance portals, to reach their customers. In this regard, financial institutions should fully understand the magnitude and intensification of technology risks from these systems. There is also a dire need for adequate and robust risk management systems as well as operating processes to manage these risks (Siayor, 2010; MAS, 2013).

In such a competitive and dynamic technological environment, the time needed for appropriate reaction on risk is decidedly shortened. The lack of appropriate preparation may lead the company to collapse, thus appropriate reaction on risk constitutes about possibilities of survival and development of enterprise (Pieplow, 2012).

### 1.2 Financial Information Systems

Technology is everywhere and information is power. Information Systems use technologies that allow information to be transmitted, stored, and accessed nearly instantaneously, in what has become an age of modern communications and global interconnectedness. The strategic use of IS within enterprise financial environments, not only provides a competitive edge, but is quickly being regarded as a resource of necessity rather than of convenience, in surviving what is now considered, a globalized marketplace (Hughes, 2012).

The main obligation of FIS is to keep updated and under control not only the revenue, the expenditure, the production and the human resources/payroll cycle of a firm, but also the general ledger and the reporting system of a firm. Hence, financial information systems offer: (1) operational assistance to a firm (keeping track of transactions), (2) knowledgeable support (using computerised tools for quick and easy support in investments), (3) managerial aid (controlling financial resources) and (4) strategic development of the organisation (establishing long-term investments goals and providing long-range forecasts of the firm’s financial performance). The above features of FIS are achieved and integrated with the widely use of Distributed Data Processing (DDP) configuration, as well as the use of Enterprise Resource Planning (ERP) system applications. ERP systems integrate all the operational aspects of a firm with the traditional accounting-financial functions. The corporate data are kept in databases and a Data Base Management (DBMS) is responsible for the data exploitation and sharing (Hughes, 2012; Siayor, 2010).

### 1.3 Financial Information Systems Risk Management

Schou and Hernandez (2014) indicate that Financial Information System (FIS) risk means the possibility of negative effects on the financial result and capital achievement of business objectives, operation in accordance with regulations, and reputation of a financial institution. This is due to inadequate information system management or other system weaknesses which negatively affect the system functionality or security, and/or jeopardise the business continuity of the financial institution. PwC (2013) indicated that, regulators have required financial institutions to implement technology risk management programs for many years, boards and audit committees are also demanding greater visibility into the technology risks facing their institutions, and how those risks are being addressed. For too long, financial institutions have viewed technology risk management as a defensive tactic or regulatory compliance activity. Technology risk management often consists of various siloed and fragmented processes working alongside one another to provide a compliance capability that supports technology audits and regulatory examinations. Given today’s environment of rapid change and intensifying regulatory scrutiny, these fragmented approaches to technology risk management cannot be sustained (PwC, 2013; Schou and Hernandez, 2014).

According to Kouns and Minoli (2011), there is no perfect protection against malicious attacks on data and information on ISs due to the fact that even the most advanced security systems are targeted by more and more complicated threats. A threat may take the form of any condition, circumstance, incident or person with the potential to cause harm by exploiting vulnerability in a system. The source of the threat can be natural, human or environmental. Humans are significant sources of threats through deliberate acts or omissions which could inflict extensive harm to the organisation and its information systems. Despite massive investments in IS risk management technologies, information is never truly secure where it resides. For most organizations, the value of the information associated with an Information Systen (IS) exceeds the value of the security technology associated with the IS, although, the resources and costs in maintaining compliance, while keeping up with the never ending flow of regulatory requirements, can be overwhelming; It is important for an organization to take into account what risk might mean to their strategies, and the outcomes those strategies will produce, while calibrating an appropriate balance between risk and reward that corresponds with the organization's appetite for risk and mission success (Kouns and Minoli 2011; Hughes, 2012).

Fenz, (2011) noted that minimizing negative impact on an organization and need for sound basis in decision making are the fundamental reasons organizations implement a risk management process for their IS. Effective IS risk management must be totally integrated into the System Development Life Cycle (SDLC). An Information System (IS) system's SDLC has five phases: initiation, development or acquisition, implementation, operation or maintenance, and disposal. In some cases, it may occupy several of these phases at the same time. However, the risk management methodology is the same regardless of the SDLC phase for which the assessment is being conducted. Risk management is an iterative process that can be performed during each major phase of the SDLC (Fenz, 2011).

Cortez (2011) indicated that in order to protect themselves, organizations have to take all due precautions, their approach must be defensive and proactive. Managing information system threats is achieved through a process of risk management that enables an organization to have relatively safe and stable operating conditions. The process of risk management depends completely on each individual organization. That is why there are as many ways of managing risks as there are organizations. There are four different approaches possible (i) Informal or technical approach without systematic/structural methods;(ii) General approach, which refers to a choice of standardized protection mechanisms for every part of the information system; (iii) Exact analysis, which consists of identification and evaluation of information material, threats and their level of danger; (iv) Combined approach, which provides an exact analysis for the most exposed parts/systems, and a basic analysis for less vulnerable parts (Cortez, 2011; Schou and Hernandez, 2014).

The fourth approach is the most useful according to Bessis (2010); in theory, identification and evaluation of information material are necessary parts of the process of risk management. What follows is uncovering and evaluating threats with the help of past experiences, and identifying the weaknesses of information material, which could eventually be misused. Furthermore, calculating of the possibility of such an attack, and its consequences, are also necessary. An exact system analysis is therefore inevitable in order to find the most appropriate mechanisms of protection for the information system. Such an analysis provides information for the management (Bessis, 2010; Cortez, 2011).

The management makes decisions based on provided information and research. The main goal is to achieve a balance among risks and costs, which arise due to the implementation of preventive and protective measures. After analysing the system, the organization chooses the most appropriate way for managing risks (Bessis, 2010).

Information systems risk management framework are established to manage risks in a systematic and consistent manner. The framework should encompass the following attributes: roles and responsibilities in managing technology risks; identification and prioritisation of information system assets; identification and assessment of impact and likelihood of current and emerging threats, risks and vulnerabilities; implementation of appropriate practices and controls to mitigate risks; and periodic update and monitoring of risk assessment to include changes in systems, environmental or operating conditions that would affect risk analysis (MAS, 2013).

Siayor (2010), indicates that risk management is a corner stone of good corporate governance and therefore results in better service delivery, more efficient and effective use of scarce resources and better project management. It has to do with identification, analysis and control of such risks that threaten resources, assets, personnel and the earning capacity of a company. Cortez, (2011) echoes this sentiments and goes ahead to state that risk management is the logical development and implementation of a plan to deal with potential losses. It is important for an organization to put in place risk management programmes so as to manage its exposure to risks as well as protect its assets. The essence is to prepare ahead of time on how to control and finance losses before they occur Siayor (2010), continues to say that risk management is a strategy of pre-loss planning for pre-loss resources. Risk management the processes by which organizations methodologically address the risks to their activities with the goal of achieving sustained benefit within each activity and across the portfolio of all activities.

The risk management process is a continuous activity involving these basic steps: (i) understanding the mission of the organization, (ii) risk assessment to identify the risks associated with the mission, (iii) categorizing and prioritizing the risks, (iv) design processes, (v) training and checks (controls) for top level risks, (vi) monitoring internal control effectiveness and (vii) making improvements as required and iteration of some of the steps (Siayor, 2010; MAS, 2013; Cortez, 2011).

Daly et al., (2011) indicate that BNs have been used to analyse risky situations particularly, Bayesian Networks (BNs) represent a formalism used in the risk analyses domain due to their capacity to deal with probabilistic data and to model the dependencies between events. BNs are an ideal decision support tool for a wide range of problems and have been applied successfully in a number of different settings where evidence is incomplete, contradictory or disparate. Unlike many other risk analysis methods, they make use of a range of data types, concepts and assumptions for which a range of evidence of varying quality exists (Daly et al., 2011). According to Feng and Xie, (2011) a distinctive feature of the Bayesian approach is that it permits the investigator to use both sample (data) and prior (expert-judgment) information in a logically consistent manner in making inferences by using Bayes' theorem to produce a 'post data' or posterior distribution for the model parameters (Feng and Xie, 2011; Daly et al., 2011).

## 1.4 Problem Statement

The financial industry is changing subject to the dynamic and ever changing technological environment, to gain competitive edge, product development and in order to keep in pace with the needs and preferences of consumers. As a result, they are with not only more but greater risks (both internal and external). As such there is a dire need to put in place adequate and robust risk management program as well as operating processes to manage these risks (MAS, 2013).

Technological progress generates dependencies which evoke growth of diversities, complexity, and non-descriptiveness of quantity of risk factors. This exposes financial information systems to serious threats that can have adverse impacts on organizational operations, mission, functions, reputation, organizational assets, individuals, and other organizations. Increased investments and consequent return on investment on FISRM becomes more significant, concentrating on searching optimal proportion between threats, costs of information systems protections and return on investment. This brings out the need for sustainable, dynamic and adaptable risk management program in the financial sector.

Risk management as a scientific methodology including those applied to large organizations as FISMA and RISK-IT have been criticized as being shallow and disregarding incomplete, disparate data from different sources. The risk management methodology is based on scientific foundations of statistical decision making: indeed, by avoiding the complexity that accompanies the formal probabilistic model of risks and uncertainty, the current Financial Information System Risk Management (FISRM) is a process that attempts to guess rather than formally predict the future on the basis of statistical evidence. It is highly subjective in assessing the value of assets, the likelihood of threats occurrence and the significance of the impact. In addition to this, researchers have only addressed the aspect of risk management in regard to already developed systems.

## 1.5 Justification

The existing risk management programs focuses on post implementation risks and specific phases of the software life cycle, without recognizing that risks in one stage can have an impact on other stages. This necessitates the proposed study in order to propose a generic approach that may be deployed to mitigate risks from the early stages of financial information systems development for daily financial institution operations until the post-implementation phases.

In a modern technological environment where information systems are characterized by complexity, situations of non-effective operation should be anticipated. Often system failures are a result of insufficient planning or equipment malfunction, indicating that it is essential to develop techniques for predicting and addressing a system failure. Particularly for safety–critical applications such as financial information systems, risk analysis should be considered a necessity.

A significant drawback of current models is that they are generally designed for use as a posteriori, and have difficulty handling risks with no prior data, or where data is incomplete or disparate. Additionally, with every method using its own terminology, it is difficult to combine several methods, in the aim of taking advantage of each of them. This research seeks to overcome this draw back and propose a better framework.

The proposed system unlike the existing system provides a method of integrating all the risk factors, their proposed mitigation techniques and other related factors (such as cost) and the relationship between them and a way of modelling and predicting the possible outcome without actual implementation. It also proposes a the use of more than one technique in the approach; that is BN, statistics, Artificial Intelligence (AI) and the inclusion of both all available information (complete, incomplete and disparate) from experts and passed experiences. Further it has the ability to learn, if modelled correctly and over time this approach as compared to those currently in use that are based on scientific foundations of statistical decision making is way more exhaustive with prediction capabilities hence economical. This is valuable in the risk management industry. All these coupled with FISRM from the initial stages of SDLC gives risk managers an upper hand in managing risk early before it compounds and thereby saving on resources (monetary or otherwise).

## 1.6 Research Objective

### 1.6.1 General Objective

The use of Bayesian Networks in risk management of financial information systems.

### 1.6.2 Specific Objectives

The specific objectives of this study are:

i. To assess the current state of financial information system risk management by financial institutions.

ii. To analyse financial information system risks and management programs used by financial institutions

iii. To analyse the application of Bayesian networks in risk management in financial information systems
iv. To propose and evaluate a generic approach that can be deployed to manage risks using Bayesian networks throughout the system development life cycle

### 1.7 Research Questions

The research questions of this study are:

i. What is the current state of financial information system risk management by financial institutions?
ii. What is the analysis of financial information system risks and management programs used by financial institutions
iii. How can Bayesian Networks be used in risk management in financial information systems?
iv. What generic approach can be deployed to manage risks using Bayesian networks throughout the system development life cycle?

### 1.8 Research Assumptions

Having a generic approach supporting the Information Systems Risk management process improves the product coming from the various Information Systems Risk management phases resulting to an overall efficient and effective risk management program. This generic approach however acts a general guideline and provides for modification to suit specific financial institution’s needs.

### 1.9 Significance of the Study

The study will be of significance to the following ways to the various stakeholders:

Financial institution Information System Managers: they, under the standard of due care and ultimate responsibility for mission accomplishment, must ensure that the necessary resources are effectively applied to develop the capabilities needed to accomplish the mission. They will benefit from this research that proposes new approach to IS risk management through prediction to improve the efficiency and effectiveness of risk management. They (managers) can then devise appropriate strategies, resource allocation and make amends in their current approach in the light of new information and also by use of incomplete and disparate data.

External Stakeholders: These include investors, customers and regulatory bodies with vested interest in the financial institutions who show an increasingly technological focus. This will also be the firms whose strategies are aligned towards using technologies that have been proven to deliver superior performance and reduced risks. Additionally, the level of risks faced will provide an indicator of the management quality as determined by its ability to design effective solutions to apply to IS challenges; also they will learn to appreciate service related difficulties faced by the providers which will create mutual satisfaction. To some extent they will play a role in risk management based on their relation with FIS: this study will be enable them analyze the various risk management factors and their relationships (direct or otherwise) and also interrogate the financial institutions of interest risk management approaches.

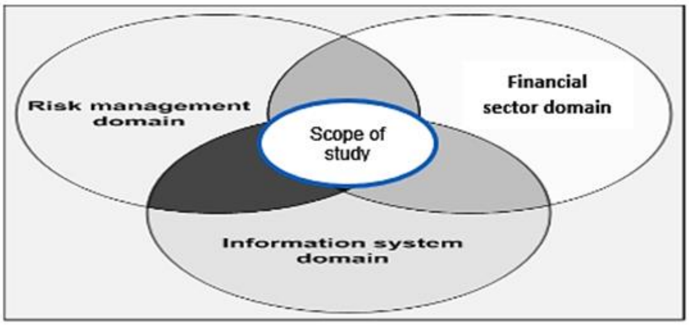
Consultants, IT security practitioners: These are professionals and specialists are responsible for proper implementation of IS requirements and assessing the effectiveness of IS solutions in the financial institutions and giving guidance to management on how to optimize these systems; support or use the risk management process, identify and assess new potential risks and implement new security controls as needed to safeguard their IT systems. The research will provide this group with valuable information not only on risks and risk management in the FIS development phases but also new risk management approaches and assessment of this approach. Knowledge of actual risks, relationship between variables, mitigation approaches and the ability to model and predict the various risk management measures and their consequences will better place consultants in a position to better advice and devise financial effective solutions to the risks.

Researchers and academics: This research through its findings, proposal, conclusion and recommendations will provide significant contribution to the body of knowledge in regards to risk management and especially FIS risk management. By filling the research gap that exists especially in Kenya in regards to FISRM and recommending areas of further research this study forms a basis for further research on newer, modern and better IS risks management approaches in the world of business and academia.

## 1.10 Scope of the Study

This research work is based on the Information System Risk management Domain as clearly shown in figure 1.1. It defines the different concepts and the boundaries of the work. The objective of financial information risk management (FISRM) is thus to safe guard essential constituents of a financial information system (FIS), from all threats which could arise accidentally or deliberately, by implementing risk management (RM) strategies.

*Figure 1.1: Scope related to the FISRM domain*



There are many definitions of information system most of them based on the domain applied, as such for this research we will adopt a technical definition of IS in a computerized environment. Therefore, an information system can be defined technically as a set of interrelated components that collect (or retrieve), process, store, and distribute information to support decision making and control in an organization. In addition to supporting decision making, coordination, and control, information systems may also help managers and workers analyze problems, visualize complex subjects, forecasting, operational activities and create new products. Three activities in an information system produce the information that organizations need;   input, processing, and output. In terms of the components that undertake this activity, they can be classified into five basic resources of people, hardware, software, communications and data. When such a system as defined is applied in financial institution it is referred to as a financial information risk management system (Hardcastle & Ventus., 2008; MAS, 2013).

One well accepted description of risk management is risk management is a systematic approach to setting the best course of action under uncertainty by identifying, assessing, understanding, acting on and communicating risk issues (Berg, 2010).

A standard definition of risk management is presented by AS/NZS 4360:1999 as the term applied to a logical and systematic method of establishing the context, identifying, analysing, evaluating, treating, monitoring and communicating risks associated with any activity, function or process in a way that will enable organizations to minimize losses and maximize opportunities. Risk management is as much about identifying opportunities as avoiding or mitigating losses (AS/NZS 4360:1999; Standards Association of Australia and Standards New Zealand, 2013).

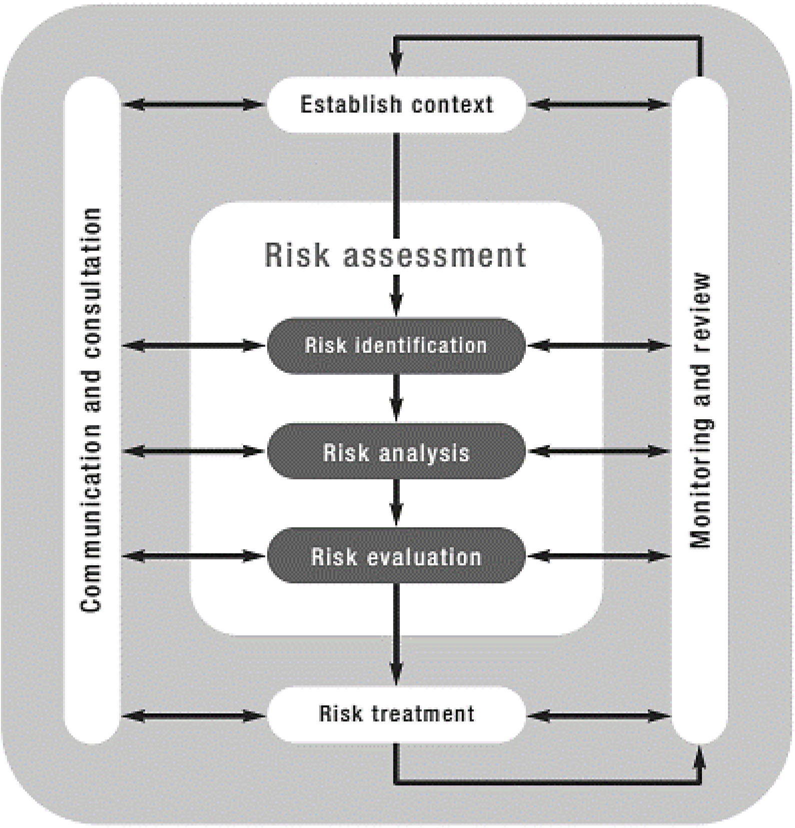
## 1.11 Theoretical Framework

ISO 31000:2009, Risk management - Principles and guidelines, provides principles, framework and a process for managing risk (figure1.2). This risk standard provides the theoretical basis for this research because if the various benefits it offers. It can be used by any organization regardless of its size, activity or sector. Also using ISO 31000 can help organizations increase the likelihood of achieving objectives, improve the identification of opportunities and threats and effectively allocate and use resources for risk treatment. It also provides guidance for internal or external audit programmes. Organizations using it can compare their risk management practices with an internationally recognised benchmark, providing sound principles for effective management and corporate governance (AS/NZS ISO, 2009; ISO, 2012; Rollason and Haines; 2012; Standards Association of Australia and Standards New Zealand, 2013).

Being a proactive approach, rather than a compliance approach, it provides a consistent approach that can be tailored to any type of operation in any location and integrated with other standards and guidelines. The International Standard Risk Management Principles and Guidelines (ISO 31000:2009) has been tried and tested across numerous industries; it is thus a reliable methodology for applying a risk based approach. Additionally, Organizations with existing risk management processes can use this Standard to critically review, align and improve their existing practices. Those whose risk management framework has been based on other standards (including traditional ones) will thereby benefit from the additional concepts and practices in this Standard (ISO, 2012; Rollason and Haines; 2012).

This standard is however not intended to promote uniformity of risk management across organizations. The design and implementation of risk management plans and frameworks will need to take into account the varying needs of a specific organization, its particular objectives, context, structure, operations, and processes.

*Figure 1.2: Conceptual model of the risk management process*



Source: airmic, et al., (2010).

Establishing the context entails the definition of the (internal and external) context of financial institutions, definition of the criteria against which risk will be evaluated and the definition of the structure for the rest of the process. This includes identifying your organisation’s objectives (vision, mission, and goals), the environment you operate in and the factors which influence your effectiveness. These factors can include financial, political, operational, cultural, public perception, and legal. This provides not only information about the organisation but its capability and the strategies it relies on to fulfil its objectives. It is used to set the scope, define the objectives of the risk management process and develop risk criteria tailored to the organisation's needs (Deloitte and Touche, 2012).

ISO (2012) presents risk assessment to be made up of three steps; risk identification, analysis, and evaluation. Risk identification identifies possible risks, their sources and potential impacts. The sources will assist development of preventive risk treatment and the impacts will assist development of reactive strategies. This is the determination of threats and vulnerabilities to the financial institution's IS environment (internal and external). Risk analysis involves the quantification of the potential impact and consequences of risks on the overall business and operations. It involves evaluation of existing controls, estimation of the magnitude of the consequences and the likelihood of the event of risks, classification risks whether minor or major. The institution should develop a threat and vulnerability matrix to assess the impact of the threat to its' information system environment. The matrix will also assist in prioritising IT risks. The organization should then take decisions about risks, based on the outcome of risk analysis. Risk evaluation considers questions such as, what is an acceptable or intolerable level of risk, which risks need treatment and what are the priorities? (Pandey and Mustafa, 2012; Deloitte and Touche, 2012; ISO, 2012).

Having identified and analysed the risk, it is necessary to identify the most appropriate treatment. For each type of risk identified, the institution should develop and implement risk mitigation and control strategies that are consistent with the value of the information system assets and the level of risk tolerance. This entails; identifying options for the treatment of risks (having positive or negative consequences), selecting the most appropriate option, the costs of risk controls should be balanced against the benefits to be derived. This can include controlling the risk, reducing the likelihood, reducing the consequences, transferring the risk, accepting the risk and avoiding the risk (airmic, et al, 2010; Standards Association of Australia and Standards New Zealand, 2013).

Monitoring and Review, communicate and consult:  This two parallel tasks should be Performed all along the Preceding Processes. The financial institution should maintain a risk register which facilitates the monitoring and reporting of risks. Risks of the highest severity should be accorded top priority and monitored closely with regular reporting on the actions that have been taken to mitigate them. The institutions should update the risk register periodically, and institute a monitoring and review process for continuous assessment and treatment of risks (MAS, 2013). To facilitate risk reporting to management, the institutions should develop IS risk metrics to highlight systems, processes or infrastructure that have the highest risk exposure. In determining the IS risk metrics, the institutions should consider risk events, regulatory requirements and audit observations. Monitoring and review - ensures that the risk management activities and processes are actually working effectively and any gaps are identified and addressed (MAS, 2013).

Risk parameters may shift as the IS environment and delivery channels change. Thus, the institutions should review and update the risk processes accordingly, and conduct a re-evaluation of past risk-control methods with renewed testing and assessment of the adequacy and effectiveness of risk management processes. Management of the IS function should review and update its risk control and mitigation approach, taking into account changing circumstances and variations in the institutions risk profile. Communication and consultation needs to involve stakeholders allowing the process to be both transparent and inclusive. One of the key areas in relation to communication is the development of a shared vocabulary to ensure that everyone is using the same terms in the same way in the same sector for instance financial sector (MAS, 2013; Pieplow, 2012).

## 1.12 Operational Definitions of Terms

**Risk:** the effect of uncertainty on objectives;

$$Risk = likelihood*impact \qquad \dots\dots\dots\dots (1.1)$$

**Risk management:** human activity which integrates recognition of risk, risk assessment, developing strategies to manage it and mitigation of risk using managerial resources. The strategies include transferring the risk to another party, avoiding the risk, reducing the negative effect of the risk and accepting some or all of the consequences of a particular risk. The objective of risk management is to reduce different risks related to a pre-selected domain to a level accepted by society.

**Information Systems:** These are capabilities offered to financial institution by computers, software applications and telecommunications to deliver data, information and knowledge to individuals and processes.

**Information system risk** means the possibility of negative effects on the financial result and capital, achievement of business objectives, operation in accordance with regulations, and reputation of a financial institution due to inadequate information system management or other system weaknesses which negatively affect the system functionality or security, and/or jeopardise the business continuity of the financial institution

**Risk management:** The process that allows IT managers to balance the operational and economic costs of protective measures and achieve gains in mission capability by protecting the Information systems and data that support their financial institution' missions.

**Risk ontology -** formal representation of knowledge as a set of concepts within a domain, and the relationships between those concepts. It includes rules about how concepts interact, and provide a basis for calculations and analytics. It provides a basis for establishing consensus on the meaning of risk terms, and a model that explains their use.

**Traceability**: The degree to which a relationship can be established between two or more products of the development process, especially products having a predecessor-successor or master-subordinate relationship to one another.

## 1.13 Research Organization

The research is organised into seven sections. The first chapter introduces the research topic by presenting a background, the problem statement, research question and the theoretical framework upon which the research is based. The second chapter presents risk management in the context of financial institutions and details the phases of risk management process based on international standards and best practices. Chapter three goes through the entire systems development life cycle and the risk encountered in each phase through the use of risk taxonomy. Bayesian networks are then presented in detail in chapter four with various examples of application in the context of information system risk management. The research methodology presented in chapter five followed by summary, conclusion and recommendations of the research.

### 1.14 Summary

IS risk management is a crucial issue in all financial institutions and needs to be investigated further in view of the fact that financial management is a risky business. Every transaction in a financial institution has financial consequences to the financial institution itself and its customers. Consequently, deployment of IS as a medium for transacting necessitates an equivalent development of a generic model to guard against the risk of loss. Such an approach should be incorporated right from the outset and should have risk maintenance capabilities. This study therefore seeks to address this and propose a generic approach for risk identification and management that can be deployed to mitigate risks from the early stages of financial information systems development for daily financial institution operations until the post-implementation phases with the use of Bayesian Network.

# CHAPTER TWO

## 2.0 FINANCIAL INFORMATION SYSTEM RISK MANAGEMENT

### 2.1 Introduction

Information system risk means the possibility of negative effects on the financial result and capital, achievement of business objectives, operation in accordance with regulations, and reputation of a financial institution due to inadequate information system management or other system weaknesses which negatively affect the system functionality or security, and/or jeopardise the business continuity of the financial institution

### 2.2 Overview of Risk Management

Risk is inherent in the business environment and therefore affects the design, implementation and use of information processing systems. Most definitions relate risk to uncertainty. However, Berg, (2010), present the definition of risk that has been adopted in this study as the uncertainty that surrounds future events and outcomes. Risk is defined by the ISO Guide 73 standard as the effect of uncertainty on objectives (International Organization for Standardization 2009a). Risk is therefore connected to a consequence and a likelihood. It is the expression of the likelihood and impact of an event with the potential to influence the achievement of an organization's objectives. This means that for each risk, two calculations are required: its likelihood or probability; and the extent of the impact or consequences.

Pandey and Mustafa (2012), noted that risk management is an integral part of every organization’s strategic management; it is processes by which organizations methodological address the inherent risks in its activities with the goal of achieving sustain benefits from each activity and across all portfolios of activities. It is viewed as a corner stone of good corporate governance and therefore results in better service delivery, more efficient and effective use of scarce resources and better project management (Pandey and Mustafa, 2012).

Since risk management is directed at uncertainty related to future events and outcomes, it is implied that all planning exercises encompass some form of risk management. There is also a clear implication that risk management is everyone's business, since people at all levels can provide some insight into the nature, likelihood and impacts of risk (Lang, 2011). Risk management is about making decisions that contribute to the achievement of an organization's objectives by applying it both at the individual activity level and in functional areas. It assists with decisions such as the reconciliation of science-based evidence and other factors; costs with benefits and expectations in investing limited public resources; and the governance and control structures needed to support due diligence, responsible risk-taking, innovation and accountability (MAS, 2013; Lang, 2011).

### 2.3 Risk Management Concepts

Gaidow and Boey, (2009) presented basic requirements and concepts for any successful enterprise-wise risk management program, that form the fundamental overall risk management concepts for these study. They are:

The process of risk management is an iterative one. With each repetition of the cycle there may be changes in the risk criteria and hence a progressive improvement of the risk management process may be achieved. This may result in increasing the benefits from the application of the risk management approach. In general, the outcomes from the implementation of risk management may include more effective decisions, more effective allocation and use of resources, higher standard of customer service, and more flexibility in meeting objectives (Madill, 2003; Pieplow, 2012).

Risk management has to be an integral part of any management practice. It can be applied at any organizational level from strategic through operational, functional and tactical to project. Starting with the strategic aspect guarantees that the lower level aspects are accurately placed within the strategic context. The most benefit is obtained by employing risk management right from the beginning of an activity. But it may be applied to help the decision-making in specific situations or specific risk areas. Risk management may exhibit its relevance in times of major changes for an organization as well as in everyday routine operations (Siayor, 2010; MAS, 2013).

Lang (2011) expressed that risk management has to be an integral part of quality management. It has to be incorporated in the existing organizational structure at all levels.

This will enable managers and staff to identify a wider range of options and strive for better outcomes. Risk management will facilitate greater responsibility and flexibility in the decision making process. A structured risk management framework may also stimulate continuous improvement and innovative thinking (Lang, 2011).

Risk management has to be an essential part of corporate governance, i.e. direction, executive action, supervision and accountability. The tools and techniques of risk management give any manager at any level a systematic approach to managing risks within their corporate responsibilities. Risk management may also provide some protection in the case of adverse results. Corporate governance activities may be improved by establishing links between risks, returns and resources, for example by applying risk management in the efficient allocation of resources. To establish a risk management framework an organisation needs a policy, a support mechanism and an implementation program (Madill, 2003; MAS, 2013).

## 2.4 Principles of risk management

The AS/NZS ISO 31000:2009, (2010) introduced 11 principles of risk management as a guide for organizations to benchmark their risk management programs. A risk management program:

i. Creates and protects value; good risk management contributes to the achievement of an agency's objectives through the continuous review of its processes and systems.

ii. Is an integral part of organisational processes; risk management needs to be integrated with an agency's governance framework and become a part of its planning processes, at both the operational and strategic level.

iii. Is part of decision making; the process of risk management assists decision makers to make informed choices, identify priorities and select the most appropriate action.

iv. Explicitly address uncertainty; by identifying potential risks, agencies can implement controls and treatments to maximise the chance of gain while minimising the chance of loss.

v. Is systematic, structured and timely; the process of risk management should be consistent across an agency to ensure efficiency, consistency and the reliability of results.

vi. Based on the best available information; to effectively manage risk it is important to understand and consider all available information relevant to an activity and to be aware that there may be limitations on that information. It is then important to understand how all this information informs the risk management process.

vii. Is tailored; an agency's risk management framework needs to include its risk profile, as well as take into consideration its internal and external operating environment.

viii. Take into account human and cultural factors; risk management needs to recognise the contribution that people and culture have on achieving an agency's objectives.

ix. Is transparent and inclusive; engaging stakeholders, both internal and external, throughout the risk management process recognises that communication and consultation is key to identifying, analysing and monitoring risk.

x. Is dynamic, iterative and responsive to change; the process of managing risk needs to be flexible. The challenging environment we operate in requires agencies to consider the context for managing risk as well as continuing to identify new risks that emerge, and make allowances for those risks that no longer exist.

xi. Facilitate the continual improvement of organisations; agencies with a mature risk management culture are those that have invested resources over time and are able to demonstrate the continual achievement of their objectives.

### 2.5 Overview of risk management Standards and Practices

Over the years there has been a major interest in improving our ability to deal with uncertainty, and especially with its negative impact at the organisation level. This has led to the development and application of tools, techniques, processes and methodologies which are typically classified under the label of "risk management". These standards provide guidance and advice, and encourage the adopting organisations to adapt them to their own needs. They include and are not limited to: CRAMM, AS/NZS 4360:2004; Risk management; Project Risk Analysis & Management (PRAM) Guide, 2nd edition; ISO 3100; CAN/CSA-Q850-97

(Gaidow and Boey, 2009; Madill, 2003; Standards Association of Australia and Standards New Zealand, 2013).

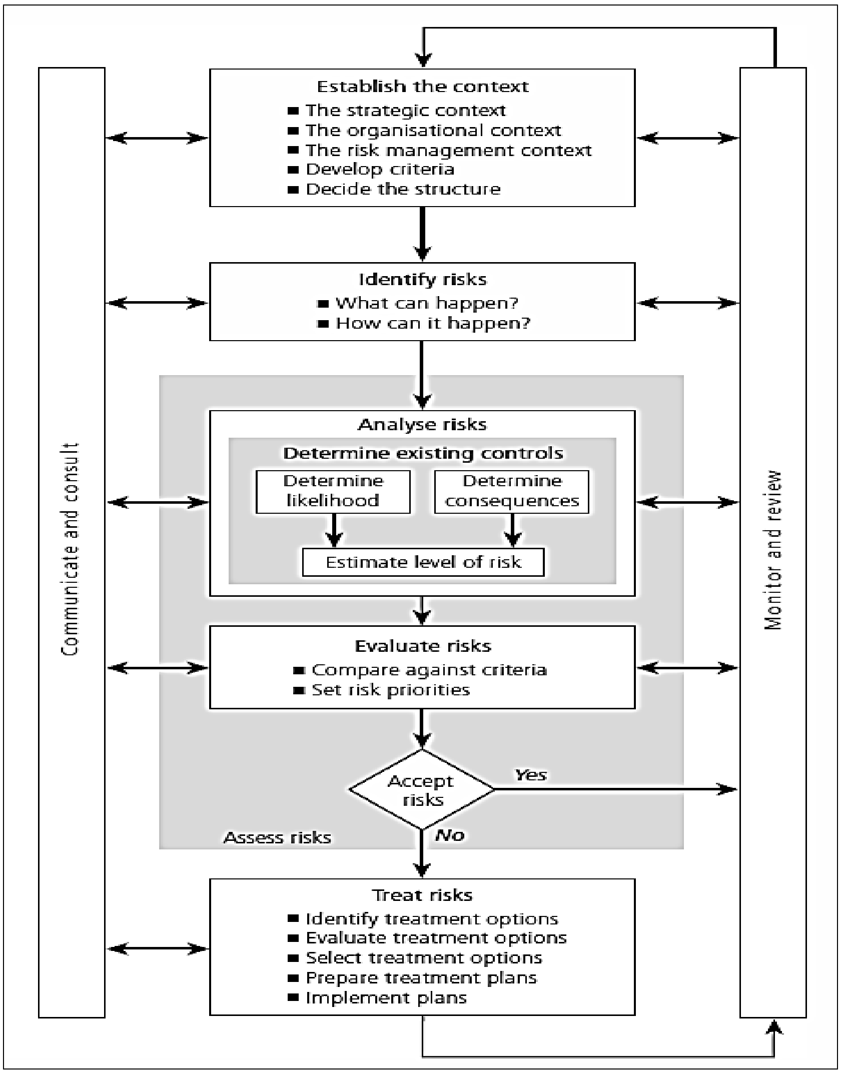
A review of the process steps described by the selected standards identified the following main steps: planning, identification, analysis, treatment and control. Terminology differs between the standards, though the structure of the process in each case is similar. For example, in some standards analysis is called assessment; and in some cases analysis is broken down into estimation (of probability and consequences of the risk events) and evaluation (determining the overall magnitude of the risk event, from which its priority is derived). There is a wide consensus regarding the main steps and activities of a generic risk management process, there is still room for a comprehensive document which will not only combine the best elements of the existing standards, but also provide broad coverage of the issues related to instituting such a process (Madill, 2003; Pieplow, 2012). This research therefore seeks not only to introduce an Artificial Intelligence (AI) based Bayesian Network (BN) approach but also that financial information system risk management (FISRM) be done from the initial phases of information system (IS) development.

## 2.6 Information Systems Risk Management Process

The process of risk management depends completely on each individual organization, which is why there are as many ways of managing risks as there are organizations. It is used by leaders, managers and professional personnel to recognize the weaknesses in their information system, in order to ensure trust, integrity and accessibility of all components of the system. Even though the information system’s operation is explained thoroughly in theory, its type depends on various factors, such as the size of the organization, the interests of the management, how qualified the personnel is, and whether the organization is financially capable to establishing and maintaining such a system (MAS, 2013).

When looking into the risk management process, there are several different methodologies found from scientific research as well as standards and guidelines reports such as the ISO 27000 series and NIST, AS/NZS publications. Most of them share similarities while slightly differing from each other. Today there exist literally hundreds of ISRM methods and standards targeted to professionals. ISRM activities usually follow an overall process composed of classical steps generally found in traditional ISRM methods, this research adopts the AS/NZS 4360 which has been the basis of modern risk management standards and guidelines. Standards provide generic guidance throughout all steps of the risk management process and could be tailor made to suit most if not all organizations. As such the presentation and exploration of this risk management framework in this section has been adapted and presented with financial information system risk management in mind (Pieplow, 2012; MAS, 2013; Standards Association of Australia and Standards New Zealand and Standards New Zealand, 2013).

*Figure 2.1: Risk Management Process*



Source: Standards Association of Australia and Standards New Zealand, (2013)

The first phase is establishing the context which occurs within the framework of an organization’s strategic, organizational and risk management context. This needs to be established to define the basic parameters within which risks must be managed and to provide guidance for decisions within more detailed risk management studies. This sets the scope for the rest of the risk management process (Madill, 2003).

In this phase the financial institution defines the relationship between the organization and its environment, identifying the organization’s strengths, weaknesses, opportunities and threats

(SWOT analysis). The context includes the financial, operational, competitive, political (public perceptions/image), social, client, cultural and legal aspects of the organization's functions. It identifies the internal and external stakeholders, and consider their objectives, take into account their perceptions, and establish communication policies with these parties.

This step is focused on the environment in which the organization operates. The organization should seek to determine the crucial elements which might support or impair its ability to manage the risks it faces. Strategic analysis may be undertaken. It should be endorsed at the executive level, set the basic parameters and provide guidance for the more detailed risk management processes. There should be a close relationship between an organization's mission or strategic objectives and its management of all the risks to which it is exposed (Madill, 2003).

Deloitte and Touche, (2012) explain the four crucial phases for establishing organization context: Firstly, before a risk management study is commenced, it is necessary to understand the organization and its capabilities, as well as its goals and objectives and the strategies that are in place to achieve them, also known as the organization context. Secondly, is to establish the risk management context where the goals, objectives, strategies, scope and parameters of the activity, or part of the organization to which the risk management process is being applied, should be established. The process should be undertaken with full consideration of the need to balance costs, benefits and opportunities. The resources required and the records to be kept should also be specified. Thirdly is the development of risk evaluation criteria where the organization decides the criteria against which risk is to be evaluated. Decisions concerning risk acceptability and risk treatment may be based on operational, technical, financial, legal, social, humanitarian or other criteria. These often depend on an organization's internal policy, goals, objectives and the interests of stakeholders. Criteria may be affected by internal and external perceptions and legal requirements. It is important that appropriate criteria be determined at the outset. Although risk criteria are initially developed as part of establishing the risk management context, they may be further developed and refined subsequently as particular risks are identified and risk analysis techniques are chosen, i.e. the risk criteria must correspond to the type of risks and the way in which risk levels are expressed. The last activity involves defining the structure. This involves separating the activity or project into a set of elements. These elements provide a logical framework for identification and analysis which helps ensure significant risks are not overlooked. The structure chosen depends on the nature of the risks and the scope of the project or activity (Deloitte and Touche, 2012; Standards Association of Australia and Standards New Zealand, 2013).

This phase involves the development of criteria against which risk is to be evaluated. The criteria usually depend on the interests of the stakeholders and the objectives of the organisation. Here the acceptable level for each risk has to be considered (Standards Association of Australia and Standards New Zealand, 2013).

Risk identification: this step seeks to identify the risks to be managed. At this step of the risk management process one has to apply a well-structured and systematic approach and try to identify all risks, which may potentially arise. Comprehensive identification using a well-structured systematic process is critical, because a potential risk not identified at this stage is excluded from further analysis. Identification should include all risks whether or not they are under the control of the organization. The aim is to generate a comprehensive list of events which might affect each element of the structure, these are then considered in more detail to identify what can happen. Having identified a list of events, it is necessary to consider possible causes and scenarios. There are many ways an event can be initiated. It is important that no significant causes are omitted. Approaches used to identify risks include checklists, judgments based on experience and records, flow charts, brainstorming, systems analysis, scenario analysis and systems engineering techniques. The approach used will depend on the nature of the activities under review and the types of risk. Risks beyond the organizations control are to be identified as well (Deloitte and Touche, 2012; Pieplow, 2012).

Risk analysis is the next phase; here one considers the risk consequences (impact or magnitude of effect) and likelihood (measured by frequency or probability) of risk occurrence to combine them into the level of risk (Madill, 2003). The objectives of analysis are to separate the minor acceptable risks from the major risks, and to provide data to assist in the evaluation and treatment of risks. Factors which affect consequences and likelihood may be identified. Risk is analysed by combining estimates of consequences and likelihood in the context of existing control measures. A preliminary analysis can be carried out so that similar or low-impact risks are excluded from detailed study. Excluded risks shall, where possible, be listed to demonstrate the completeness of the risk analysis. The risk level is discussed within the context of existing or non-existing controls (Madill, 2003).

Once the analysis is done by examining existing controls Identifying the existing management, technical systems and procedures to control risk and assess their strengths and weaknesses. Appropriate tools as well as approaches such as inspections and control self-assessment techniques (‘CSA’) should be used. The magnitude of consequences of an event, should it occur, and the likelihood of the event and its associated consequences, are assessed in the context of the existing controls. Consequences and likelihood are combined to produce a level of risk. Consequences and likelihood may be determined using statistical analysis and calculations. Alternatively where no past data are available, subjective estimates may be made which reflect an individual’s or group’s degree of belief that a particular event or outcome will occur. To avoid subjective biases the best available information sources and techniques should be used when analysing consequences and likelihood. Wherever possible, the confidence placed on estimates of levels of risk should be included (Deloitte and Touche, 2012).

Risk analysis may be undertaken to various degrees of refinement depending upon the risk information and data available. Analysis may be qualitative, quantitative or a combination of these, depending on the circumstances. In practice, qualitative analysis is often used first to obtain a general indication of the level of risk. Later it may be necessary to undertake more specific quantitative analysis (Madill, 2003). Sometimes it is appropriate to consider likelihood to be composed of two elements, usually referred to as frequency of exposure and probability. Frequency of exposure is the extent to which a source of risk exists, and probability is the chance that when that source of risk exists, consequences will follow. Caution must be exercised in situations where the relationship between the two elements is not completely independent, i.e. where there is a strong relationship between frequency of exposure and probability. Since some of the estimates made in quantitative analysis are imprecise, a sensitivity analysis should be carried out to test the effect of changes in assumptions and data (Madill, 2003; Deloitte and Touche, 2012).

Risk analysis is to avoid bias and hence is to be based on the best available sources of information and data management techniques. Examples of information sources includes a lists past records, published literature, market research, relevant individual and industry practice and experience, various models, and expert judgements. The techniques may include interviews, questionnaires, expert group discussions, computer modelling, statistical analysis, and decision-making tools. There are three types of methods applicable in risk analysis (in order of complexity): qualitative, semi-quantitative, and quantitative. Usually one starts with the qualitative analysis to get a rough approximation of the level of risk and then proceeds with a more accurate quantitative analysis (Pandey and Mustafa, 2012).

Risk evaluation involves comparing the level of risk found during the analysis process with previously established risk criteria against which risks are compared and considered on the same basis. Thus qualitative evaluation involves comparison of a qualitative level of risk against qualitative criteria, and quantitative evaluation involves comparison of numerical level of risk against criteria which may be expressed as a specific number, such as fatality, frequency or monetary value. The output of a risk evaluation is a prioritized list of risks for further action (Madill, 2003).

The objectives of the organization and the extent of opportunity which could result from taking the risk should be considered. Decisions shall take into account the wider context of the risk and include consideration of the tolerability of the risks borne by parties other than the organization which benefits from it. If the resulting risks fall into the low or acceptable risk categories they may be accepted with minimal further treatment. Low and accepted risks should be monitored and periodically reviewed to ensure they remain acceptable (Standards Association of Australia and Standards New Zealand, (2013).

According to Deloitte and Touche (2012), risk evaluation results in a ranked list of risks. Then, these risks are classified as acceptable or unacceptable. Acceptable risks are to be monitored and their acceptable status reviewed periodically. Unacceptable risks have to be prioritised by management for treatment. The risk evaluation has to consider the big picture including the stakeholder’s objectives and risk tolerability, the degree of control over each risk, the cost, the benefits and potential opportunities. Management is also to respond by allocating responsibilities in the risk treatment process with respect to the level of risk.

Risk treatment/mitigation involves identifying the range of options for treating risk, assessing those options, preparing risk treatment plans and implementing them Gaidow and Boey, (2009) bring out the various existing risk mitigation options which are not necessarily mutually exclusive or appropriate in all circumstances, which include; risk avoidance, risk transfer, risk control, and risk retaining.

Risk avoidance, an organization avoids the risk by deciding not to proceed with the activity likely to generate risk (where this is practicable). Inappropriate risk avoidance may increase the significance of other risks. In relation to IS systems development, avoidance includes trading off risk for performance or other capability, and it is a key activity during requirements analysis. Avoidance requires understanding of priorities in requirements and constraints. Are they mission critical, mission enhancing, or nice to have? (Chornous and Ursulenko, 2013).

Risk transfer in full or in part which involves another party bearing or sharing some part of the risk. Mechanisms include the use of contracts, insurance arrangements and organizational structures such as partnership and joint ventures (Madill, 2003). The transfer of a risk to other parties, or physical transfer to other places, will reduce the risk for the original organization, but may not diminish the overall level of risk to society. Where risks are transferred in whole or in part, the organization transferring the risk has acquired a new risk, in that the organization to which the risk has been transferred, may not manage the risk effectively. After risks have been reduced or transferred, there may be residual risks which are retained. Plans should be put in place to manage the consequences of these risks if they should occur, including identifying a means of financing the risk. Risks can also be retained by default, which is when there is a failure to identify and/or appropriately transfer or otherwise treat risks (Chornous and Ursulenko, 2013).

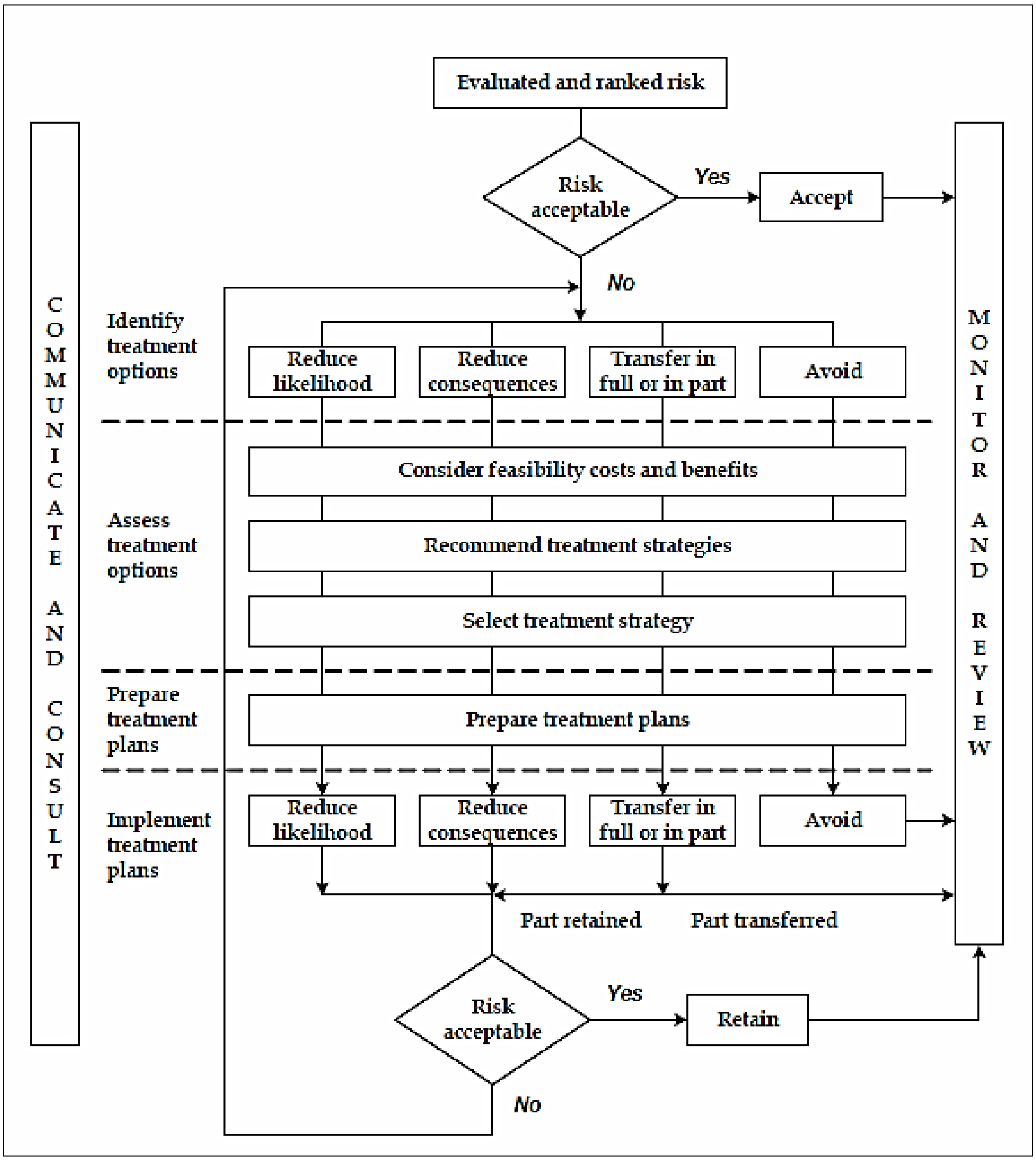
Risk control which is the reduction of consequence and likelihood of occurrence may be referred to as risk control. A priori formulated risk criteria have to form the basis of this optimisation procedure while the specific circumstances and the established risk context will determine the most suitable criterion. Together, they will yield the solution to the risk reduction problem. Risk control involves determining the relative benefit of new controls in the light of the effectiveness of existing controls. Controls may involve effectiveness policies, procedures or physical changes. Options should be assessed on the basis of the extent of risk reduction, and the extent of any additional benefits or opportunities created. A number of options may be considered and applied either individually or in combination. Selection of the most appropriate option involves balancing the cost of implementing each option against the benefits derived from it. In general, the cost of managing risks needs to be commensurate with the benefits obtained (Berg, 2010). In general the adverse impact of risks should be made as low as reasonably practicable, irrespective of any absolute criteria. If the level of risk is high, but considerable opportunities could result from taking the risk, such as the use of a new technology, then acceptance of the risk needs to be based on an assessment of the costs of risk treatment, and the costs of rectifying the potential consequences versus the opportunities afforded by taking the risk.

Risk Retaining, usually residual risk after the completion of risk reduction or risk transfer procedures. Risk may be retained by default when it is not treated by the organisation.

The successful implementation of the risk treatment plan requires an effective management system which specifies the methods chosen, assigns responsibilities and individual accountabilities for actions, and monitors them against specified criteria. If after treatment there is a residual risk, a decision shall be taken as to whether to retain this risk or repeat the risk treatment process (Deloitte and Touche, 2012).

Risk mitigation strategies and methodologies should be consistent throughout the organization to provide confidence that risk mitigation efforts expended in one place are not being undermined by weaknesses allowed elsewhere (Gaidow and Boey, 2009). This also help to ensure consistent practice across the organization and enhance interoperability and versatility. The Figure 2.2 depicts the risk treatment process

*Figure 2.2: Risk Treatment/Mitigation Process*



Adopted from Gaidow and Boey, (2009).

Assessing treatment options involves considering their feasibility, benefits and cost, to recommend treatment strategies, and to select a treatment strategy. Assessing risk treatment options is a process which has to be conducted with respect to the extent of risk level reduction, of the number of newly created opportunities, of the size of the additional benefits and with respect to the risk evaluation criteria including budget constraints. Usually a single risk treatment option cannot be the solution for a specific problem. A number of options have to be considered and applied together in combination. For example, reduction of risk likelihood, reduction of risk consequences, risk transfer, and risk retention, if applied simultaneously may provide a better solution (Borek, Parlikad, Webb and Woodall 2013). When assessing risk treatment options one has to base the decision making on the balance between an option's cost implementation and benefits obtained from it. As a rule the cost has to be lower or at most commensurate with the benefits. Exceptions of this rule cover risks of rare likelihood with catastrophic (severe) consequences. Such risks have to be treated despite the potential or even real danger of being identified as unjustified in a financial context. Further on, if a high-level risk undertaking could be associated with a considerable number of new opportunities emerging from it, then the assessment would have to include risk treatment cost and the risk consequences rectification cost. These two costs have to be weighed against the impact of the aforementioned opportunities. Prepared treatment plans show how the selected treatment strategies have to be implemented. They have to clearly delegate responsibilities, provide time schedules, describe the expected treatment effects, secure adequate resourcing, determine performance measures, and establish a rigorous review process. Plans have to include performance criteria against which the implementation of the risk treatment options is to be tested. Treatment plans usually contain critical milestones needed in the implementation monitoring (Gaidow and Boey, 2009; Chornous Ursulenko, 2013; Borek, et al., 2013).

Implementing treatment plans requires the existence of a management system capable of identifying the techniques to be used, assigning the responsibilities and accountabilities to individual level, and monitoring the process against specified criteria (Gaidow and Boey, 2009).

Monitor and review is not just a step, but an ongoing process embedded in the risk management process. It deals with the performance of the risk management system and the potential changes affecting it. It is necessary to monitor risks, the effectiveness of the risk treatment plan, strategies and the management system which is set up to control implementation. Risks and the effectiveness of control measures need to be monitored to ensure changing circumstances do not alter risk priorities. Berg, (2010). Few risks remain static. Ongoing review is essential to ensure that the management plan remains relevant. Factors which may affect the likelihood and consequences of an outcome may change, as may the factors which affect the suitability or cost of the various treatment options. It is therefore necessary to regularly repeat the risk management cycle. Review is an integral part of the risk management treatment plan (Berg, 2010).

The organization's objectives, its internal structures and systems, and the environment in which it operates, are continually evolving. As a result, the risks the organization faces are continually changing. A sound system of information risk mitigation will include the regular re-evaluation of the nature and extent of the risks to which the organization is exposed, plus periodic adjustment to ensure the organization continues to steer the line between allowing risks to grow out of hand and constraining operational effectiveness. The assumptions made in the previous risk assessment (hazards, likelihood and consequence), the effectiveness of controls and the associated management system as well as people need to be monitored on an on-going basis to ensure risk are in fact controlled to the underlying criteria. For an efficient risk control the analysis of risk interactions is necessary (Borek, et al., 2013).

Risk monitoring is the continuous process of tracking and evaluating the risk management process by metric reporting, enterprise feedback on watch list items, and regular enterprise input on potential developing risks. (The metrics, watch lists, and feedback system are developed and maintained as an assessment activity.) The output of this process is then distributed throughout the enterprise, so that all those involved with the program are aware of the risks that affect their efforts and the system development as a whole (MAS, 2013).

According to Madill, (2003), monitor and review is not just a step, but an ongoing process embedded in the risk management process. It deals with the performance of the risk management system and the potential changes affecting it. Risks change with time and circumstances. Hence the need to monitor them and their environments, the implementation of risk treatment plans, the system set up to control the risks and the established contexts and risk priorities.

No matter how diligently an organization strives to ensure it has all appropriate controls in place, protection failures will arise from time to time. Organizations need to monitor for protection failures so they can deal with incidents as they arise and contain the harm those incidents cause. Organizations also need to keep the number and nature of their incidents under review so they can learn the available lessons. Incidents provide a rare objective indicator of the real level of risk being experienced, and should be used to benchmark and adjust the risk mitigation controls in place (Gaidow and Boey, 2009; Borek, et al., 2013).).

Risks of the highest severity should be accorded top priority and monitored closely with regular reporting on the actions that have been taken to mitigate them (MAS, 2013).

Review is a continuous process and an integral part of the risk management plan. It makes sure the plan stays relevant and up-to-date. It introduces all changes in the risk management process. The inevitable regular repetition of the risk management cycle is based on the review process.

It is important to understand that the concept of risk is dynamic and needs periodic and formal review. New risks and their impact on the organization have to be taken into account. This step requires the description of how the outcomes of the treatment will be measured. Milestones or benchmarks for success and warning signs for failure need to be identified. The review period is determined by the operating environment (including legislation) (Gaidow and Boey, 2009).

In view of changes in IT environment and delivery channels, risk parameters may change. Thus, the risk processes should be reviewed and enhanced accordingly. Re-evaluation of past risk-control methods with renewed testing and assessment of the adequacy and effectiveness of risk management processes should be conducted. Management should review and update its risk control and mitigation approach, taking into account changing circumstances and variations in its risk profile (Gaidow and Boey, 2009; Borek, et al., 2013).

Communication and consultation are an important consideration at each step of the risk management process. Clear communication is essential for the risk management process, i.e. clear communication of the objectives, the risk management process and its elements, as well as the findings and required actions as a result of the output. It is ongoing and lasts as long as the whole risk management process. It is important to develop a communication plan for both internal and external stakeholders of the financial institutions at the earliest stage of the process. This plan should address issues relating to both the financial information risk itself and the process to manage it. Communication and consultation involve a two way dialogue between stakeholders with efforts focused on consultation rather than a one way flow of information from the decision maker to other stakeholders (Pieplow, 2012). Effective internal and external communication is important to ensure that those responsible for implementing risk management, and those with a vested interest understand the basis on which decisions are made and why particular actions are required (ACHS, 2013).

Consultation has to be given priority rather than simply passing information from decision makers to the other participants in the process. Perceptions of risk can vary due to difference in assumptions and concepts and the needs, issues and concerns of stakeholders as they relate to the risk or the issues under discussion. Stakeholders are likely to make judgments of the acceptability of a risk based on their perception of risk. Since stakeholders can have a significant impact on the decisions made, it is important that their perceptions of risk, as well as their perceptions of benefits, be identified and documented and the underlying reasons for them understood and addressed (Madill, 2003).

Risk management is an integral element of organization´s management. However, for its successful adoption it is important that in its initial stages, the reporting on risk management is visible through the framework. The requirements on the reporting have to be fixed in a qualified and documented procedure, e. g., in a management handbook (Berg, 2010).
To facilitate risk reporting to management, IS risk metrics should be developed to highlight systems, processes or infrastructures that have the highest risk exposure. An overall information system risk profile of the organization should also be provided to the board and senior management. In addition, risk events, regulatory requirements and audit observations should be considered in determining the information system risk metrics (MAS, 2013).

Each stage of the risk management process should be documented. It will contain results, plans, reviews, assumptions, methods, data, etc. AS/NZS 4360:2009 prescribes appropriate documentation as required for the proper management of risk. Documentation should include assumptions, methods, data sources and results. The reasons for documentation are as follows: to demonstrate the process is conducted properly, to provide evidence of a systematic approach to risk identification and analysis, to provide a record of risks and to develop the organisation’s knowledge database, to provide the relevant decision makers with a risk management plan for approval and subsequent implementation, to provide an accountability mechanism and tool, to facilitate continuing monitoring and review, to provide an audit trail, and to share and communicate information (Standards Association of Australia and Standards New Zealand, 2013).

Pieplow, (2012) explains that documentation is essential to demonstrate that the process has been systematic, the methods and scope identified, the process conducted correctly and that it is fully auditable. Documentation provides a rational basis for management consideration, approval and implementation including an appropriate management system. This document is a basis for communication throughout the organization and for the on-going monitor and review processes. It can also be used with other supporting documents to demonstrate regulatory compliance. Decisions concerning the extent of documentation may involve costs and benefits and should take into account the above factors (Pieplow, 2012).

### 2.7 Governance, Policies, and Controls

Developing and maintaining strong governance, policies, and controls over the risk management framework is fundamentally important to its effectiveness. Even if model development, implementation, use, and validation are satisfactory, a weak governance function will reduce the effectiveness of overall model risk management. A strong governance framework provides explicit support and structure to risk management functions through policies defining relevant risk management activities, procedures that implement those policies, allocation of resources, and mechanisms for evaluating whether policies and procedures are being carried out as specified. Notably, the extent and sophistication of a financial institution’s governance function is expected to align with the extent and sophistication of model usage (OCC, 2012; Borek, et al., 2013).

# CHAPTER THREE

## 3.0 INFORMATION SYSTEMS DEVELOPMENT PROCESS

### 3.1 Overview

Technology has developed at a rapid pace but the most important aspect of any system is human know-how and the use of ideas to harness the computer so that it performs the required tasks. This process is essentially what system development is all about. According to the Institute of Beynon-Davies (2013), in business, systems development refers to the process of examining a business situation with the intent of improving it through better procedures and methods.

This section describes the standard phases and major processes of Development Lifecycle (SDLC); additionally, it provides an overview of other SDLC methodologies (Beynon-Davies; 2013).

There are many SDLC methodologies that can be used by an organization to effectively develop an information system. A traditional SDLC, is a linear sequential model (also known as waterfall method) in which the system will be delivered in its final stages of the development life cycle. Another SDLC method uses the prototyping model, which is often used to develop an understanding of system requirements without actually developing a final operational system. More complex systems may require more iterative development models. More complex models have been developed and successfully used to address the evolving complexity of advanced and sometimes large information system designs (Wang and Wang, 2012). Examples of these more complex models are the rapid application development (RAD) model, the joint application development (JAD) model, the prototyping model, and the spiral model. The expected size and complexity of the system, development schedule, and length of a system’s life will affect the choice of which SDLC model to use, application development tools, by the software architecture within which the application will operate, or by the “build versus buy” decision, that vary from organization to organization. In many cases, the choice of the SDLC model will be defined by an organization’s acquisition policy (Broad, 2013; Davendranath, 2013; Wang and Wang, 2012).

This section incorporates risk management into the linear sequential model of SDLC because this model is the simplest of the various models, and it is an appropriate platform for this discussion. However, it is important to note that the concepts discussed can be adapted to any SDLC model.

### 3.2 Systems Development Life Cycle (SDLC)

The traditional, full, sequential work pattern life-cycle model is divided into six phases. Depending on the size and complexity of a project, alternative work patterns may be selected that will result in the combining, extending or overlapping of specific phases. Not every project will require that every phase be executed. The SDLC cycle describes a broad and diverse set of activities for addressing information systems efforts. It consists of a set of steps or phases in which each phase of the SDLC uses the results of the previous one (Broad, 2013).The content in this section is organized according to a generic system development lifecycle. While no two development efforts are exactly alike, all projects generally progress through these phases influenced by the project characteristics (Broad, 2013).

The SDLC is document driven which means that at crucial stages during the process, documentation is produced. A phase of the SDLC is not complete until the appropriate documentation or artifact is produced. These are sometimes referred to as deliverables.

A deliverable may be a substantial written document, a software artifact, a system test plan or even a physical object such as a new piece of technology that has been ordered and delivered. This feature of the SDLC is important to the successful management of an IS project.

The SDLC can also be viewed from a more process oriented perspective in which some activities are undertaken concurrently. The overlap of processes emphasizes the parallel nature of some of the activities and presents activities such as system maintenance as an alternative to a complete re-design of an existing system (Broad, 2013; Boyde, 2014).

The SDLC Process is a repeatable, predictable process intended to improve project productivity and system quality and to enhance the management of a computer systems project. A project team produces a set of deliverables (the SDLC deliverable set) which documents both the system design and the validation activities related to implementing or installing the system. By instituting and using the SDLC Process, the organization should implement quality systems (according to specifications and with minimal issues) and should incur minimal project waste (cost/schedule overruns, rework) (Billgruener, 2007; Davendranath, 2013).

### 3.2.1 System Initiation/Concept Phase

Boyde (2014) indicates that system concept development actually starts the life cycle when a need to develop or significantly change a system is identified. Once a business need, based on operational requirements, is identified and documented the approaches for meeting it must be reviewed for feasibility and appropriateness. The need may involve development of a new system or modification of an existing system.

In this phase, the Business Case and Proposed Solution developed during Project Origination are re-examined to ensure that they are still appropriately defined and address an existing organizational need. A validation effort is made to provide the project team with the basis fora detailed schedule defining the steps needed to obtain a thorough understanding of the business requirements and an initial view of staffing needs. In addition, a high level schedule is developed for subsequent system development lifecycle phases. Approvals and funding are needed before beginning the Planning phase (Davendranath, 2013).

### 3.2.2 System Planning Phase

The Planning phase begins after the project has been defined and resources have been committed to the project. A project plan is developed that documents the approach to be used and includes the discussion of methods, tools, tasks, resources, project schedules, and user input. It is the process of understanding why the system should be built and defining its requirements. It also includes feasibility study from several different perspectives, technical, economic, and organization feasibility aspects (Davendranath, 2013).

### 3.2.3 System Requirements Analysis Phase

This phase formally defines the detailed functional user requirements using high-level requirements identified in the Initiation and Feasibility Phases. They need to be measurable, testable, and relate to the business need or opportunity identified in the Initiation Phase. Functional user requirements are formally defined in a Functional Requirements Document

(FRD) and are delineated in terms of data, system performance, security, and maintainability requirements for the system. All requirements are defined to a level of detail sufficient for systems design to proceed (Boyde, 2014).

From a business perspective the needs of the business are captured in as much detail as possible. The Project Manager leads the Project Team in working with the Customers to define what it is that the new system must do. By obtaining a detailed and comprehensive understanding of the business requirements, the Project Team can develop the Functional Specification that will drive the system design (NYS, 2010).

The purposes of this phase are to: Complete business process reengineering of the functions to be supported, e.g., verify what information drives the business process, what information is generated, who generates it, where does the information go, and who processes it. Develop detailed data and process models including system inputs and outputs. Develop the test and evaluation requirements that will be used to determine acceptable system performance (Michigan Tech Information Technology Services & Security, 2010).

### 3.2.4 System Design Phase

This phase builds upon the work performed during System Requirements Analysis. During this phase, the system is designed to satisfy the functional requirements identified in the previous phase and results in a translation of the functional requirements into a complete technical solution. This solution dictates the technical architecture, standards, specifications and strategies to be followed throughout the building, testing, and implementation of the system. The completion of System Design also marks the point in the project at which the Project Manager should be able to plan, in detail, all future project phases (Marshall and Brainerd, 2010)

The external physical characteristics of the system are designed, the operating environment is established, major subsystems and their inputs and outputs are defined, and processes are allocated to resources. Everything requiring user input or approval must be documented and reviewed by the user. The internal physical characteristics of the system are specified and a detailed design is prepared. Subsystems defined during the external design are used to create a detailed structure of the system. Each subsystem is partitioned into one or more design units or modules. Detailed logic specifications are prepared for each module. This phase exactly determines how the system operates in terms of process, data, hardware, network infrastructures, user interface, and other important factors in the system environment. The Design phase ends with a formal design walk-through with the user and approval of the design by the system owner (Deloitte and Touche, 2012).

Since problems in the design phase can be very expensive to solve in later stages of the software development, a variety of elements are considered in the design to mitigate risk. These include: Identifying potential risks and defining mitigation techniques, design features, performing a security risk assessment, developing a conversion plan to migrate current data to the new system, determining the operating environment, defining major subsystems and their inputs and outputs, allocating processes to resources (Michigan Tech Information Technology Services and Security, 2010).

### 3.2.5 System Development/ Acquisition Phase

According to Lazaros and Prodromos (2011), coding and debugging is done in this phase, where, the design is implemented by the software engineer. The design described in the previous phase serves as the blueprint for the system to be built, providing most of the information the software engineer will need. The software engineer will interpret the design and develop the code. Even when the software engineer is also the designer, it is important to have a detailed design, because it is easy to overlook minor details that can result in a major error (Lazaros and Prodromos, 2011).

Debugging is the process of locating and removing errors from the code. Most current programming languages allow compiling a “debug” version of the code. The “debug” version allows stepping through code, setting breakpoints, viewing current variable values, and offers debug information about the code that helps the software engineer locate problems. After the code is stable, the production version of the code is compiled and used for system testing (Heidrich et al., 2013).

Effective completion of the previous stages is a key factor in the success of the Development phase. In this phase the Project Team builds and tests the various modules of the application, including any utilities that will be needed during System Acceptance and System Implementation. As system components are built, they will be tested both individually and in logically related and integrated groupings until such time as a full system test has been performed to validate functionality. Documentation and training materials are also developed during this phase (Broad, 2013).

### 3.2.6 System Integration and Testing

System integration occurs when distinct software modules are linked together and are capable of functioning as a unit. When there are multiple software engineers on a project, all the developers are expected to code to an accepted standard; if they do, and the design is good, there will likely be very few problems, if any, at this point. Unfortunately, this is not always the case. A common cause of system breakdown is a software engineer deciding that something needs to be done differently without informing the other software engineers. Because modules need to work together, a common protocol must be followed (Heidrich et al., 2013).

System testing helps to locate problems, and potential problems, with a software system. It is essential to have people other than the software engineers testing the software. During this phase, the focus of system validation efforts shifts from those team members responsible for developing the application to those who will ultimately use the system in the execution of their daily responsibilities. In addition to confirming that the system meets functional expectations, activities are aimed at validating all aspects of data conversion and system deployment. For larger software projects, reporting bugs and prioritizing bug fixes will be a coordinated effort between the project manager, software engineer, and testers Davendranath, 2013).

### 3.2.7 System Deployment/Implementation

This phase is initiated after the system has been tested and accepted by the user. In this phase, the system is installed to support the intended business functions. System performance is compared to performance objectives established during the planning phase. Implementation includes user notification, user training, installation of hardware, installation of software onto production computers, and integration of the system into daily work processes. This phase continues until the system is operating in production in accordance with the defined user requirements (Lazaros and Prodromos, 2011).

### 3.2.8 System Maintenance, Operations and Support

The system operation is ongoing. The system is monitored for continued performance in accordance with user requirements and needed system modifications are incorporated. Operations continue as long as the system responds to the organization's needs. When modifications are identified, the system may re-enter the planning phase (Boyde, 2014).

Boyde (2014) also notes that maintenance includes items such as patches and data updates, while support includes bug fixes, help for users of the software, and collecting requests for new functionality. Different types of maintenance and support may be provided based on what makes sense for the particular software product that is being created as well as on the needs of the customer.

### 3.2.9 System Disposal

The disposition activities ensure the orderly termination of the system and preserve the vital information about the system so that some or all the information may be reactivated in the future if necessary. Particular emphasis is given to proper preservation of the data processed by the system so that the data are effectively migrated to another system or archived in accordance with applicable records management regulations and policies for potential future access (McKay, 2006).

### 3.2.10 System Documentation

The life-cycle methodology specifies which documentation will be generated during each phase. Some of the products may be the basis for information collection requirements, Information Resources Management (IRM) reviews, and cost-benefit analyses (CBAs).

The outputs of SDLC documentation activities are typically categorized into two major types: process documentation, and product documentation.

Process Documentation -Process documentation communicates status and direction. It addresses the actions required for developing, implementing, and maintaining the system. Process documentation is not updated after implementation; however, it should be retained for evaluation and general reference. Examples include project plans, timelines, funds required, procedures to be followed, project review reports, requirements documents, and design documents.

Product Documentation - Product documentation describes the system itself, what it is, how it is operated, and how it is to be maintained. It is most often used by individuals who were not directly involved in the system's development and instructs them on how to effectively operate, maintain, and use the system. Modifications to the system are reflected in the documents as they occur and new versions are distributed periodically. Examples include user manuals, operations manuals, and maintenance manuals.

Some documentation remains unchanged throughout the systems life cycle while others evolve continuously during the life cycle. Other documents are revised to reflect the results of analyses performed in later phases. Each of the documents produced are collected and stored in a project file (McKay, 2006).

### 3.3 Other SDLC Methodologies

There are many SDLC methodologies, which can be used by an organization to effectively develop an information system. They include:

*Joint Application Development*: In a traditional waterfall methodology, the development team gathers requirements, many times through a series of interviews with the customer, and then proceeds to develop the application. Using a Joint Application Development (JAD) methodology, however, the client or end user collaborates with the developers through JAD sessions to design and develop an application. Because the development process involves greater involvement of the client, this methodology may lead to faster development and greater client satisfaction (Parsons and Oja, 2013).

Another model is the *prototype model* which is a development methodology similar to the waterfall model, in that once the requirements analysis is performed and the prototype is designed, the prototype development begins. Once created, the prototype is evaluated by the customer, who then provides feedback to the developer. The developer, in turn, refines the product according to the customer's expectation. After a number of iterations of this process, the final product is provided to the customer (Parsons and Oja, 2013).

*Rapid Application Development (RAD)* according to Broad (2013), is a development methodology that creates an application more quickly by employing techniques aimed at speeding application development, such as the use of fewer formal methodologies and reuse of software components. In exchange for faster development, some compromises in functionality and performance may be realized. It is important to ensure, however, that this exchange for a faster product delivery does not result in compromises being made in the selection and specification of the security controls necessary to provide adequate security for the information and the information system, and the mission function they support (Broad, 2013).

*The Spiral Model* is a development methodology that combines the features of the prototype and waterfall models, and is often favoured for large, expensive, and complicated projects. The spiral model process generally involves defining requirements and creating an initial design, and constructing and evaluating the first prototype. This same process is then repeated for subsequent prototypes until the refined prototype represents the product desired. The final system is constructed based on the final prototype, and is evaluated and maintained in a production environment. The major distinguishing feature of the spiral model is that it creates a risk-driven approach to the software process rather than a primarily document-driven or code-driven process. It incorporates many of the strengths of other models and resolves many of their difficulties (Parsons and Oja, 2013).

### 3.4 Risks Faced During the SDLC

Although only few publications provide an explicit definition for risk as a quantitative measure of probability and impact, risk is commonly understood as a function of probability and impact: Literature focuses on software development projects, ranging from methodological recommendations for developing code, environmental and socio-economic aspects to project management approaches. Every project of IS development has some specific kinds of risk, however, some categories of risks are usual for all the kinds of projects, no matter how complex their carrying out is. Deloitte and Touche (2012), noted that risks, in general, can be classified into three categories; known risks with known consequences, known risks with unknown consequences, unknown risks.

Known risks, with known consequences, are events which the project team knows, and whose probability of appearance is high. Known risks, with unknown consequences, are known to the project team, but their influence on the project is not known. Unknown risks represent risks which cannot be identified, therefore there is no way to predict their consequences and make a plan of activities if these events realize. It is necessary, even from the general aspect, to include this kind of risk in this plan (Deloitte and Touche, 2012; Parsons and Oja, 2013).

Risks can originate from internal and external sources. Internal risks depend on the very project nature, organizational questions, staff, resources, and so on. External risks are out of control of the project team and the organization including political and legal (Manikandan, Anbuoli and Saikishore).

### 3.5 Risk Identification Approaches

It is undeniable that the systematic use of risk management into the project development process will have a considerable negative impact on project risks level. In an attempt to promptly, effectively and easily identify risk, managers of software projects have been using various methods Pieplow, (2012) categorised risks into four groups;

The first one is the Ad-hoc Approach, which provides an assessment of risks when the initial symptoms appear on the project, as well as their mitigation with unofficial way. The second approach is called Informal Approach and includes a discussion with people, who are directly or indirectly involved with the project, concerning the several risk issues that appear (or will possibly appear) and the recording and documentation of the risks for future use. The third is the Periodic Approach and, as it can be understood from its title, involves the use of repetitive procedures for the identification and specification (quantitatively and qualitatively) of the risks. Finally, the fourth approach is the Formal Approach for the identification of the various risks. According to this approach, a thorough and in-depth assessment of each risk by independent individuals is performed (Pieplow, 2012; Parsons and Oja 2013; McManus, 2012).

## 3.6 Risk Taxonomy

Central to the risk identification method is the software development taxonomy. The taxonomy provides a framework for organizing and studying the breadth of software development issues. Hence, it serves as the basis for eliciting and organizing the full breadth of software development risks both technical and non-technical. They also provide a consistent framework for the development of other risk management methods and activities (Stern and Arias, 2011; McManus, 2012).

Risk taxonomies are lists of problems that have occurred on other projects and can be used as checklists to help ensure all potential risks have been considered. The Risk Taxonomy follows the SDLC and provides a framework for organizing data and information. The taxonomy-based identification method provides the organization developing software with a systematic interview process with which to identify sources of risk (Menezes et al., 2013; McManus, 2012).

## 3.7 The Software Development Risk Taxonomy

Because the development of large-scale financial information system codes is an often difficult, complicated, and sometimes uncertain process, success depends on identifying and managing risk. The Software Engineering Institute (SEI) taxonomy of software development maps the characteristics of software development and hence of software development risks. The Task-Based Questionnaire (TBQ) consists of a list of non-judgmental questions to elicit issues and concerns (that is, potential risks) and risks in each taxonomic group. The questionnaire ensures that all risk areas are systematically addressed, while the application process is designed to ensure that the questions are asked of the right people and in the right manner to produce optimum results. The method adopted here presents a disciplined and systematic way to identify risk in a software-dependent system development. This method allows risks to be identified without justification and without a proposed solution. This is the first step in establishing vital communication within an organization. The taxonomy also provides a consistent framework for the development of other risk management methods and activities (Stern and Arias, 2011; McManus, 2012).

The software taxonomy is organized into three major classes:

i. Product Engineering. The technical aspects of the work to be accomplished.
ii. Development Environment. The methods, procedures, and tools used to produce the product.
iii. Program Constraints. The contractual, organizational, and operational factors within which the software is developed but which are generally outside of the direct control of the local management.

These taxonomic classes are further divided into elements and each element is characterized by its attributes. The figure 3.1 contains a schematic of the taxonomy described from the software development risk perspective.

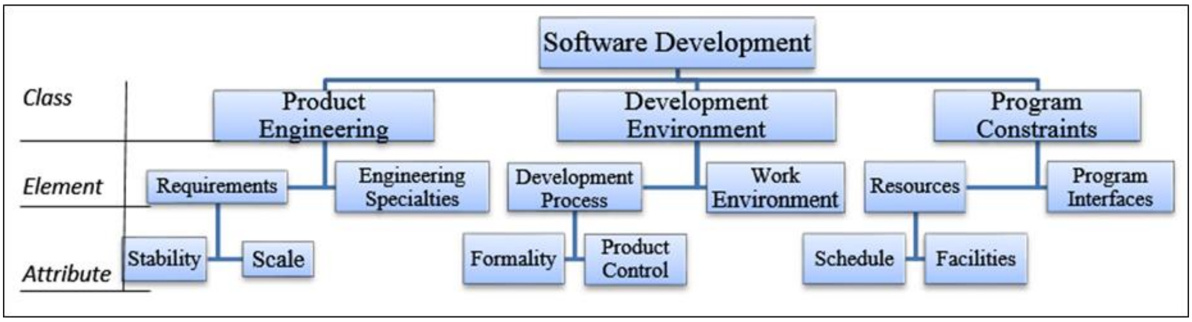


*Figure 3.1: Risk taxonomy software development risk* (Stern and Arias, 2011)

### 3.7.1 Product Engineering Class

The product engineering class consists of the intellectual and physical activities required to build the product to be delivered to the customer. It includes the complete system hardware, software, and documentation. The class focuses on the work to be performed, and includes the following elements:

*Requirements* is the definition of what the software product is to do, the needs it must meet, how it is to behave, and how it will be used. This element also addresses the feasibility of developing the product and the scale of the effort.

Design; entails the translation of requirements into an effective design within project and operational constraints

*Code and Unit Test* is the translation of software designs into code that satisfies the requirements allocated to individual units.

*Integration and Test*; is the integration of units into a working system and the validation that the software product performs as required.

*Engineering Specialties*; product requirements or development activities that may need specialized expertise such as safety, security, and reliability (Stern and Arias, 2011; McManus, 2012).

### 3.7.2 Development Environment Class

The development environment class is concerned with the project environment in which a software product is engineered. This environment consists of the following elements:

*Development Process* which is the definition, planning, documentation, suitability, enforcement, and communication of the methods and procedures used to develop the product.

*Development System*; where the tools and supporting equipment used in product development, such as computer-aided software engineering (CASE) tools, simulators, compilers, and host computer systems.

*Management Process*; is the planning, monitoring, and controlling of budgets and schedules; controlling factors involved in defining, implementing, and testing the product; the project manager's experience in software development, management, and the product domain; and the manager's expertise in dealing with external organizations including customers, senior management, matrix management, and other contractors (Menezes et al., 2013).

*Management Methods*; this is where the methods, tools, and supporting equipment that will be used to manage and control the product development, such as monitoring tools, personnel management, quality assurance, and configuration management.

*Work Environment*; is the general environment within which the work will be performed, including the attitudes of people and the levels of cooperation, communication, and morale (Jonasson, 2012; Carl, 2014).

### 3.7.3 Program Constraints Class

The program constraints class consists of the "externals" of the project, the factors that are outside the direct control of the project but can still have major effects on its success. Program constraints include the following elements: resources which are the external constraints imposed on schedule, staff, budget, or facilities; contract which are the terms and conditions of the project contract; and program Interfaces which are the external interfaces to customers, other contractors, corporate management, and vendors.

One of the drivers of the evolution of software engineering, as a discipline, has been the desire to identify reliable, quantifiable ways to manage software development risks (the possibility of suffering harm or loss, or "the product of uncertainty associated with project risks times some measure of the magnitude of the consequences", stemming from, for example; uncertain or inaccurate requirements, requirements that change too rapidly, overly optimistic scheduling, institutional turmoil, including too much employee turnover, poor team performance (Menezes et al., 2013; Carl, 2014).

## 3.8 Sources of System Development Risks

The taxonomy attempts to organize the sources of system development risk for financial information systems around three principal aspects of the software development activity: development cycle risks, development environment risks, programmatic environment risks

### 3.8.1 Development Cycle Risks

*a) System Requirements analysis risks*

Requirements analysis, definition and management are intrinsic elements of life-cycle management. Risk attributes of the requirements risk element are associated with both the quality of the software requirements specification and also the difficulty of implementing software that satisfies the requirements.

In projects that start from poorly articulated requirements, there is inherently far more risk that imprecisely expressed expectations will not be met. Technically difficult or imprecise requirements, coupled with the inability to negotiate relaxed requirements or budgets or schedules is a well-recognized source of software engineering risk (Jonasson, 2012; Carl 2014).

The following attributes will be employed to illuminate the nature of the risks that are associated with the “requirements” elements: (predictability, completeness, clarity, accuracy, precedence, execution performance expectations, proportionality, evolvability (The failure to recognize and adequately address the continuous evolution of requirements).

*b) System Design Risks*

Design encompasses those steps through which requirements are translated into an actionable development plan. We distinguish three steps: software architecture (abstract or conceptual design), specification, and design per se. The system architecture should be influenced by the scientific domain and the mathematical attributes of the application (e.g., initial value problem, steady-state problem, time evolution problem). With only the requirements and architecture, many system implementations are admitted; with a complete design and complete specification, there are far fewer options. It is important therefore to vet these documents (usually referred to as a “baseline”) both ways: with the sponsor/customer to ensure that requirements (even unstated) and expectations will be met, and with the implementation team to ensure that they are confident of successful implementation. Finally, it is important that specifications and design be documented and kept up-to-date (this is referred to as baseline management); otherwise the work breakdown structure, scheduling, and budgeting (which should be based upon them) will be faulty. This is typically assessed in a critical design review (CDR) (Marshall and Brainerd, 2010; Carl, 2014).

Another sometimes overlooked design risk is the impact of design on testing. Difficulty in testing may begin with failure to include test features, especially those important to users, in the design. The following attributes characterize different aspects of the risks inherent in the design element. It is also important to recognize that while documentation is very important, there is a risk of becoming too enamoured with this aspect of project management at the expense of continuous validation against changing needs. These include; difficulty, modularity, usability, maintainability, portability, reliability (Stern and Arias, 2011).

*c) Implementation Risks*

This element addresses the sources of project risk associated with how the coding will be done, that is, how the design will be translated into unit specifications and ultimately units of code. Attributes of this element describe the nature of risks associated with the quality and stability of system or interface specifications, and coding constraints or even styles that, if left unspecified, may exacerbate future maintenance and extensibility problems. System specifications are as important to the project success as are the more obvious higher level requirements explored earlier. Project teams dominated by physical or natural scientists or engineers may not recognize this, which often exacerbates the impacts of these risks. They are; specifications, project plan, scale of effort (Stern and Arias, 2011; Jonasson, 2012).

*d) System Testing and Evaluation Risks*

Testability risks are an attribute of design risk, not a class by itself. Owing to the importance of the verification and validation of the system, it has been included as a source of risk at a higher level in this taxonomy. Just as information systems must have requirements and a design codified into a specification document, it must also have a documented test plan. Most testing will address specifications, but important validation. All test plans should include a test coverage matrix documenting just what is being tested and how. The main consequence of the sources of risk cited below is that it will not be possible to demonstrate that the code is actually fit to purpose. They are; verification, unit testing, integration testing, interoperability testing, validation (Menezes et al., 2013).

### 3.8.2 Development Environment Risks

This class addresses the sources of risk inherent in an environment and the processes used to develop a software application. The risks here are usually intrinsic, but in some instances the choice of development environment, or some of its features, is beyond the control of the code development team. This environment includes the development philosophy, (e.g., Capability Maturity Model [CMM], agile), workflow management model (e.g., incremental, iterative, spiral, and others), the development system, project management methods, and work environment. The risk elements associated with the development environment are characterized by Menezes et al., (2013) as:

*a) Development Process Risks*

This element refers to risks that can be experienced through a process or processes by which the development team proposes to satisfy the customer's requirements. The process is the sequence of steps leading from the initial requirements gathering and specification to the final delivered soft-ware product. Development processes themselves have attributes. Most conform to some degree to a development philosophy like CMM, ISO, or agile. Most development processes can also be identified with a workflow management model (called "development models" in the original SEI risk taxonomy) (Stern and Arias, 2011).

The development philosophy typically describes the approach to processes used to create a software products. Examples include formal methods favoured by CMM or ISO, and agile methods, which are often encountered in information system code development projects. CMM-endorsed processes emphasize a formal approach to the customary development phases (i.e., life-cycle elements) of requirements analysis, product design, product creation, testing, delivery, and maintenance (sometimes called "production," ending ultimately in the eventual retirement or decommissioning of the application). It includes both general management processes such as costing, schedule tracking, and personnel assignment, and also project-specific processes such as feasibility studies, design reviews, and regression testing. Importantly, advanced CMM organizations collect and utilize metrics about their own development processes with a view to process improvements. Agile methods, on the other hand, focus at a philosophical level on software development practices and people, not processes. Note that some agile methods are very prescriptive; the reference above is intended to capture the shared philosophical basis of this development methodology [Agile Software]. The lack of a methodology has been recognized as a risk in and of itself. Specific sources of risk associated with the absence of a development methodology include the absence of coherent change control, no project planning, and no repeat-able processes or practices. Of course, the adoption of or more likely, the imposition of a methodology incompatible with the goals of the project or the development team is also a source of risk (Jonasson, 2012; Menezes et al., 2013).

Workflow management models describe different approaches to the management and organization of the development project workflow elements cited above (Beck 1999). Various models have been proposed for this: waterfall (the original conceptual model for software development), incremental, iterative, evolutionary, spiral (emphasizing prototyping), and others. At the opposite end of the spectrum from the waterfall model are approaches like extreme programming (XP) and rapid application development (RAD). Note that the software engineering literature often aligns these workflow management models with certain development methodologies, e.g., RAD with agile (Carl, 2014).

This element groups risks that result from a development philosophy and/or workflow management approach that; does not reflect what is known at the beginning of the project, is not suited to the activities necessary to accomplish the project goals, is poorly communicated to the project staff and lacks enforceability

*b) Development System Risks*

The development system risk element addresses those risks related to the choices of hardware and software tools used in application development. The purpose of these tools is to facilitate application development and in some cases (such as integrated development environments) to reduce performance risk as well (for example, by introducing automated product control) (Stern and Arias, 2011).

*c) Management Process Risks*

This is the category of risks associated with planning, monitoring, and controlling budget and schedule; controlling factors involved in defining, implementing, and testing the software application; managing project personnel; and handling external organizations, including the customer, senior management, matrix management, and other contractors. It is widely recognized that management actions determine, and management is ultimately responsibility for, much of the risk associated with software development projects. Management processes must support the following central objectives: recruit the right staff, match them to the right tasks, and keep them motivated, help teams jell. Moreover, management commitment has been cited as the number *one* risk to long term project success (Jonasson, 2012; Menezes et al., 2013).

*d) Management Methods Risks*

This element refers to the risks associated with methods adopted for managing both the development of the product and program personnel. These include risks related to quality assurance, configuration management, staff development with respect to program needs, and maintaining communication about program status and needs. The continuity of management support over the life of the project is an important facet of this element. Continuity is especially challenging in view of the fact that many important scientific code development projects have a production phase that spans careers that is, decades long (Stern and Arias, 2011; Jonasson, 2012).

*e) Work Environment Risks*

This element refers to risks arising from subjective aspects of the environment such as the amount of care given to ensuring that stakeholders, including the management, users, sponsors, and the development team itself, are kept informed of program goals and information, the way they work together, their responsiveness to staff inputs, and the attitude and morale of the program personnel. A well-functioning development team has already been identified as a critical success factor for software development projects (Menezes et al., 2013; Gaol, Mars and Saragih 2014).

### 3.8.3 Programmatic Risks

Programmatic risks refer to those project risks emanating from external forces acting on system development projects. These are sources of risk that are usually outside the direct control of the code development team, that is, extrinsic risks.

a) *Resources Risks* - this element addresses sources of project risk arising from resource dependencies or constraints that the project must honour. These dependencies/constraints include schedule, staff, budget, and facilities.

b) *Contract Risks* - Risks associated with the program contract are classified according to contract type, restrictions, and dependencies.

c) *Program Interface Risks* - This element consists of the various interfaces with entities and organizations outside the development program itself (Menezes et al., 2013; Stern and Arias, 2011).

*Table 3.1: Taxonomy for Sources of Software Development Risks*

| **A. Development Cycle Risks** | **B. Development Environment Risks** | **C. Programmatic Risks** |
|---|---|---|
| 1. *Requirements Risks*<br>a. Predictability<br>b. Evolvability<br>c. Completeness<br>d. Clarity<br>e. Accuracy<br>f. Precedence | 1. *Development Process Risks*<br>a. Repeatability<br>b. Suitability<br>c. Control of Process<br>d. Familiarity with Process or Practice<br>e. Environment Change Control | 1. *Resources Risks*<br>a. Schedule<br>b. Staff<br>c. Budget<br>d. Facilities<br>e. Management Commitment |
| *2. Design Risks*<br>a. Difficulty<br>b. Modularity<br>c. Usability<br>d. Maintainability<br>e. Portability<br>f. Reliability | *2. Development System Risks*<br>a. Hardware Capacity<br>b. Development System Capability<br>c. Suitability<br>d. Usability<br>f. Reliability<br>g. Target-Unique System Support<br>h. Security | *2. Contract Risks*<br>a. Contract Type<br>b. Restrictions<br>c. Dependencies |
| 3. *Implementation Risks*<br>a. Specifications<br>b. Project Plan<br>c. Scale of Effort | *3. Management Process Risks*<br>a. Contingency Planning<br>b. Project Organization<br>c. Management Experience<br>d. Program Interfaces<br>e. Reward Systems | *3. Program Interface Risks*<br>a. Customer Communication<br>b. User Commitment<br>c. Corporate Communication<br>d. Vendor Performance<br>e. Political |
| *4. Test and Evaluation Risks*<br>a. Verification<br>i. Unit Testing<br>ii. Integration Testing<br>iii. Interoperability Testing<br>b. Validation | *4. Management Methods Risks*<br>a. Monitoring<br>b. Personnel Management (Staffing and Training)<br>c. Quality Assurance<br>d. Configuration Management | |
| | *5. Work Environment Risks*<br>a. Quality Attitude<br>b. Cooperation<br>c. Communication<br>d. Morale<br>e. Trust | |

Source: SEI, (2009)

### 3.9 Integration of Risk Management into SDLC

Effective risk management must be totally integrated into the SDLC. In some cases, an IT system may occupy several of these phases at the same time. However, the risk management methodology is the same regardless of the SDLC phase for which the assessment is being conducted. Minimizing negative impact on an organization and need for sound basis in decision making are the fundamental reasons financial institution implement a risk management process for their Information systems. Risk management is an iterative process that can be performed during each major phase of the SDLC (Stern and Arias, 2011; Carl 2014).

*Table 3.2: Characteristics of SDLC phases and risk management.*

| **Phase** | **Details** | **Support from Risk Management Activities** |
|---|---|---|
| Initiation | The need for an IT system is expressed and the purpose and scope of the IT system is documented | Identified risks are used to support development of system requirements |
| Development or Acquisition | The IT system is designed, purchased, programmed, developed, or otherwise constructed | Risks identified during this phase can be used to support the analyses of the IT system that may lead to architecture and design trade-offs during system development |
| Implementation | The system security features should be configured, enabled, tested, and verified | The risk management process supports the assessment of the system implementation against its requirements and within its modelled operational environment. Decisions regarding risks identified must be made prior to system operation |
| Operation or Maintenance | The system performs its functions. Typically the system is being modified on an ongoing basis through the addition of hardware and software and by changes to organizational processes, policies, and procedures | Risk management activities are performed for periodic system reauthorization (or reaccreditation) or whenever major changes are made to an IT system in its operational, production environment (e.g., new system interfaces) |
| Disposal | This phase may involve the disposition of information, hardware, and software. Activities may include moving, archiving, discarding, or destroying information and sanitizing the hardware and software | • Risk management activities are performed for system components that will be disposed of or replaced to ensure that the hardware and software are properly disposed of, that residual data is appropriately handled, and that system migration is conducted in a secure and systematic manner |

Source: Stern and Arias, (2011)

# CHAPTER FOUR

## 4.0 BAYESIAN BELIEF NETWORKS

### 4.1 Overview

Decision making is an important aspect of software processes management. Most organizations allocate resources based on predictions. Improving the accuracy of such predictions reduces costs and helps in efficient resources management. The application of BNs was considered impractical until recently due to the difficulty of computing the joint probability distribution even with a small number of variables. However, due to recent progresses in the theory of and algorithms for graphical models, Bayesian networks have gained importance while dealing with uncertainty and probabilistic reasoning. BNs are an ideal decision support tool for a wide range of problems and have been applied successfully in a large number of different settings such as medical diagnosis, credit application evaluation, software troubleshooting, safety and risk evaluation (Johnson, 2009; Fenton and Neil, 2011).

Bayesian networks, also known as Bayesian Belief Networks - BBN - (or Bayes nets for short) belong to the family of probabilistic graphical models (GMs) which allows the inference of a future event based on prior evidence. These graphical structures are used to represent knowledge about an uncertain domain. In particular, each node in the graph represents a random variable, while the edges between the nodes represent probabilistic dependencies among the corresponding random variables (Fenton and Neil, 2011). These conditional dependencies in the graph are often estimated by using known statistical and computational methods. Hence, BNs combine principles from graph theory, probability theory, computer science, and statistics. The variables (factors) in a BN may be at different temporal and spatial scales and the data represented in the network may originate from diverse sources such as empirical data, expert opinion and simulation outputs. As such, BNs can be effectively incorporated in a traditional risk management framework through explicitly displaying the causal web of interacting factors and the probabilities of multiple states of predictor and response variables (Johnson, 2009; Fenton and Neil, 2011; Rasmussen et al, 2013).

Holmes, (2010) indicated that Bayesian Belief Networks exploit the distributional simplifications of the network structure by calculating how probable certain events are, and how these probabilities can change given subsequent observations, or predict change given external interventions. BNs enable reasoning under uncertainty and combine the advantages of an intuitive visual representation with a sound mathematical basis in Bayesian probability. With BNs, it is possible to articulate expert beliefs about the dependencies among different variables and to propagate consistently the impact of evidence on the probabilities of uncertain outcomes (Holmes, 2010).

As decision support tools, Bayesian networks can be used to analyze complex problems, prioritize hazards, and support decision-making in an adaptive process, where knowledge is incomplete. Bayesian networks are ideal for assisting decision making where evidence is incomplete, contradictory or disparate. Unlike many other risk analysis methods, they make use of a range of data types, concepts and assumptions for which a range of evidence of varying quality exists. When this evidence is assembled in concert, the overall weight from individual threads can support in prioritizing, and where relevant, managing risks (Daly et al., 2011).

## 4.2 Probability Functions

The likelihood of an event E is indicated by Probability Function P(E); The sum of the probabilities of all elementary outcomes within sample space S, P(S) = 1, with values between 0 and 1 where:
P(E) = 1: the event is CERTAIN to occur, P(E) = 0: the event is certain NOT to occur and anything in between represents a level of belief of the certainty or uncertainty of an event to occur.

Considering the laws of probability, the probability of any event A is $0 \leq P(A) \leq 1$;

- Law of Addition: $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
- Law of Multiplication: $P(A \cap B) = P(A) \times P(B)$, in the real sense

$P(A \cap B) = P(B) \times P(A|B)$, where $P(A|B)$ is probability of A given B has occurred
If A and B are statistically independent,

P(B|A) = P(B), then

$$P(A \cap B) = P(A) \text{ x } P(B|A) = P(A)\, P(B) \qquad \text{………… (4.1)}$$

Bayes' Theorem is a trivial consequence of the definition of conditional probability, but it is very useful in that it allows us to use one conditional probability to compute another (Ho, 2008).

Given that A and B are events in sample space S, and P(B), ≠ 0, conditional probability is defined as:

$P(A \cap B) = P(A|B)\, P(B)$

$P(A \cap B) = P(B|A)\, P(A)$

$$P(B \mid A) = \frac{P(A|B)P(B)}{P(A)} = \frac{P(A \cap B)}{P(A)}$$

$P(B|A)\, P(A) = P(A|B)\, P(B)$

………… (4.2)

In general, to determine uncertainties based on evidence two approaches to estimate parameters are used; one frequentist approach and Bayesian approach. The frequentist approach is based only on observed data and an adopted model characterized by scientific objectivity while the Bayesian approach works by appropriately combining prior intuition or knowledge with information from observed data characterized by subjective nature of prior opinion. Each approach is valid when applied under specific circumstances neither approach uniformly dominates the other (Ho, 2008).

In the frequentist approach, Relative frequency λ (proportion of times an outcome occurs) is given as

$$\lambda = \frac{Number\ of\ successful\ trials}{Total\ number\ of\ trial\ (N)}$$

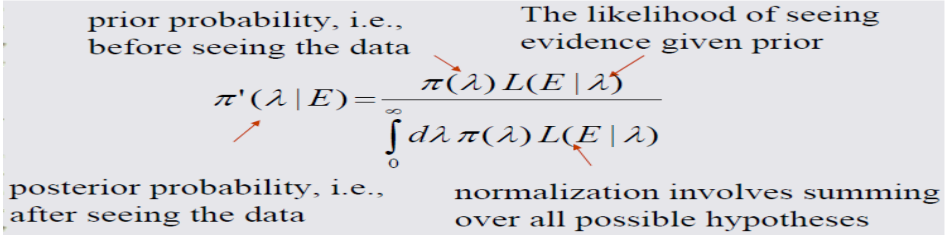
………… (4.3)

For $N \rightarrow \infty$, the relative frequency tends to stabilize around some number known as probability estimates.

Frequentist statistics will completely break down if; no data or no experience history (N =0), the advent of new technology, in the event of rare events/occurrences, when there is no failure record.

Bayesian statistics measures degrees of belief by using intuition knowledge (prior belief), updating it by evidence (likelihood) to obtain a posterior belief P(B|A) = P(A|B)*P(B) / P(A) to process knowledge, P(Cause|Effect) = P(Effect|Cause) P(Cause) / P(Effect)

Ho (2008), summarizes this as



prior probability, i.e., before seeing the data

The likelihood of seeing evidence given prior

$$\pi'(\lambda \mid E) = \frac{\pi(\lambda)\, L(E \mid \lambda)}{\int_0^\infty d\lambda\, \pi(\lambda)\, L(E \mid \lambda)}$$

posterior probability, i.e., after seeing the data

normalization involves summing over all possible hypotheses

… (4.4)

*Example 1:*

Ho, (2008), stated that to ensure successful system implementation, after a new system is developed it has to undergo a battery of tests each more rigorous than the previous to determine their reliability levels. Suppose we have a new untested system. We estimate, based on prior experience, that 80% of chance the reliability (probability of successful run) $R_1$ = 0.95, and 20% of chance that $R_2$ = 0.75. We run a test and find that it operate successfully. What is the probability that the reliability level is $R_1$.

$S_i$ = event System test results in a success

$$P(R_1 | S_1) = \frac{P(R_1)P(S_1 | R_1)}{P(R_1)P(S_1 | R_1) + P(R_2)P(S_{|1} | R_2)}$$

………… (4.5)

=0.8*0.95/(0.8*0.95 + 0.2*0.75) = 0.835 (updated from 0.8)

Our prior probability of the first system being successful has been updated to 0.8 which is slightly lower based on the difference our prior wasn't far off the mark, that is there is a higher chance that the reliability level is $R_{1.}$

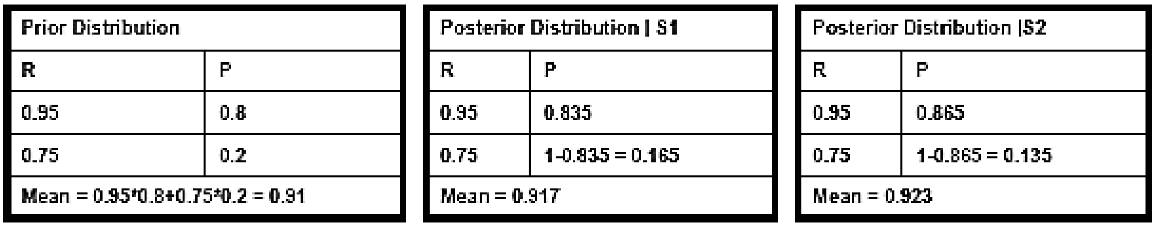
Supposed we conduct the second test and assume that it is also successful the probability that the reliability level is $R_1$ is computed as

$$P(R_1 | S_2) = \frac{P(R_1)P(S_2 | R_1)}{P(R_1)P(S_2 | R_1) + P(R_2)P(S_2 | R_2)}$$ ………… (4.6)

= 0.835*0.95/(0.835*0.95 + (1-0.835)*0.75)

= 0.79325/(0.79325+0.12375) = 0.865

*Table 4.1: Prior and posterior Distributions*



| Prior Distribution | |
|---|---|
| R | P |
| 0.95 | 0.8 |
| 0.75 | 0.2 |
| Mean = 0.95*0.8+0.75*0.2 = 0.91 | |

| Posterior Distribution \| S1 | |
|---|---|
| R | P |
| 0.95 | 0.835 |
| 0.75 | 1-0.835 = 0.165 |
| Mean = 0.917 | |

| Posterior Distribution \|S2 | |
|---|---|
| R | P |
| 0.95 | 0.865 |
| 0.75 | 1-0.865 = 0.135 |
| Mean = 0.923 | |

Based on the calculations above and given the prior we conclude that after two successful tests there is a higher chance with a mean of approximately 0.9 that the reliability level is $R_{1.}$

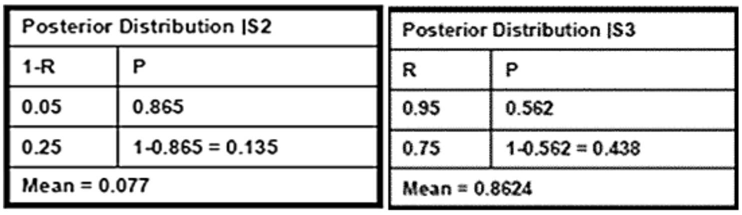
Lets further determine what is P(R1|S3) is if the third test S3 is a failure: that is, S3 = event System test results in a failure

$$P(1-R_1) | S_3) = \frac{P(1-R_1)P(S_3 | 1-R_1)}{P(1-R_1)P(S_3 | 1-R_1) + P(1-R_2)P(S_3 | 1-R_2)}$$ ………… (4.7)

P(1-R1|S3)= 0.865*0.05/(0.865*0.05 + (1-0.865)*0.25)

=0.04325/(0.04325+0.03375)=0.562

*Table 4.2: Posterior Distributions*



| Posterior Distribution \|S2 | |
|---|---|
| 1-R | P |
| 0.05 | 0.865 |
| 0.25 | 1-0.865 = 0.135 |
| Mean = 0.077 | |

| Posterior Distribution \|S3 | |
|---|---|
| R | P |
| 0.95 | 0.562 |
| 0.75 | 1-0.562 = 0.438 |
| Mean = 0.8624 | |

Upon running a third test which is a failure, we note that the probability of the reliability level being $R_1$ significantly reduces to 0.562 with a mean of 0.86, this indicates that the third test has more bearing on reliability level than the first two tests.

## 4.3 Bayesian Network Theory and Propagation

The underlying theory of BNs combines Bayesian probability theory and the notion of conditional independence to represent dependencies among variables. The essence of

Bayesian methods is a mathematical rule explaining how we should change our existing beliefs in the light of new evidence. Being probabilistic, Bayesian networks readily incorporate uncertain with uncertainties being reflected in the conditional probabilities defined for linkages which classical models do not permit. A Bayesian network is particularly useful as Bayesian inference provides a probability based approach that can update knowledge when new information becomes available. Bayesian networks exploit the distributional simplifications of the network structure by calculating how probable certain events are, and how these probabilities can change given subsequent observations, or predict change given external interventions. Bayes' Theorem was developed by the Rev. Thomas Bayes, an 18th century mathematician and theologian, and was first published in 1763 (Koski and Noble, 2012; Holmes, 2010).

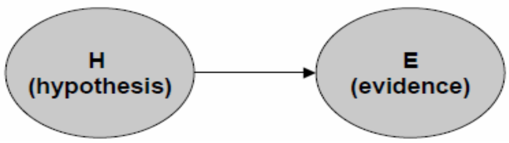


*Figure 4.1: casual view of evidence*

$$P(H|E,c) = \frac{P(H|c)*P(E|H,c)}{P(E|c)} \qquad \text{………… (4.8)}$$

We update our belief in hypothesis H given on additional evidence E with background context c. Left-hand term - P(H|E,c) - is known as posterior probability or the probability of H after considering the effect of E on c. The term P(H|c) is called the prior probability of H given c alone. The term P(E|H,c) is called the likelihood and gives the probability of the evidence assuming the hypothesis H and the background information c is true. Finally, the last term P(E|c) is independent of H and can be regarded as a normalizing or scaling factor. Formally, we start with a prior probability P(H) for the hypothesis H. The likelihood, for which we also have prior knowledge, is formally the conditional probability of E given H, which we write as P(E|H). Bayes’s theorem provides the correct formula for updating our prior belief about H in the light of observing E. In other words Bayes calculates P(H|E) in terms of P(H) and P(E|H). There exists various expressions of the same equation which do not change the meaning nor outcome depending on the context including:

$$P(H \mid E) = \frac{P(E \mid H)P(H)}{P(E)} = \frac{P(E \mid H)P(H)}{P(E \mid H)P(H) + (E \mid notH)P(notH)}$$

………… (4.9)

The most common mathematical definition of Bayes theory being if A and B are events and P(B), the probability of event B, is greater than zero, then:

$$P(B|A) \quad = \quad \frac{P(\mathrm{A}|\mathrm{B})\, P(\mathrm{B})}{P(\mathrm{A})} = \frac{P(\mathrm{A}|\mathrm{B})\, P(\mathrm{B})}{P(\mathrm{A}|\mathrm{B})\, P(\mathrm{B}) + P(\mathrm{A}|\neg\mathrm{B})\, P(\neg\mathrm{B})}$$

………… (4.10)

*Where: P(A) is prior probability of event A; P(B) is prior probability of event B; P(B|A) is conditional probability of B given A; P(A|B) is conditional probability of A given B.*

*Example: 1-* Assume in a thousand lines of code (KLOC) there is one defect D. Then:

P(D) = 0.001,

So P(not D) = 0.999.

Also assume a test for the defects has 100% sensitivity (i.e. no false negatives) and 95% specificity (meaning 5% false positives). Then if E represents the Boolean variable “Test positive for the defects”, we have:

*P(E | not D) = 0.05*

*P(E | D) = 1*

Now suppose a randomly selected line of code tests positive for defect. What is the probability that that line of code actually has a defect? Using Bayes theory:

$$P(B|A) \quad = \quad \frac{P(\mathrm{A}|\mathrm{B})\, P(\mathrm{B})}{P(\mathrm{A})} = \frac{P(\mathrm{A}|\mathrm{B})\, P(\mathrm{B})}{P(\mathrm{A}|\mathrm{B})\, P(\mathrm{B}) + P(\mathrm{A}|\neg\mathrm{B})\, P(\neg\mathrm{B})}$$

$$= \frac{1 \times 0.001}{1 \times 0.001 + 0.05 \times 0.999} = 0.01963$$

………… (4.11)

So there is a less than 2% chance that that line of code actually has a defect.

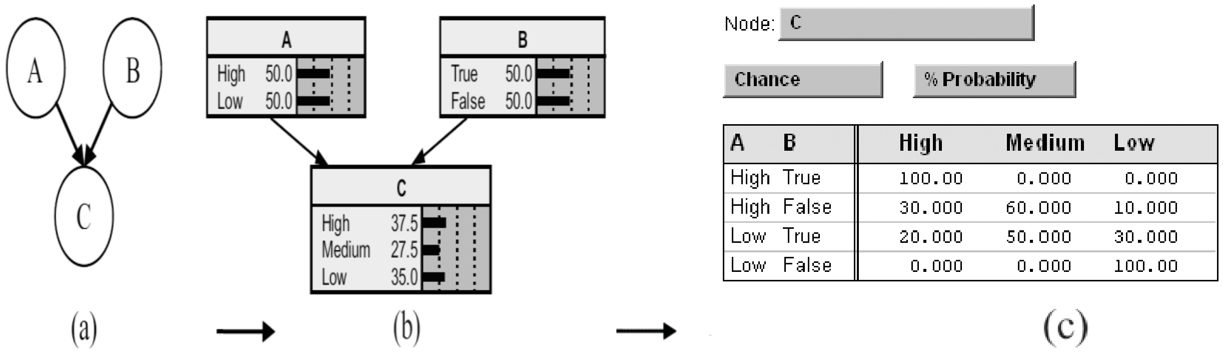
While Bayes theorem is the only rational way of revising beliefs in the light of observing new evidence, it is not easily understood by people without a statistical/mathematical background. Moreover, the results of Bayesian calculations can appear, at first sight, as counter-intuitive (for instance the above example most people would give an incorrect “intuitive” answer of 95%). In many cases, lay people only accept Bayes theorem as being ‘correct’ and are able to reason correctly, when the information is presented in alternative graphical ways, such as using event trees and frequencies. But these alternative presentation techniques do not scale up to more complex problems (Fenton and Neil, 2011; Daly et al., 2011).

### 4.3.1 Bayesian Network Structure

A Bayesian network consists of two main parts; (1) a graphical structure that defines a set of dependence and independence statements over a set of random variables representing entities of a problem domain and (2) a set of CPDs specifying the strengths of the dependence relations encoded in the graphical structure. A Bayesian Network consists of a directed acyclic graph of nodes and links or arcs that conceptualize a system. The values of the nodes are defined in terms of different, mutually exclusive states. The relationships between nodes are described by conditional probability distributions that capture the dependences between variables. (Holmes, 2010; Rasmussen, 2013).

If there is a link going from node A to node C, then A is said to be a parent node of C, and C is said to be a child node of A. In the figure 4.2 (a) parent nodes A and B represent the causal factors of child node C. The states of nodes A to C, arbitrarily selected for ease of demonstration. The conditional relationship between parent nodes A and B and child node C is defined by a conditional probability table (CPT). The figure 4.2 is interpreted as the probability that C will be in its High, Medium and Low states, given the states of A and B.

*Figure 4.2: Bayesian Network Structure*

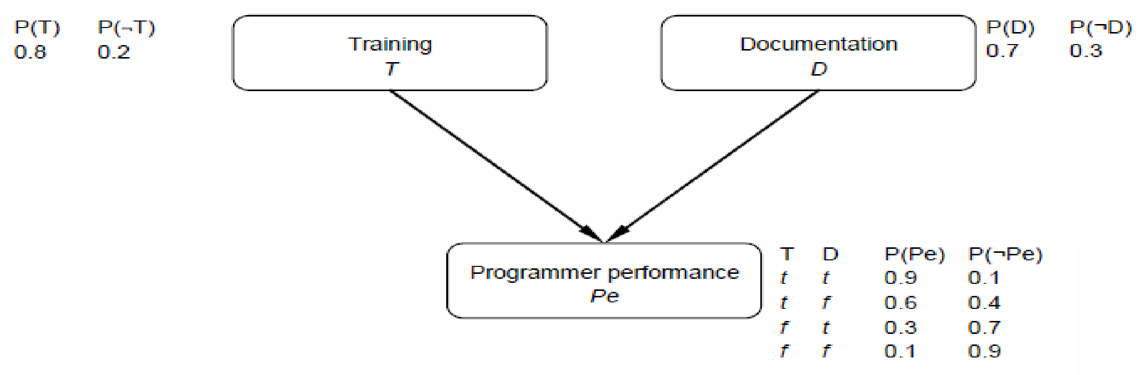


| A | B | High | Medium | Low |
|---|---|---|---|---|
| High | True | 100.00 | 0.000 | 0.000 |
| High | False | 30.000 | 60.000 | 10.000 |
| Low | True | 20.000 | 50.000 | 30.000 |
| Low | False | 0.000 | 0.000 | 100.00 |

### 4.3.2 BN sample application in system development

Radlinski, (2008) indicated that programmer performance was determined by the programmers training and the quality of documentation. Considering this two factors, figure 4.3 is a model for programmer performance based on the training and documentation quality.

*Figure 4.3: Sample application in system development*



To calculate the probability that the programmer will achieve good performance based on prior probabilities without entering any observations:

$$\begin{aligned}
&P(Pe) = P(T) * P(D) * P(Pe|T, D) + \\
&P(T) * P(\neg D) * P(Pe|T, \neg D) + \\
&P(\neg T) * P(D) * P(Pe|\neg T|, D) + \\
&P(\neg T) * P(\neg D) * P(Pe|\neg T, \neg D) = \\
&0.8 * 0.7 * 0.9 + 0.8 * 0.3 * 0.6 + 0.2 * 0.7 * 0.3 + 0.2 * 0.3 * 0.1 = \\
&0.504 + 0.144 + 0.042 + 0.006 = 0.696.
\end{aligned}$$

………… (4.12)

From the equation it is predicted that the probability that the programmer will achieve good performance is 0.696. This probability is based only on prior knowledge and forms the basis for initial computations and predictions, it however changes (updated) in light of new knowledge or information.

*Observation 1:* Suppose the programmer receives good documentation; note that documentation does not cause any change in the training node because they are conditionally independent given programmer performance, we update our prior belief using bayes theorem as;

$$\begin{aligned}
&P(Pe\,|\,D) = P(T) * P(Pe\,|\,T, D) + P(\neg T) * P(Pe\,|\,\neg T, D) = \\
&0.8 * 0.9 + 0.2 * 0.3 = 0.72 + 0.06 = 0.78.
\end{aligned}$$

………… (4.13)

Based on this new observation, the probability of achieving good performance is revised to 0.78. In this regard the management would have to carry out corrective measures in regards to training by either hiring better trained programmers or carrying out in house training seeing. Note that these nodes do not work in isolation for instance there would be other nodes to help in making decisions considering in-house training vs hiring highly trained programmers.

*Observation 2:* Let us now further assume that although good documentation was received, a bad programmer performance has been observed, by applying Bayes' rule:

$$P(T \mid D, \neg Pe) = \frac{P(\neg Pe \mid T, D) * P(T)}{P(\neg Pe \mid D)} = \frac{0.1 * 0.8}{1 - 0.78} \approx 0.364 \qquad \ldots\ldots\ldots\ldots (4.14)$$

The model revises the likelihood of good training given this new observation to around 0.364 which is much lower than initially assumed (0.78). Hence the most likely explanation for the programmer performance is good training rather than receiving good documentation. This particular type of (backward) inference is called explaining away. Classical statistics alone does not enable this type of reasoning and "what-if" analysis (Fenton and Neil, 2011).

In the more complex real-life BN models that found in financial information system development, it is impossible to perform the Bayesian inference calculations manually. Fortunately, there are various efficient algorithms (exact and approximate) implemented in BN toolkits to do this (Radlinski, 2008).

Different types of nodes can be included in a BN including:

- Nature nodes which are variables that can be controlled by actions of the decision-maker and are used to represent the empirical or calculated parameters and the probabilities that various states will occur.

- Decision nodes represent control variables or events that can directly be implemented by the decision maker (for example, risk control measures). These nodes typically represent the suite of available management actions. Decision nodes should always be accompanied by utility nodes.

- Utility nodes represent the value of the decisions or outcomes. They can be linked directly to the decision node, or to the outcome node (for example, benefits in). The utility nodes are used to assess the optimal decision rules in the network that will maximize the sum of expected values of the utility nodes (Holmes, 2010; Koski and Noble, 2012).

### 4.3.3 Discretization of Nodes

In order to represent continuous relationships in a Bayesian network, a continuous variable must be divided or discretized into a set of states (Holmes, 2010). States can be qualitative or quantitative, categorical (e.g. absent vs. Present; 0 vs 1) or continuous (represented as a set of discrete intervals), where numerical ranges are assigned (e.g. 0 to 3, 3 to 10). Nodes can be discretized according to guidelines, existing classifications or percentiles of data. There is no limit on how many states can be defined, but it is important to note that as the number of states increase, so do the number of probabilities to be estimated (Pollino and Henderson 2010; Holmes, 2010).

### 4.3.4 Representing the joint probability distribution

Most commonly, BNs are considered to be representations of joint probability distributions. There is a fundamental assumption that there is a useful underlying structure to the problem being modeled that can be captured with a BN, i.e., that not every node is connected to every other node. If such domain structure exists, a BN gives a more compact representation than simply describing the probability of every joint instantiation of all variables. Sparse Bayesian networks (those with relatively few arcs, which means few parents for each node) represent probability distributions in a computationally tractable way (Ho, 2008). Consider a BN containing the *n* nodes, $X_1$ to $X_n$, taken in that order.

A particular value in the joint distribution is represented by $P(X_1 = x_1; X_2 = x_2 \ldots\ldots\ldots X_n = x_n)$, or more compactly, $P(x_1; x_2 \ldots\ldots\ldots x_n)$. The chain rule of probability theory allows us to factorize joint probabilities so:

$$\begin{aligned} P(x_1, x_2, \ldots, x_n) &= P(x_1) \times P(x_2|x_1) \ldots, \times P(x_n|x_1, \ldots, x_{n-1}) \\ &= \prod_i P(x_i|x_1, \ldots|, x_{i-1}) \end{aligned} \quad \ldots\ldots\ldots\ldots (4.15)$$

Consider that the structure of a BN implies that the value of a particular node is conditional only on the values of its parent nodes, this reduces to

$$P(x_1|,x_2,\ldots,x_n) \quad = \quad \prod_i P(x_i|Parents(X_i))$$

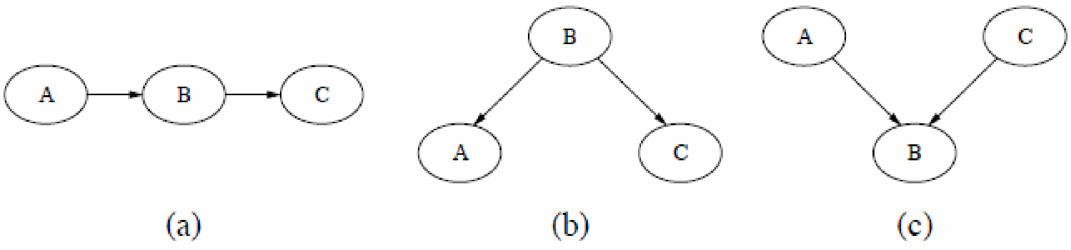
Provided:

$$Parents(X_i) \subseteq \{X_1,\ldots,X_{i-1}\}.$$ ………… (4.16)

### 4.3.5 Conditional Independence Assumptions in Bayesian Networks

Another way to view a Bayesian network is as a compact representation for a set of conditional independence assumptions about a distribution. These conditional independence assumptions are called the local Markov assumptions. While the researcher will not go into the full details here, the relation between conditional independence and Bayesian network structure is important for understanding how BNs work.

*Figure 4.4: Conditional Independence*



(a) Causal chain; (b) common cause; (c) common effect.

Causal chains - Consider a causal chain of three nodes, where *A* causes *B* which in turn causes *C*, as shown in figure (a). Causal chains give rise to conditional independence.

$$P(C|A \wedge B) = P(C|B)$$ ………… (4.17)

This means that the probability of *C*, given *B*, is exactly the same as the probability of *C*, given both *B* and *A*. Knowing that *A* has occurred doesn’t make any difference to our beliefs about *C* if we already know that *B* has occurred. This conditional independence is also written as:

$$A \perp\!\!\!\perp C|B.$$ ………… (4.18)

Common causes - Two variables *A* and *C* having a common cause *B* is represented in Figure (b). Common causes (or common ancestors) give rise to the same conditional independence structure as chains:

$$P(C|A \wedge B) =| P(C|B) \equiv A \perp\!\!\!\perp C|B \qquad \text{…………} (4.18)$$

Common effects – this is represented by a network v-structure, as in Figure (c). This represents the situation where a node (the effect) has two causes. Common effects (or their descendants) produce the exact opposite conditional independence structure to that of chains and common causes. That is, the parents are marginally independent $(A \perp\!\!\!\perp C)$, but become dependent given information about the common effect (i.e., they are conditionally dependent):

$$P(A\|C \wedge B) \neq P(A|B) \equiv A \not\perp\!\!\!\perp C|B \qquad \text{…………} (4.19)$$

Thus, if we observe the effect, and then, say, we find out that one of the causes is absent, this raises the probability of the other cause which is just the inverse of explaining away.

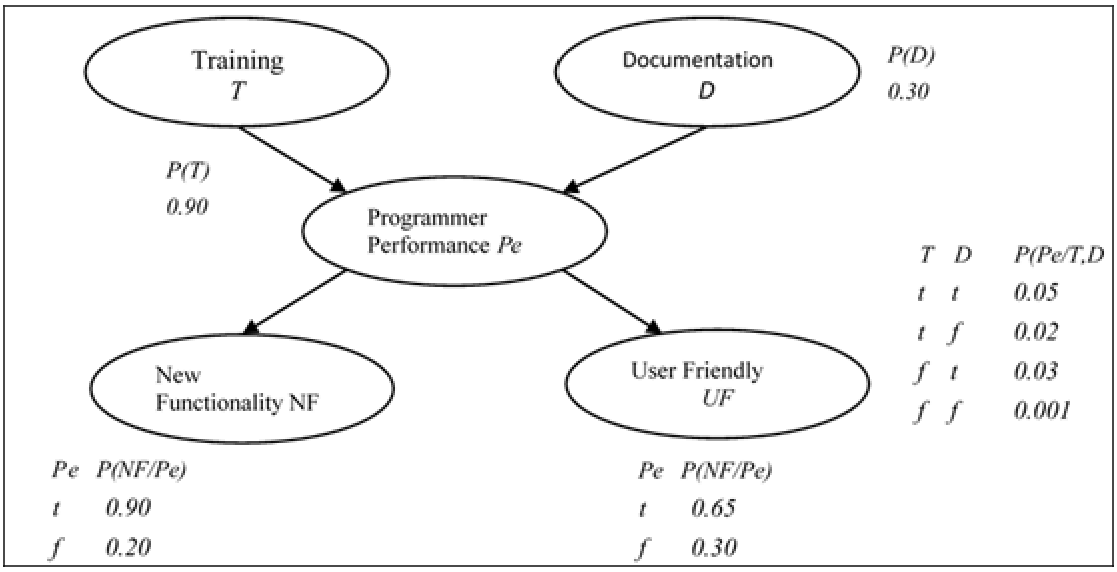
*Example:*

In a BN structure nodes should be connected directly if one affects or causes the other, with the arc indicating the direction of the effect for instance programmer performance is affected by the programmers training and the quality of training, consequently, the programmers performance will determine whether or not a new functionality will be added and the user friendliness of the module. Note that this BN topology is a tentative representation of the domain used for demonstration in this study based on existing literature and represent only one approach of the many existing.

The example shown in Figure 4.5 is aimed as going through the vital BN processes.

Consider our earlier programmer performance example with the following nodes and priors. Assuming that all nodes have binary discrete codes to represent the domain;

*Figure 4.5: BN for programmer performance*



In BN any node without parents is called a root node, while any node without children is called a leaf node. Any other node (non-leaf and non-root) is called an intermediate node. Given a causal understanding of the BN structure, this means that root nodes represent original causes, while leaf nodes represent final effects. In the example, the causes training and Documentation are root nodes, while the effects new functionality and user friendliness are leaf nodes. Conventionally, BN structures are usually laid out so that the arcs generally point from top to bottom. This means that the BN “tree” is usually depicted upside down, with roots at the top and leaves at the bottom.

Once the BN structure has been laid out, we now need to quantify the relationships between connected nodes, this is done by specifying a conditional probability distribution for each node, and this takes the form of a conditional probability table (CPT). We need to look at all the possible combinations of values of those parent nodes. Each such combination is called an instantiation of the parent set. For each distinct instantiation of parent node values, we need to specify the probability that the child will take each of its values (Barber, 2012). For instance the parent nodes for programmer performance are Training and documentation and take the possible joint values (t,t; t,f; f,t; f,f). The conditional probability table specifies in order the probability of programmer performance for each of these cases to be: (0:05; 0:02; 0:03; 0:001). Since these are probabilities, and must sum to one over all possible states of the programmer performance variable, the probability of low programmer performance is already implicitly given as one minus the above probabilities in each case; i.e., the probability of low programmer performance in the four possible parent instantiations is (0:95;0:98;0:97;0:999).

Root nodes also have an associated CPT, although it is degenerate, containing only one row representing its prior probabilities. In the example the prior for getting quality documentation is given as 0.3, indicating that 30% of all documentation received is desired quality while 90% of all programmers have had high quality training. The size of the CPT is, exponential in the number of parents. Thus, for Boolean networks a variable with n parents requires a CPT with $2^{n+1}$ probabilities.

BN are used to reason about the domain particularly when we observe the value of some variable, we would like to condition upon the new information. The process of conditioning (also called probability propagation or inference or belief updating) is performed via a flow of information through the network. Note that this information flow is not limited to the directions of the arcs. In the probabilistic system, this becomes the task of computing the posterior probability distribution for a set of query nodes, given values for some evidence (or observation) nodes (Nielsen and Jensen, 2013).

Bayesian networks provide full representations of probability distributions over their variables. That implies that they can be conditioned upon any subset of their variables, supporting any direction of reasoning (Ho, 2008). The types of reasoning in BN include: diagnostic, predictive, and inter-causal reasoning.

Diagnostic reasoning; reasoning from symptoms to cause, such as system tests realize that new functionality was not added then we update our belief about the programmers' performance and whether he has had quality training. This reasoning occurs in the opposite direction to the network arcs (Barber, 2012).

Predictive reasoning; reasoning from new information about causes to new beliefs about effects, following the directions of the network arcs. For example, the programmer may notice that the quality of documentation is low before coding, the programmer knows this will increase the chances of developing a non-user friendly system (Nielsen and Jensen, 2013).

Inter-causal reasoning/ explaining away; involves reasoning about the mutual causes of a common effect. Suppose that there are exactly two possible causes of a particular effect, represented by a v-structure in the BN; for instance training and documentation have a common effect, programmer performance. These two causes are independent of each other, suppose however we observe that programmer performance was low, this will raise our probability for both possible causes of programmer performance, increasing the chances that both the training and the document quality were low. Suppose then that we discover that the documentation quality was low, this will explain the low programmer performance and in turn lowers the probability of poor training. So, even though the two causes are initially independent, with knowledge of the effect the presence of one explanatory cause renders an alternative cause less likely. In other words, the alternative cause has been explained away (Ho, 2008; Barber, 2012).

Since any nodes may be query nodes and any may be evidence nodes, sometimes the reasoning does not fit neatly into one of the types described above, they can be combined in any way (Nielsen and Jensen, 2013).

## 4.4 Features of Bayesian Networks

### 4.4.1 Inference

A belief distribution in a Bayesian network experience changes when a new knowledge arrives to the network. The changes are computed through the inference process. The basic task of any probabilistic inference system can be regarded as a task to compute the posterior probability distribution for a set of query variables, given the exact values for some evidence variables (Poole and Ramon, 2014). One of the most important features of Bayesian networks is the fact that they provide an elegant mathematical structure for modeling complicated relationships among random variables while keeping a relatively simple visualization of these relationships. They also have the ability to update the beliefs of each random variable via bi-directional propagation of new information through the whole structure. This was initially achieved by an algorithm proposed by Pearl (1988) that fuses and propagates the impact of new evidence providing each node with a belief vector consistent with the axioms of probability theory (Daly, 2011; Larrañaga et al., 2013).

Once the basic assumptions as to how variables interact with each other is formed (i.e. the probabilistic model is constructed) all questions of interest are answered by performing inference on the distribution (Nielsen and Jensen, 2013).

Inference, or model evaluation, is the process of updating probabilities of outcomes based upon the relationships in the model and the evidence known about the situation at hand. After inference, the updated probabilities reflect the new levels of belief in (or probabilities of) all possible outcomes coded in the model. These beliefs are mediated by the original assessment of belief performed by the author of the model (Daly, 2011).

Because a Bayesian network is a complete model for the variables and their relationships, it can be used to answer probabilistic queries about them. For example, the network can be used to find out updated knowledge of the state of a subset of variables when other variables (the evidence variables) are observed. This process of computing the posterior distribution of variables given evidence is called probabilistic inference. A Bayesian network can thus be considered a mechanism for automatically applying Bayes‟ theorem to complex problems. In the application of Bayesian networks, most of the work is related to probabilistic inferences (Poole and Ramon, 2014).

Any variable updating in any node of Bayesian networks might result in the evidence propagation across the Bayesian networks. How to examine and execute various inferences is the important task in the application of Bayesian networks. Various types of inference algorithms exist for Bayesian. Each class offers different properties and works better on different classes of problems, but it is very unlikely that a single algorithm can solve all possible problem instances effectively (Daly, 2011).

In Bayesian networks, inference can be classified into four popular categories identified as: forward inference, backward inference, inter-causal inference, and mixed inference.

Forward Inference - also called predictive inference (from causes to effects). The inference reasons from new information about causes to new beliefs about effects, following the directions of the network arcs.

Backward Inference - also called diagnostic inference (from effects to causes). The inference reasons from symptoms to cause, Note that this reasoning occurs in the opposite direction to the network arcs (Grover, 2012).

Intercausal Inference - Intercausal inferences is also called explaining away (between parallel variables). The inference reasons about the mutual causes (effects) of a common effect (cause).

Mixed inference - also called combined inference. In complex Bayesian networks, the reasoning does not fit neatly into one of the types described above. Some inferences are a combination of several types of reasoning (Fenton and Neil, 2013).

For complex models in Bayesian networks, there are single-connected networks, multiple connected, or event looped networks. It is possible to use some methods, such as Triangulated Graphs, Clustering and Join Trees; to simplify them into a polytree. Once a polytree is obtained, the inference can be executed by the following approaches. Polytrees have at most one path between any pair of nodes; hence they are also referred to as singly-connected networks.

Some complex applications are too challenging for exact inference, and require approximate solutions (Larrañaga et al., 2013).

Inference can also be categorized into Exact and approximate inference algorithms. Exact inference algorithms include (Polytree Algorithm, Clustering, Conditioning, Arc Reversal, Elimination, Symbolic, Differential Method). Besides these general exact inference algorithms, there are some exact special case inference algorithms including quickscore for two-level networks with noisy- OR gates, and algorithms exploiting local structures in the distributions such as causal independency, context-specific independencies. In general all exact Bayesian network inference algorithms share a running time exponential in the size of the largest clique of the triangulated moral graph, which is also called the induced width of the graph. For graphs with many loops, this parameter is large and so rules out the use of exact inference algorithm.

Faced with the intractability of exact inference to large, complex networks, many researchers have investigated approximate inference algorithms. Approximate inference include stochastic simulation algorithms, model simplification methods, search-based methods and loopy belief propagation (Larrañaga et al., 2013; Koski and Noble, 2012).

In a cyclic network, the propagation or inference process will face problem in reaching a stable equilibrium state. The message passing scheme (algorithm for performing inference, it calculates the marginal distribution for each unobserved node, conditional on any observed nodes) will cause the inference process goes indefinitely. There are several methods that can be used to solve the problem of the cycle in a Bayesian network. They are clustering, conditioning and stochastic simulation Grover, 2012). Clustering involves forming compound variables in such way that the resulting network of cluster is singly connected. Conditioning involves breaking the communication pathway along the loops by instantiating a selected group of variables. Stochastic simulation involves assigning each variables a definite value and having each processor inspect the current state of its neighbours, compute the belief distribution of its variables, and select one value at random from the computed distribution. From the three approaches to handle a loop in Bayesian network, stochastic simulation gives the best estimation of the posterior probability. However, it suffers from the complexity of the calculation (Koski and Noble, 2012; (Fenton and  Neil, 2013).

Belief updating can be done using a number of exact and approximate inference algorithms, choosing different algorithms can affect the efficiency of both the knowledge engineering process and the automated reasoning in the deployed system. However, most existing BN software packages use essentially the same algorithm and it is quite possible to build and use BNs without knowing the details of the belief updating algorithms (Daly, 2011).

### 4.4.2 Learning

A belief network gives a probability distribution over a set of random variables. We cannot always expect an expert to be able to provide an accurate model; often we want to learn a network from data. Learning a belief network from data can mean many different things depending on how much prior information is known and how complete the data set is. In the simplest case, the structure is given, all of the variables are observed in each example, and only the probabilities must be learned. At the other extreme, you may not even know what variables should be hypothesized to account for the data, and there may be missing data, which cannot be assumed to be missing at random (Poole and Mackworth, 2010)

The goal of learning a BN is to determine both the structure of the network (structure learning) and the set of CPTs (parameter learning). Since the number of possible structures is extremely huge, structure learning often has high computational complexity. Thus, heuristic and approximate learning algorithms are a more realistic solution.

The structure and conditional probabilities necessary for characterizing a BN can be provided either externally by experts, which is time consuming and prone to error, or by automatic learning from a database of samples. The task of learning a BN can be divided into two subtasks; structural and parameter learning (Grover, 2012).

Structural learning, i.e., identification of the topology of the BN; The structure of the BN can be constructed manually by the subject expert or through structure learning algorithms - PC (Path Condition) and NPC (Necessary Path Condition) algorithms. The basic idea of these constraint-based algorithms is to derive a set of conditional independence and dependence statements (CIDs) by statistical tests among the nodes of the BN (Grover, 2012).

Parameter Learning: The CPTs (or parameters) can be specified, based on the knowledge of the domain expert, by the process of parameter elicitation. The past data may also be used as the basis for learning the parameters using efficient algorithms. The Expectation Maximization (EM) algorithm is particularly suitable for batch parametric learning, while Adaptation algorithms are useful for sequential parameter updates. This entails estimation of the numerical parameters (conditional probabilities) for a given network topology by estimating the CPT at each node, given the link structures and the data. Parameter learning is based on Bayesian learning algorithms that aim to find the maximum likelihood for the CPTs in a given BN. Of course, “sufficient” observations are needed to enable an estimation of conditional probabilities and the availability of “enough” observed data is precisely a limitation in many management issues (Daly, 2011). If there are lots of missing observations, BNs can use complex learning algorithms to learn the tables. The distribution of the missing data needs to be defined and may be dependent on the states of other variables or they can be randomly distributed. BNs can yield good prediction accuracy using learning algorithms, even if sample sizes are small (Vaněk, 2012; Larrañaga et al., 2013).

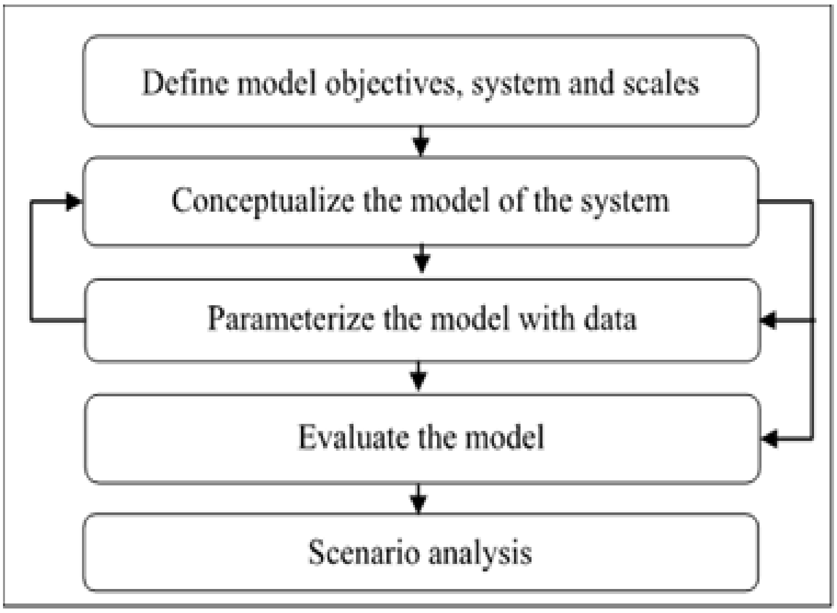
### 4.4.3 Decision-making

To incorporate decision making capabilities, the BN is converted to an influence diagram (ID) by adding decision nodes and utility nodes. The values taken by the decision nodes inform the actions which must be chosen by the decision maker. A utility node quantifies the usefulness of the outcomes resulting from the actions of decision (Bashar et al., 2010; Fenton and Neil, 2013).

## 4.5 Bayesian Network development

Prior to Bayesian network development some underlying process in the risk management will be used to generate a database of observed cases as well as domain expert experience and knowledge. The task of BN development therefore will be to fuse these information sources in order to induce a representative model of the underlying process. Model development is an iterative process that may need to be repeated several times before a valid and useful BN is established (Farmani et al., 2009; Feng and Xie, 2011). The Figure outlines the major generic steps in constructing a BN.

*Figure 4.6: BN development process*


Define model objectives, system and scales
Conceptualize the model of the system
Parameterize the model with data
Evaluate the model
Scenario analysis

Source: Ticehurst et al., (2008).

### *i) Model objectives*

Any model development process should start with a definition of the model's objective and the scope of the system to be considered. First of all, there needs to be agreement about the aim of the model, the system under consideration and the issues involved. Model developers generally need to decide on the selection of stakeholders that will be consulted in the modeling process. These could range from system developers to end users.

Various stakeholders may consider a multitude of issues related to the system, which could lead to different modeling objectives for different stakeholders. For instance where scientists may be interested in increasing their understanding of the system, decision makers may be more concerned with prediction or forecasting. The issues considered in the model will affect the management decisions that will be included in the Bayesian network. Engagement with end-users is required to ensure that management scenarios to be considered are relevant to stakeholders (Grover, 2012).

The definition of the system under consideration may also differ between stakeholders and even between the different disciplines involved in developing a Bayesian model. Agreement is needed about the spatial and temporal scales that are relevant to the system. The scope of the system needs to be defined in terms of the assets or values that will be considered in the modeling. This first phase of model development should result in a clear picture of the system that is to be modeled, its scale and scope, the discrete environmental condition or endpoint, which stakeholders will be involved and the management scenarios that are relevant to the system (Rasmussen et al., 2013).

### *ii) Conceptual model development*

When the model's objectives are defined, a conceptual BN can be developed. The initial conceptualization includes; identifying the important system variables; and establishing the links between variables. Identifying the variables ('nodes') that are important for the system that is being modeled is typically based on a literature review, expert opinion and consultation with stakeholders. Included nodes should at least be measurable, observable or predictable and should have unambiguous definitions. Nodes should be defined such that all model users understand what variable is represented. Once the variables are chosen, the links between them need to be identified. It is recommended that the number of parent nodes is kept to three or fewer, to limit the size of the CPT (Fenton and  Neil, 2013)

The identification of nodes and the links between them should result in a conceptual influence diagram representing the system under consideration. Conceptual models should capture the objective and scales of the model, provide a clear (graphical) representation of the system and address stakeholder concerns and needs. Conceptual models can assist with clarifying system understanding and identifying priorities and knowledge gaps (Rasmussen et al., 2013).

*iii) Parameterizing the model*

This phase involves assigning states and probabilities to each variable. The states for each node represent the potential values or conditions that the node can assume. States can be of different types, such as one numerical value, an interval, a probability distribution or a categorical definition (Bessiere et. al., 2013). The state types and the number of states for nodes are based on the type and quality of data available, and on the level of model parsimony desired by model developers and its users. Both node state types and 'coarseness' are fine-tuned at the model evaluation stage. The initial starting values for each node can be elicited from literature, using existing data sets or models or by discussions with experts or stakeholders.

Once the state type and number of states have been defined, the conditional probabilities for the states of each child node are specified for all combinations of states of their parent nodes. A prior expectation of the probability of a node being in a certain state can be elicited from known frequencies, or can assume a uniform distribution to represent total uncertainty. The estimation of probabilities associated with each state can be elicited from experts, obtained from existing process models, learned from data or a combination of these three sources. Uncertainties associated with each relationship are quantified in the probability distribution (Pollino et al, 2007).

*iv) Model evaluation and testing*

After developing the model's structure and estimating the conditional probabilities, the BN needs to be evaluated. Model evaluation tools include qualitative feedback from experts and stakeholders, or by comparing model predictions with literature data or with results from similar models. Quantitative model evaluation should include sensitivity analyses and assessments of predictive accuracy. Predictive accuracy refers to a quantitative evaluation of the model, by comparing model predictions with observed data (Pollino et al, 2007). Sensitivity analysis tests the sensitivity of model outcomes to variations in model parameters. Sensitivity analysis in BNs can measure the sensitivity of outcome probabilities to changes in input nodes or other model parameters, such as changes in node's type of states and their coarseness. Sensitivity analysis can be performed using two types of measures; entropy and Shannon's measure of mutual information (Barton et al, 2008).

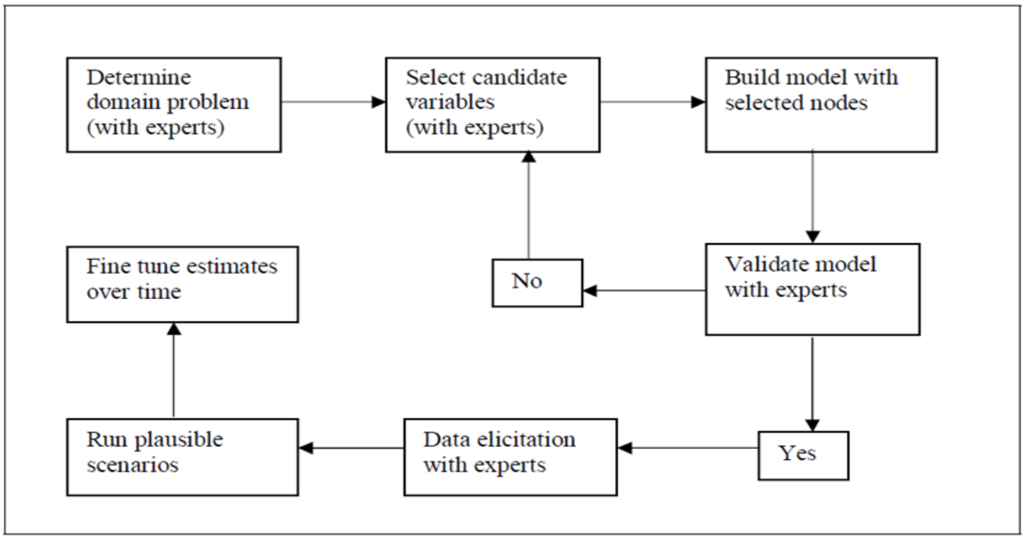
An additional empirical approach to sensitivity analysis, based on changing each of the parameters and observing the related changes in the posterior probabilities. This approach can be used to identify the most 'sensitive set' of variables in the BN; those that are most influential in affecting change and those that are most affected by variations in parameters.
Note that assessing the influence of every single parameter can be a time-consuming process, especially in large networks (Grover, 2012; Poole and Ramon 2014).

*v) Scenario analysis*

BNs can be useful decision support tools as they allow an assessment of the relative changes in outcome probabilities, associated with changes in management actions or system parameters. By specifying the state for one or more input nodes, the impacts on other nodes can easily be predicted. In addition to prediction, BNs can be used for diagnostic analyses. By selecting a specific state of an output node, the probability that the input nodes need to be in a particular state can be observed (Daly et al., 2011).

Galan Caballero (2005) provides a simple flow chart (figure 4.7) for realizing a BN. The first step is to define the domain problem that specifies the purpose of the BN. This is followed by identifying the variables or nodes that are important for the domain problem. Next, the relationships among the variables or nodes are identified and represented in a graphical structure. The model structure obtained is then validated with the experts. If there is no agreement among the experts on the model structure, the process goes back to the previous step until an agreement is reached. The last three steps include eliciting expert opinion (also referred to as quantifying the network), running plausible scenarios with the network (network application), and fine tuning the estimates over time (maintaining the network). Building a Bayesian network is an iterative process and also a tradeoff between a very detailed and rich model to obtain "accurate" results on the one hand, and the cost and complexity of obtaining the detailed probabilities and maintaining the network afterwards on the other hand (Adusei-Poku, 2005).

*Figure 4.7: Flow chart for realizing a BN*


Determine domain problem (with experts)
Select candidate variables (with experts)
Build model with selected nodes
Fine tune estimates over time
No
Validate model with experts
Run plausible scenarios
Data elicitation with experts
Yes

## 4.6 Application of Bayesian Networks in Various Domains

A lot of the original applications of BN in risk management were in the medical field and to some extent, this is the domain where Bayesian network applications dominate today. However, there are now many uses in diverse domains, including biology, natural language processing and forecasting. Part of the popularity of Bayesian networks must stem from their visual appeal, as it makes them amenable to analysis and modification by experts. However, it is the generality of the formalism that makes them useful across a wide variety of circumstances. Since 2001, BNs have been used to analyze risky situations. Particularly, BNs represent a formalism use in the risk analyses domain due to their capacity to deal with probabilistic data and to model the dependencies between events (Daly *et al*, 2011).

This section aims to look at some typical applications of Bayesian networks across many different domains:

Bayesian networks have had considerable applications in many fields both in academia and industry. The major application area in both fields has been diagnosis, which lends itself very naturally to the modelling techniques of Bayesian networks. In the academic fields, it has been applied to problems in medical diagnosis, in heuristic search, in ecology, in data mining and in intelligent trouble shooting systems. (Ershi, Jiang Shen and Dou, 2013; Stephens, 2013).

Industrial application of Bayesian technology spans several fields including medical and mechanical diagnosis, risk and reliability assessment, and financial risk management. An example of medical diagnosis is the Heart Disease Program developed by the MIT laboratory for Computer Science and Artificial Intelligence. This program assists physicians in the task of deferential therapy in the domain of cardiovascular disorders. One mechanical diagnostic application is the computer trouble shooting SASCO project by University of Aalborg, Denmark and Hewlett Packard. This system is used in several of Hewlett Packard's printers (Steenbergen, Gelder, Miraglia, and Vrouwenvelder 2013; Acton, 2013).

In risk and reliability assessment, Philips Consumer Electronics uses BN technology to predict software defects in its consumer electronics (Fenton et al., 2001). Some examples in financial risk management include the credit risk prediction tool BayesCredit and the iRisk tool for operational risk prediction (Neil et al., 2005).

In the maritime field, some work aims at developing BN approaches to consider the human and organizational factors in a risk analysis. Norrington et al. (2007) describe the process of the experts‟ judgments to build a BN. A significant BN approach was developed by Trucco et al. (2008) which demonstrates the correlation between sources events of a collision accident.

*Medicine* - As noted previously, there are many applications of Bayesian networks in medicine, but some of the more famous applications are; an early implementation of a system for diagnosis in internal medicine was the quick medical reference (QMR). This system was reformulated in a Bayesian network implementation, with three levels of nodes; background, diseases and symptoms. Known as QMR-DT, it had a very large number of nodes and arcs, as a result, algorithms had to be developed that could perform inference in this dense network. Another more specific diagnostic system comes from the Pathfinder project, which is used in the diagnosis of lymph-node diseases. ALARM network which was used for the monitoring of patients in intensive care situations (Daly *et al*, 2011).

In forecasting, Bayesian networks can be very useful in predicting the future based on current knowledge. One of the most well-known of these is the HailFinder network which is used to forecast severe weather. Also in the weather forecasting domain is the sea breeze prediction system which uses learned structure and probability. In modelling for human understanding, the use the sparse candidate (SC) algorithm to learn the structure of 800 genes using 76 samples. These ideas have been built on by other researchers who look at the problem of small sample sizes prevalent with biological data and examine techniques to characterize the sensitivity and specificity of results (Daly et al., 2011; Scutaria and Nagarajanb, 2013*).*

## 4.7 Bayesian Networks and IS Risk Management

The risk assessment and management cycle and the process used to build a Bayesian network are highly complementary where the outcome of each part of the risk assessment cycle can be formalized within a Bayesian network. Where appropriate, risk management strategies, and the probability of their success, can be built and tested within the Bayesian network (Hart and Pollino 2009).

In the process of risk analysis for information systems, models are built in order to analyze and better understand the risk factors and their causal relationships in real-world information systems. Establishing an appropriate model suitable for the target risk problem is a crucial task that will ultimately influence the effectiveness of risk analysis results (Fenton and Neil, 2013). In the existing literature, most the approaches either assumed that the structure of the model was provided by domain expert experience and knowledge, or assumed that the structure was chosen from some general well-known class of model structures, thus, the results of risk analysis were relatively subjective. To overcome these drawbacks, not only expert have the experience and knowledge that needs to be taken into account, but also, the database of observed cases from information systems should be utilized in the process of modeling. When analyzing risk, communication of uncertainties is essential. Sources of uncertainty can include imperfect understanding or incomplete knowledge of the state of a system, randomness in the mechanisms governing the behavior of the system, or a combination of these factors. A Bayesian network is particularly useful as Bayesian inference provides a probability based approach that can update scientific knowledge when new information becomes available. Therefore, how to fuse the database of observed cases with domain expert experience and knowledge for inducing a representative model for observed information systems is a critical issue in risk analysis. Through structure learning and parameter learning, a Bayesian network (BN) can be developed to simultaneously define the risk factors and their causal relationships (Cavusoglu et al., 2009; Feng and Xie, 2011; Fenton and Neil, 2013).

## 4.8 Comparison between Bayesian Techniques and other Approaches

### 4.8.1 Bayesian Networks vs. Classical Probability Theory

In this section I will compare the Bayesian and classical view of probability on two of the important aspects of probability, namely the meaning of the probability and the meaning of conditional independence.

The Bayesian approach views probability as a person's degree of belief in an event x occurring given the information available to that person. A probability of 1 corresponds to the belief in the absolute truth of a proposition, a probability of 0 to the belief in the proposition's negation, and the intervening values to the partial belief or knowledge.

Classical probability theory considers the probability of an event x as the physical probability of the event x occurring. The probability values are acquired through a number of repeated experiments. The larger the number of experiments performed, the more accurate the value of the probability (Koski and Noble, 2012).

Thus, the classical approach relies on the existence of the experiments and is not willing to attach any probability value to an event that is not a member of a repeatable sequence of events. The Bayesian approach, on the other hand, consider a probability as a person's degree of belief, a belief can be assigned to unique events that are not members of any repeatable sequence of events. Although Bayesian approach is willing to assign a probability value to this event, the assignment of this subjective probability should be considered carefully. It must be based on all the information available to the individual who makes the prediction.

This information may include those items that are known to be true, deducible in a logical sense and empirical frequency information (Daly et al., 2011).

In regards to the meaning of conditional independence the classical probability checks the conditional independence through the equality of the joint probability of the events and the product of the individual events. The problem with this checking is that the result of the joint probability calculation does not provide psychological meaning to the user or developer of the knowledge-based system about the dependency between the events. Human cannot easily attach numerical values to an event but can easily determine whether two events are independent from looking at the cause-effect relationship between the events involved. The Bayesian approach, on the other hand, bases its conditional independence concept around the human reasoning process. Treating conditional independence using conditional probabilities rather than joint probabilities not only mirrors the human reasoning process but also provides the capability for knowledge based systems to use the recursive and incremental updating of the belief value. Bayes theorem provides us with greater ability to quantify the probability model of a situation by a method close to the human reasoning process (Pollino, and Henderson, 2010; Fenton and  Neil, 2013).

### 4.8.2 Bayesian Networks vs. Rule-based Systems

A Rule-based system consists of a library of rules of the form: if (assertion) then action. Such rules are used to elicit information or to take appropriate actions when specific knowledge becomes available. The main difference between BNs and rule based systems is that rule based systems model expert’s way of reasoning while BNs model dependencies in the domain. Rules reflect a way to reason about the relationships within the domain and because of their simplicity, they are mainly appropriate for deterministic problems, which is not usually the case in software engineering. Estimates are a probabilistic assessment of a future condition and that is the main reason why managers do not obtain good estimates (Koski and Noble, 2012).

Another difference is that the propagation of probabilities in BNs uses a global perspective in the sense that any node in a BN can receive evidences, which are propagated in both directions of the edges. In addition, simultaneous evidences do not affect the inference algorithm (Fenton and Neil, 2013).

### 4.8.3 Bayesian Networks vs. Neural Networks (NN)

Neural networks, can be used for classification and its architecture consists of an input, an output and possibly several hidden layers in between them; except for the output layer, nodes in a layer are connected to nodes in the succeeding layer. In software engineering, the input layer may be comprised of attributes such as lines of code, development time etc. and the output nodes could represent attributes such as effort and cost. NNs are trained with past project data adjusting weights connecting the layers, so that when a new project arrives, NNs can estimate the new project attributes according to previous patterns. (Fenton and Neil, 2013).

A difference is that NNs cannot handle uncertainty and offer a black-box view in the sense that they do not provide information about how the results are reached; however, all nodes in a BN and their probability tables provide information about the domain and can be interpreted. Another disadvantage of NN compared to BNs is that expert knowledge cannot be incorporated into a NN, i.e., BN can be constructed using expert knowledge, past data or a combination of both, while in NNs it is only possible through training with past project data (Koski and Noble, 2012).

## 4.9 Advantages/Strengths of Bayesian Networks

Rodríguez et al. (2009), notes that BNs have a number of features that make them suitable for dealing with problems in the software engineering field.

Graphical representation - BNs allow us to create and manipulate complex models to understand chains of events (causal relationships) in a graphical way that might never be realized using for example, parametric methods. Moreover, it is possible to include variables in a model that correspond to processes as well as product attributes (Fenton and Neil, 2013).

Uncertainty - Bayesian systems model probabilities rather than exact value meaning that uncertainty can be handled effectively and represented explicitly. Many areas in Software Engineering are driven by uncertainty and influenced by many factors. BN models can predict events based on partial or uncertain data, i.e., making good decisions with data that is scarce and incomplete (Dyhre and Jensen 2013).

Qualitative and quantitative modeling - BNs are composed of both a qualitative part in the form of a directed acyclic graph and a quantitative part in the form of a set of conditional probability distributions. Therefore, BNs are able to utilize both subjective judgments elicited from domain experts and objective data (e.g. past project data) (Dyhre and Jensen 2013).

Bi-directional inference - Bayesian analysis can be used for both forward and backward inference, i.e. inputs can be used to predict outputs and outputs can be used to estimate input requirements. For example, we can predict the number residual defects of the final product based on the information about testing effort, complexity. Furthermore, given an approximate value of residual defects the BN will provide us with a combination of allowable values for the complexity, testing effort etc. which could satisfy the no. of residual defects (Larrañaga et al, 2013).

Confidence Values - The output of BNs are probability distributions for each variable instead of a single value, that is, they associate a probability with each prediction. This can be used as a measure of confidence in the result, which is essential if the model is going to be used for decision support. For example, if the confidence of a prediction is below certain threshold the output could be ‘not known’ (Dyhre and Jensen 2013).

Decision support tools - BNs can facilitate learning about causal relationships between variables and can easily be converted into decision support tools to aid risk management. The graphical nature of a BN clearly displays the links between different system components and predict the effect of taking the various alternative courses of action (Koski and Noble, 2012).

A convenient feature of BNs is the ability to learn about the structure and parameters of a system based on observed data. Knowledge of the structure of a system can reveal the dependence and independence of variables and suggest a direction of causation. It evaluates the ‘optimal’ BN structure, based on the highest probability score for possible candidate structures, given the data provided and perhaps penalized for the level of complexity. Different score metrics can be used to evaluate the BN structure, varying from entropy methods to genetic algorithms. If there are lots of missing observations, BNs can use complex learning algorithms to learn the tables (Koski and Noble, 2012; Fenton and Neil, 2013).

BNs allow an injection of scientific rigor when the probability distributions associated with individual nodes are simply "expert opinions". This can both increase the reliability of the expert opinions, while also making explicit the imprecision that is inherent in such judgments (Larrañaga et al, 2013).

### 4.10 Limitations of Bayesian Networks

There are also some clear limitations to BN models. While Bayesian models are a useful way to model expert knowledge, it may be difficult to get experts to agree on the structure of the model and the nodes that are important to be included. Furthermore, experts may be challenged to express their knowledge in the form of probability distributions. Elicitation of expert knowledge requires an iterative process, to ensure that experts are comfortable with the nodes, their states and interrelationship in the BN, before they can make statements about distributions and confidence intervals of variables (Fenton and  Neil, 2013).

Some BN software packages may have limited ability to deal with continuous data. Such data generally needs to be discretized. Although discretizing is a convenient way to control the size of the network, discrete states may not capture the original distribution of the variable completely and can lead to lower precision of variable values. Barton et al (2008) show how discretization assumptions can significantly affect the outcome estimates.

Another limitation that has been defined in the literature stems from the acyclic nature of BNs. The acyclic property is required to carry out probability calculus, but implies that feedback effects cannot be included in the network (Barton et al, 2008).

Another limitation centers on the extent of the quality of the prior beliefs used in Bayesian inference processing. The usefulness of a BN is based on the reliability of its prior knowledge. An excessively optimistic or pessimistic expectation of the quality of these prior beliefs will either distort the entire network or invalidate the results. Related to this concern is the selection of the statistical distributions induced in modeling the data. Selecting the proper distribution model to describe the data has a notable effect on the quality of the resulting network (Koski and Noble, 2012; (Fenton and  Neil, 2013).

### 4.11 Summary

Bayesian analysis is a well-defined and rigorous process of inductive reasoning that has been used in many scientific disciplines. A distinctive feature of the Bayesian approach is that it permits the investigator to use both sample (data) and prior (expert-judgment) information in a logically consistent manner in making inferences by using Bayes' theorem to produce a 'post data' or posterior distribution for the model parameters. Using Bayes' theorem, prior (or initial) values are transformed to post-data views. This transformation can be viewed as a learning process. The posterior distribution is determined by the variances of the prior and sample information. If the variance of the prior information is smaller than the variance of the sampling information, then a higher weight is assigned to the prior information. On the other hand, if the variance of the sample information is smaller than the variance of the prior information, then a higher weight is assigned to the sample information causing the posterior estimate to be closer to the sample information. The Bayesian approach provides a formal process by which a-priori expert judgment can be combined with sampling information (data) to produce a robust a posteriori model.

Bayesian analysis has all the advantages of "Standard" regression and it includes prior knowledge of experts. It attempts to reduce the risks associated with imperfect data gathering. Software engineering data is usually scarce and incomplete and estimators are faced with the challenge of making good decisions using this data. Classical statistical techniques derive conclusions based on the available data. But, to make the best decision it is imperative that in addition to the available sample data we should incorporate non-sample or prior information that is relevant. Usually a lot of good expert judgment based information on software processes and the impact of several parameters on effort, cost, schedule, quality etc. is available. This information doesn't necessarily get derived from statistical investigation and hence classical statistical techniques do not incorporate it into the decision making process. Bayesian techniques make best use of relevant prior information along with collected sample data in the decision making process to develop a stronger model.

# CHAPTER FIVE

# 5.0 RESEARCH METHODOLOGY

## 5.1 Introduction

The preceding chapters have discussed in-depth the research context that is financial information system risk management using Bayesian networks. Based on the literature review findings and information elicited, research questions were posed upon which this research rests. This chapter is organised around the following areas: research design, research methodology adopted for this study to satisfy the research objectives, operationalizing and bringing the survey instruments into context and data analysis.

The researcher has used qualitative research approach. The qualitative research paradigm, also referred to as “constructivist”, “naturalistic”, “interpretative”, “post-positivist” or “post-modern perspective” approach, is an enquiry process of comprehending a social or human problem/phenomenon based on building a complex holistic picture formed with words, reporting detailed views of informants and conducted in a natural setting (Cooper and Schindler, 2011).

Creswell and Clark (2011), indicated that qualitative research is multi method in focus, involving an interpretive, naturalistic approach to its subject matter. This means that qualitative researchers study things in their natural settings, attempting to make sense of, or interpret, phenomena in terms of the meanings people bring to them. Qualitative approach is one in which the inquirer often makes knowledge claims based primarily on constructivist perspectives (i.e., the multiple meaning of individual experiences, meaning socially and historically constructed, with an intent of developing a theory or pattern) or advocacy/participatory perspectives (i.e., political, issue-oriented, collaborative or change oriented) or both. It also uses strategies of inquiry such as narratives, phenomenology, ethnography, grounded theory studies or case studies (Creswell and Clark, 2011).

## 5.2 Research Design

Creswell and Clark (2011) state that research designs are procedures for collecting, analysing, interpreting and reporting data in research studies. Rigorous research designs are important because they guide the methods and decisions that researchers must make during the study and set the logic by which interpretations are made at the end of the study.

The exploratory research design used has outlined the situation in respect to the variable being investigated. This means of research design makes it possible for data to be collected effectively without any manipulation on the research context. The research design seeks to outlay the goals of the research by stipulating practical issues that are of focus to this study (Saunders, Lewis and Thornhill 2012).

Research can be classified in terms of their purpose. Accordingly, they are most often classified as exploratory, descriptive or explanatory. The researcher has opted to use exploratory research. Exploratory research is used to develop a better understanding. Exploratory studies are a valuable means of finding out what is happening, to seek new insight, to ask questions and to assess phenomena in a new light. It is particularly useful if researcher wish to clarify the understanding of a problem (Saunders, et al., 2012).

## 5.3 Population of the Study

The target population consisted of employees of various financial institutions, with background knowledge of FIS and risk management by virtue of their positions in their organizations be it managerial or administrative. They include directors, manager, unit and departmental heads.

## 5.4 Sample and Sampling Frame

To facilitate data collection, the study's sampling frame constituted a listing of institutions from various sectors which include: banks, SACCOs, micro finance institutions and housing finance. A total of 40 respondents from various financial institutions.

An adequate sample size should allow reliability of results so that the investigation can be repeated with consistent results. A sample is a small set of data drawn from a population as Emmel (2013) noted and that the sample should be sufficiently and demonstrably representative of the population in order to allow analysis of the sample to be used. The sample size affects confidence interval, thus could, in principle, select the sample to yield any degree of confidence (Saunders, Lewis and Thornhill, 2012). For this study, a stratified purposive sampling technique was adopted for data collection from the sampled institutions and key informants; since financial institutions are discrete and in an effort to maintain confidentiality of the respondents the respondents holding positions of interest were identified. This is normally done by dividing the population into different strata on the basis of some common characteristics. To determine the sample size, the researcher used the formula:

n = N*X / (X + N - 1), where, X = Zα/22 ¬*p*(1-p) / MOE2, and Zα/2 is the critical value of the Normal distribution at α/2 (for a confidence level of 95%, α is 0.05 and the critical value is 1.96), MOE is the margin of error, p is the sample proportion, and N is the population size (Gay and Peter 2012). From this formula the minimum sample size is 39, however the researcher chose a total of 40 respondents

### 5.5 Data Collection

Because surveys make it possible to study a population too large to observe directly, it presents an excellent mechanism to collect original data. According to King (2012), the careful selection of a probability sample will provide a group of respondents whose characteristics could mirror those of the larger population. The data gathered by studying the characteristics of the sample can then be generalised to the larger population. This data is then gathered by administering a questionnaire, otherwise known as a structured scheduled interview. A questionnaire is the complete data collection instrument used by and interviewer or respondent (or both) during a survey (King, 2012; Johanson, 2013).

Primary data was used for this study and the data was collected using questionnaires that were hand delivered and also sent by e-mail. A questionnaire was prepared to understand the perspective of various financial IS stakeholders on risk management including the use of Bayesian networks. The questionnaire was designed as per the objectives of the study.

Secondary data was also used. Information was obtained from various journals, publications, websites and reports. Secondary sources helped the researcher in explaining different conclusions based on previous studies that have been conducted and concluded, while the primary data sources was information collected by the researcher herself specifically for the study.

### 5.6 Validity and Reliability of the Research Instrument

For quality control, a pre-test of the research instrument (questionnaire) was done to establish its validity. The questionnaire was given to individuals (who constitute the population of key informants) to give their opinion on the relevance of the questions using a 4-point scale of relevant, quite relevant, somewhat relevant, and not relevant. In this study the researcher is confident that the theoretical understanding of risk management of information systems is the same as in the operational sense and to that extent, there is clear connection between the theoretical and practical notion of risk management of information systems and for that reason the researcher has valid data. Additionally, numerous steps were taken to ensure the validity of the study:

- Data was collected by in-depth questionnaires from the reliable sources with knowledge of financial information systems risk management
- Questions in the questionnaire were made based on literature review and frame of reference to ensure the validity of the result.
- Data was collected within 4 weeks, within this short period of time no major event has been changed with the related topic.

According to Saunders et al. (2012), reliability refers to the degree to which data collection method or methods will yield consistent findings, similar observations would be made or conclusions reached by other researchers or there is transparency in how sense was made from the raw data. Reliability can be assessed by the following three questions (Easterby-

Smith, Thorpe and Jackson (2012), Numbers of different steps were taken to ensure the reliability of the study:

- The same type of questions was used for all the respondents in order to increase the reliability.
- The theories that have been selected for the study were clearly described and research question has been formulated based on the previous theory.
- Data has been collected based on the frame of reference that was drawn from the discussed theories. The objective is to make sure that if another investigator will follow the same procedures, the same conclusions would be made.

### 5.7 Data Processing and Analysis

After collecting all the data the process of analysis begins. The data was analysed to provide an over view of respondents perception of the various aspects of the research objectives. To summarize and rearrange the data, several interrelated procedure are performed during the data analysis stage (Zikmund, et al., 2013). The data was analysed through descriptive statistics. Graphs, pie charts and distribution tables have been used where appropriate to present the research findings so as to ensure that the research is clear and easily understandable. Statistical Package for Social Sciences (SPSS) version 22 was used for data analysis.

### 5.8 Ethical Considerations

The goal of ethics in research is to ensure that no one is harmed or suffers adverse consequences from the research activities (Cooper and Schindler, 2011). The researcher has undertaken various measures to protect the rights of the respondents by:

- Ensuring that none of the respondents was named during the research or subsequent report
- Respondents were selected to participate without compulsion
- All respondents were informed of the reason and purpose of the research; and

- Informed consent was sought from the management of the selected company and the respondents before the commencement of this research initiative.

# CHAPTER SIX

# 6.0 FINDINGS AND DISCUSSIONS

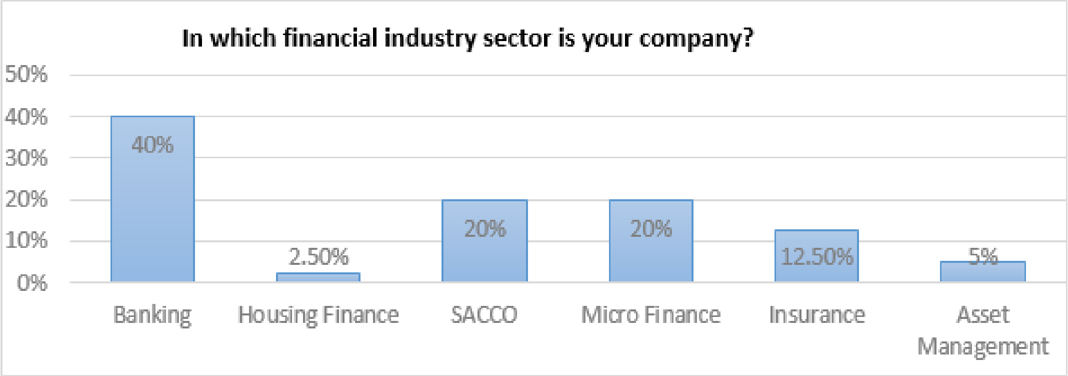
## 6.1 Introduction

This section is devoted to the analysis of primary data collected from the survey using questionnaires that were formulated based on the knowledge gathered from secondary sources to help attain the objectives of the research. Descriptive statistics was adopted to carry out preliminary data analysis and to describe the features of the data; summarise the samples and the measures, and jointly with diagrammatic and graphic analysis, provide a basis of quantitative analysis of the data.

## 6.2 Demographic Information

The survey setting was Financial institutions composed of a listing of various recognized and government approved institutions that offered financial services. The institutions are indicated in Figure 6.1

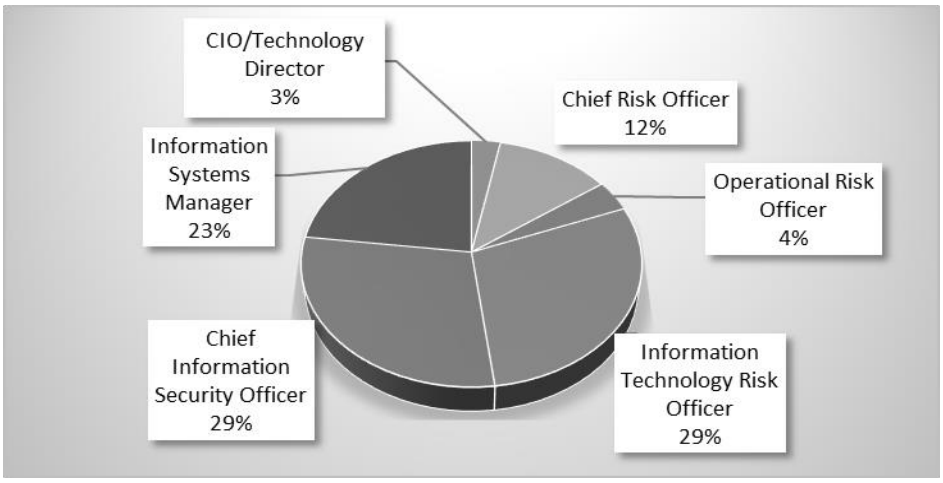
*Figure 6.1: In which financial industry sector is your company?*



These institutions are based within and around Nairobi County, most of them have branches in various towns and cities within the country with others across borders. Majority of the respondents were from banks (40%) followed by SACCOs and micro-finance institutions each with 20%.

The survey sample was selected on the basis of their organizations and organizational position and risk management roles in the financial institutions including but not limited to being on risk management committees. As such the respondents were better placed to have knowledge and access to accurate information that would be valuable in this research. Their distributions are indicated in Figure 6.2.

*Figure 6.2: Which of the following best describes your title?*



The purposeful sampling and selection of key informant was carefully done to ensure that all key financial information systems stakeholders within the institutions were adequately presented. The selected respondents had experience and access to the information required to conduct this study successfully.

The researcher sought to find out how financial institutions acquired their Financial Information Systems, the research findings are indicated in table 6.1

*Table 6.1: Information system acquisition*

| | Frequency | Percent |
|---|---|---|
| Tailor makes the system (internally or outsources) | 24 | 60.0 |
| Acquires already developed systems | 16 | 40.0 |
| Total | 40 | 100.0 |

Source: Research findings; 2014

As shown in table 6.1 above, 60% of financial institutions in Kenya develop their own tailor made financial information Systems; this is done by their employees or outsourced developers who are in the direct control of the financial institutions as opposed to 40% of the institutions who acquire already developed information systems. These institutions however tweak these systems to suit the standard operating procedures.

Organization's involvement in Information systems development is vital not only to ensure successful implementation but also to manage risks. The respondents for organizations that tailor made their Financial Information systems were asked whether they were actively involved in every stage of their financial information system development. The response is indicated in table 6.2.

*Table 6.2: Organization involvement in information system development*

| | Frequency | Percent |
|---|---|---|
| Yes | 14 | 58.33 |
| No | 7 | 29.17 |
| Don't know | 3 | 12.50 |
| Total | 24 | 100.0 |

Source: Research findings; 2014

The survey indicated that more than half (58.33%) of these organizations were actively involved in their information systems development, a significant 29.17% of the organizations were not actively involved while 12.5% of the respondents had no idea about their organizations involvement as shown in figure 6.2.

Organizations that are not actively involved in their information systems development have a higher chance of being exposed to not only more risks that could have been well managed in the initial stages, but also incur more cost in mitigating the said risks.

## 6.3 Current State of FISRM in Financial Institutions

The researcher sought to find out whether FIs have formal financial information system risk framework and assessment process, table 6.3 shows the response.

*Table 6.3: Presence formal FIS risk framework and assessment*

| | Frequency | Percent |
|---|---|---|
| Yes | 34 | 85.0 |
| No | 4 | 10.0 |
| Don’t Know | 2 | 5.0 |
| Total | 40 | 100.0 |

Source: Research findings; 2014

The survey results indicate that majority of the financial institutions 85% there existed a formal financial information system risk framework and assessment process while a minority 10% did not.

In regards to the existence of a common risk language that is broadly used and understood throughout and across my sector the research findings are presented in table 6.4.

*Table 6.4: Existence of common risk language that is broadly used and understood*

| | Frequency | Percent |
|---|---|---|
| Strongly agree | 20 | 25.0 |
| Agree | 14 | 25.0 |
| Neither agree nor disagree | 4 | 11.0 |
| Disagree | 2 | 41.0 |
| Total | 40 | 100.0 |

Source: Research findings; 2014

According to the survey responses (table 6.3), 25% of the respondents said there is a common risk language that is broadly used, understood and accepted. However, 11% of the respondents were not sure if one existed while 41% said there was no common risk language that was broadly accepted and understood throughout their sector.

Reporting and metrics undoubtedly have a great impact in demonstrating the value that an information system risk management program brings to an organization. However, reporting functionality is largely reliant on tools and technology. An organization’s process must be accurately mapped and effectively designed if the tools are to have a positive impact. Tools are not a solution in and of themselves; rather, they simply optimize an operational process and make it more efficient. The researcher sought to assess this, the responses are indicated in tables 6.4 and 6.5

*Table 6.5: IS Managers receive sufficiently regular and robust information on risk*

| | Frequency | Percent |
|---|---|---|
| Strongly agree | 22 | 55.0 |
| Agree | 16 | 40.0 |
| Disagree | 2 | 5.0 |
| Total | 40 | 100.0 |

Source: Research findings; 2014

Table 6.4 shows the survey response where 95% of the respondents agree and strongly agree that have sufficient information to assess whether the information provided on risk is sufficient.

The response to the question on the existence of defined financial information system risk performance framework and metrics that are monitored and reported to management on a regular basis are shown in figure 6.6.

*Table 6.6: Existence of defined FIS risk performance framework and metrics*

| | Frequency | Percent |
|---|---|---|
| Yes | 29 | 72.5 |
| No | 8 | 20.0 |
| Don’t Know | 3 | 7.5 |
| Total | 40 | 100.0 |

Source: Research findings; 2014

Majority of the respondents 72% said their organizations had defined financial information system risk performance framework and metrics that are monitored and reported to management on a regular basis while 20% did not, also 3% of the respondents were not sure whether or not their organizations had such a framework and metrics as shown in table 6.5.

As a program becomes more effective and efficient, it should not require substantially increased investment every year, except in the event of significant business or regulatory changes. Ideally, spending should level off and may even decrease as a program gains maturity and is optimized. This increase in investment however is an indication that most programs are only in the early stages of maturity since investments typically increase at higher rates when a program is in development. The researcher wanted to determine an estimated projection on financial investment over the next two years in FISRM activities, the findings are indicated in table 6.7.

*Table 6.7: Financial investment in FIS risk management activities*

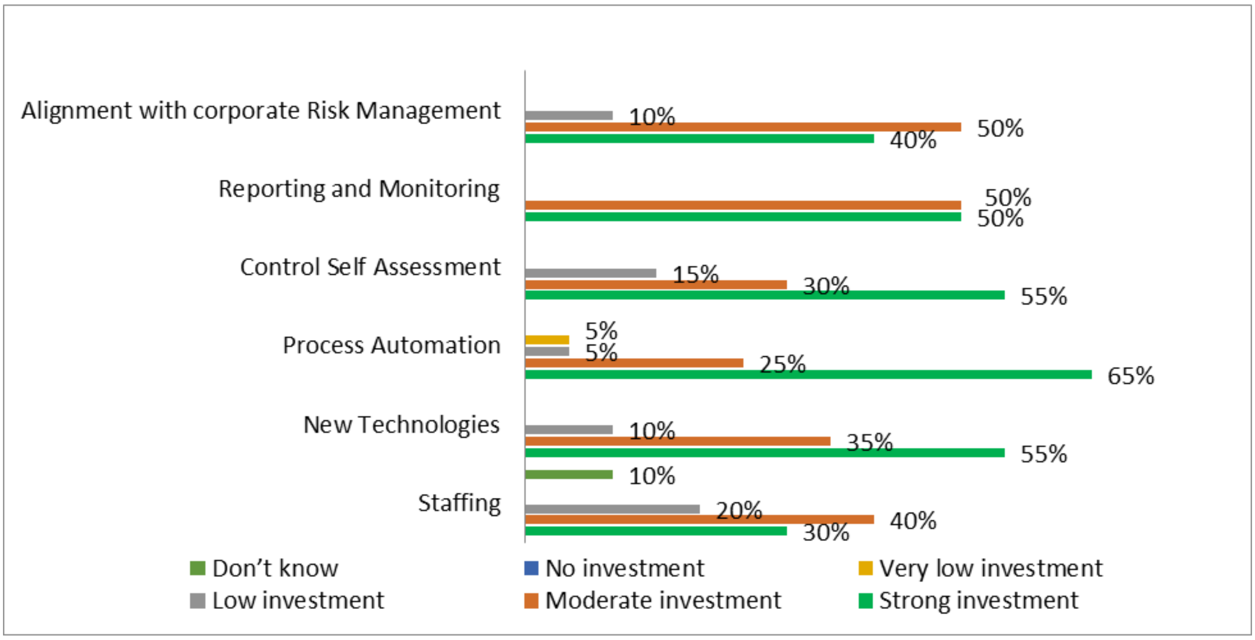
| | Frequency | Percent |
|---|---|---|
| Increase by 5-25% | 20 | 50.0 |
| Increase by more than 25% | 20 | 50.0 |
| Total | 40 | 100.0 |

Source: Research findings; 2014

The survey study found out that 50% of respondents felt their organizations would increase spending over the next 2 years, while the other half were of the opinion that investment would increase by between 5-25% as shown in Table 6.6.

The researcher went ahead to determine how these investments in risk management initiatives will be allocated, the distributions are presented in Figure 6.3.

*Figure 6.3: Financial allocation in FIS risk management activities*



The survey indicated as shown in Figure 6.3 that process automation would receive strong investment according to 65% of the respondents. Following closely was new technologies and control self-assessment. Additionally, the initiative that would receive the lowest investment the company's information system investment in risk management is concerned is staffing. The high investment in new technologies and process automation as tools and technology initiatives designed to optimize the existing processes shows us that there is a strong interest in investing in tools and technology for process automation and optimization.

The researcher wanted to determine generally if the respondents felt that the current information system risk management framework in financial institutions is sufficient. The descriptive statistics tables shows the response as analysed and presented in Table 6.8

*Table 6.8: Sufficiency of current information system risk management framework*

| | Mean | Std. Deviation |
|---|---|---|
| strongly agree | 2.07 | 1.871 |
| agree | 3.41 | 0.410 |
| Neither agree nor disagree | 4.14 | 0.851 |
| disagree | 4.42 | 0.746 |
| strongly disagree | 3.70 | 0.718 |

Source: Research findings; 2014

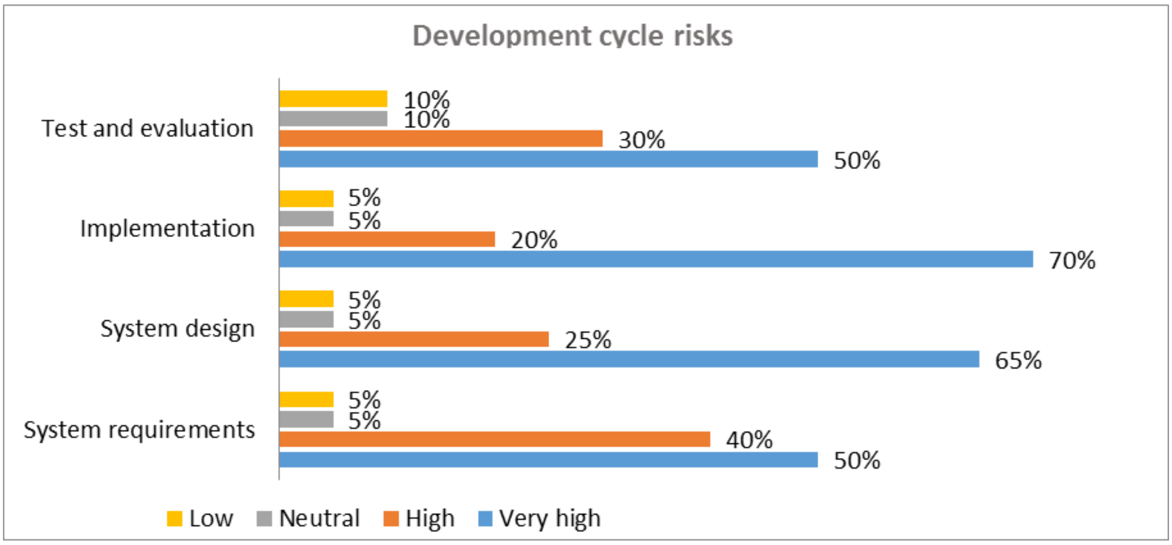
As indicated in table Majority of the respondents as indicated by a mean of 4.42 were of the opinion that current information system risk management framework in financial institutions is not sufficient, a higher number of them were neutral about its sufficiency with minority of them (mean=2.07) strongly agreeing that the current risk management program is sufficient.

### 6.4 Analysis of FIS Risks and RM Programs Used By Financial Institutions

To better manage financial information system development risk, it's not only important to identify and understand the risk innate in every stage but also the level of risk and areas that are prone and or vulnerable.

### i. Development cycle risks

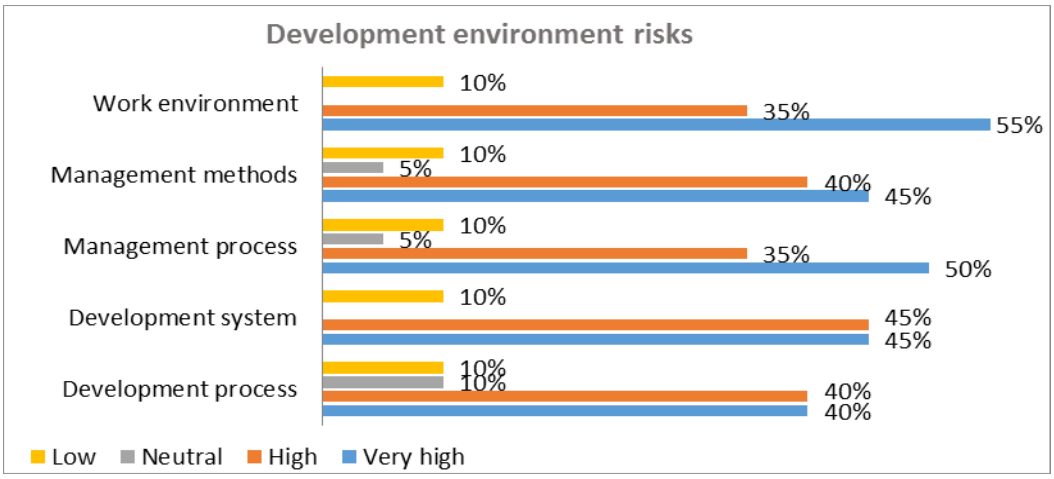
*Figure 6.4: Analysis of Development cycle risks*



Majority of the respondents as indicated in figure 6.4, 70% indicated that the implementation stage of the IS development cycle had the highest risk level followed by system design (65%); test and evaluation and system requirements (both at 50%). It is worth noting that all the stages of the development according to at least 50% of the respondents have very high risk. This is significant to this research as indicates that this stages pose significant risks that need to be managed early to avoid compounding and minimize the cost that would be involved to manage them at a later stage.

### ii. Development environment risks

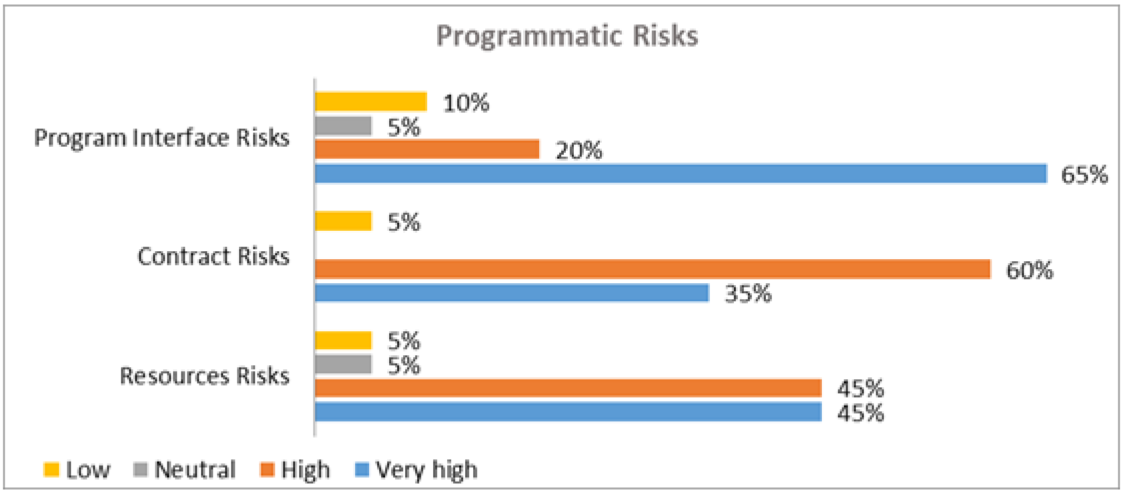
*Figure 6.5: Analysis of Development environment risks*



Work environment was considered by 55% of the respondents to be posing the highest risk as far as development environment risks were concerned followed closely by management process 50%; Figure 6.5.

### iii. Programmatic Risks

*Figure 6.6: Analysis of Programmatic risks*



Program interface risks and contract risks were considered to pose high risk levels by most respondents (65 and 60%) respondents, resources risks also were also considered to pose significant risk levels as indicated in Figure 6.6 above.

The findings above from the analysis of financial information system development risks in this section is evidence that the entire SDLC is prone to numerous significant risks most of which if not managed will have a ripple effect after implementation and most probably require more resources to manage. This echoes the need to manage this risks from the initial phases of the SDLC.

The researcher sought to find out the Standards/leading practices used in developing financial information system risk framework and assessment processes the findings are indicated in Table 6.9.

*Table 6.9: standards used in developing FIS framework and assessment processes*

| | **Frequency** | **Percent** |
|---|---|---|
| COSO ERM - Committee of Sponsoring Organizations of the Treadway | 2 | 5.0 |
| British Standard BS-6079-3:2000 | 9 | 22.5 |
| COBRA - Consultative, Objective, and Bi-functional Risk Analysis | 4 | 10.0 |
| IRAM - Information Risk Analysis Methodologies | 12 | 30.0 |
| CRAMM - CCTA Risk Analysis and Management Method | 4 | 10.0 |
| OCTAVE - Operationally Critical Threat, Asset, and Vulnerability Evaluation | 6 | 15.0 |
| Other | 3 | 7.5 |
| Total | 40 | 100.0 |

Source: Research findings; 2014

According to the survey, financial institutions use a plethora of Standards and leading practices with no preferred approach across the sector; with some institutions use more than one. IRAM was used by 30% of the institutions and the British Standard BS-6079-3:2000 by 22.5% of them as indicated in table 6.9. This findings indicate that financial institutions use a variety standards used in developing FIS framework and assessment processes. This provides a challenge when sharing information within the sector, personalization and formalization.

The researcher sought to find out what main limitation were facing the current financial information risk management program in financial institutions. The responses were analysed using means and standard deviations as shown in Table 6.10.

*Table 6.10: Limitations/Challenges of current FISRM program*

| | **Mean** | **Std. Deviation** |
|---|---|---|
| Limitation on sharing past and present knowledge/information across the financial sector | 3.9412 | 1.15316 |
| Inability to utilize all the available data/information and knowledge gathered from system risk framework and assessment process | 4.1765 | 1.11384 |
| Conformity/integration with regulations, international management standards guidelines and practices | 3.7941 | 1.27397 |
| Inflexibility in the adoption of the dynamic and rapid rate of change in technology and the overlap between old and new technologies | 4.0882 | 1.11104 |
| Inability to be customized to suit specific organizations and yet meet international standards | 3.7353 | .99419 |

Source: Research findings; 2014

From the study findings the respondents indicated that inability of the current program to utilize all the available data/information and knowledge gathered from system risk framework and assessment process was the main challenge with a mean of 4.1765. The respondents further indicated that inflexibility in the adoption to the dynamic and rapid rate of change in technology and the overlap between old and new technologies was another significant limitation with a mean of 4.0882. Also, limitation on sharing past and present knowledge/information across the financial sector and conformity/integration with regulations, international management standards guidelines and practices posed significant challenges with means of 3.9412 and 3.7941 respectively.

The Means and Standard deviations of the responses in regards to desired improvements to their current FISRM program to ensure efficiency and effectiveness were determined as shown in

Table 6.11, the means were interpreted according to the Five point Likert scale as rated by the respondents.

*Table 6.11: Desirable improvements to the current FISRM program*

| | **Mean** | **Std. Deviation** |
|---|---|---|
| Ability to make predictions/forecasts on risks, alternative causes of action, implication risk management activities on enterprise prior to implementation | 4.3824 | .65202 |
| Ability to learn and integrate past and present knowledge and information | 4.2647 | .61835 |
| A holistic/enterprise wide approach to FISRM | 4.0882 | .83003 |
| It should be able to indicate the relationships between risk, risk factors and from an organizational/holistic perspective both direct and casual relationships | 3.9412 | .91920 |
| Scalability with the growth/development of the organization, the sector and technological change | 3.8824 | 1.03762 |

From the survey findings, ability to make predictions/forecasts on risks, alternative causes of action, and implication risk management activities on enterprise prior to implementation with a mean of 4.3824 was considered as the most desired improvement. This was followed by the ability to learn and integrate past and present knowledge and with a mean of 4.2647; a holistic/enterprise wide approach to FISRM with a mean of 4.0882. Other significant modifications included the ability to indicate the relationships between risk, risk factors and from an organizational/holistic perspective both direct and casual relationships; and scalability with the growth/development of the organization, the sector and technological change.

## 4.12 Analysis of the Application of Bayesian Networks in Financial Information System Risk Management

The study sought to find out whether the management of financial information risks throughout the SDCL would be better as compared to majority of the current approach that start risk management after implementation. The findings are shown in table 6.12.

*Table 6.12: FISRM would be better if done throughout the development phases*

| | Frequency | Percent |
|---|---|---|
| Strongly agree | 26 | 65.0 |
| Agree | 14 | 35.0 |
| Total | 40 | 100.0 |

Source: Research findings; 2014

The respondents overwhelmingly agreed that most of the risks faced by financial information systems would be better managed if they were handled during the development phases of the system as indicated in Table 6.12.

The researcher then sought to find out if the respondents were aware of Bayesian Networks and their application, the findings are presented in Table 6.13.

*Table 6.13: Knowledge of BN and its application*

| | Frequency | Percent |
|---|---|---|
| Yes | 33 | 82.5 |
| No | 7 | 17.5 |
| Total | 40 | 100.0 |

Source: Research findings; 2014

Majority of the respondents 82.5% (n=33) had knowledge on Bayesian networks and their applications as indicated in Table 6.13.

The researcher then asked the respondents who had knowledge on BN whether based on their knowledge and experience they thought that incorporating BN into their FISRM would significantly improve its efficiency and effectiveness. The response is indicated in Table 6.14.

*Table 6.14: Knowledge of BN and its application*

| | Frequency | Percent |
|---|---|---|
| Yes | 26 | 78.8 |
| No | 3 | 9.1 |
| Don’t know | 4 | 12.1 |
| Total | 33 | 100.0 |

Source: Research findings; 2014

Out of the respondents that had knowledge on Bayesian networks, majority (78.8%) were of the opinion than the incorporation of BN into their FISRM would significantly improve its efficiency and effectiveness.

# CHAPTER SEVEN

# 7.0 SUMMARY, CONCLUSION AND RECOMMENDATIONS

## 7.1 Introduction

This chapter highlights the main areas of the research including summary of the research findings, conclusions, recommendations and suggestions for further research.

## 7.2 Summary of the Research Findings

### 7.2.1 Demographic Information

The survey sample was selected on the basis of their organizations and organizational position that better placed to have knowledge and access to accurate information that would be valuable in this research. They included IT risk officers, chief IS officers, IS systems managers, chief risk officers and technology directors. This sample was picked from financial institutions including banks, Microfinance, SACCOs, insurance companies, asset management firms and housing finance.

The research indicated that majority of financial institutions in Kenya develop their own tailor made financial information systems through their employees or third parties. Of these institutions, 85.3% were actively involved in every stage of their financial information system development.

Majority of financial institutions acquired developed their own systems internally of sourced third parties to develop for them. Those that purchased already developed system, the researcher found out that they highly personalized them. Majority of the respondents from institutions that developed their system were involved in the development of these systems.

### 7.2.2 The Current State of FISRM in Financial Institutions

Majority of the institutions surveyed have a formal financial information system risk framework and assessment process in place. In regards to the existence of a common risk language that is broadly used and understood throughout and across the financial sector, the response was split in half with only 50% of the respondents indicating existence.

Reporting and metrics are valuable tools in any risk management program to this end majority of the financial institutions (95%) indicated that IS Managers receive sufficiently regular and robust information on risk from RM teams. Similarly there exists a defined financial information system risk performance framework and metrics that are monitored and reported to management.

Financial investment in a RM program is an indicator of its maturity level (efficiency and effectiveness); all the respondents surveyed indicated there was going to be an increase in financial investment in FIS risk management activities in the next 2 years. Half of the institutions will increase their investments by more than 25% and the rest by between 5-25%. This is a general indication that most of the RM programs are under development are those that are at an advanced stage are at an early maturity stage.

In regards to organizations' investment allocation in the risk management program, the survey showed that process automation would receive highest investment. This were followed by technologies and control self-assessment with the lowest allocation being on staffing.

Overall, the current information system risk management framework in financial institutions is insufficient according to majority of the respondents (mean of 4.42); minority of them however with a mean of 2.07 and 3.14 strongly agreed and agreed that the current RM program was sufficient.

### 7.2.3 Analysis of FIS Risks and Risk Management Programs Used By Financial Institutions

The researcher seeks and proposes to manage risks from the initial stages of the SDLC, to do these risks levels in the various environments were assessed. In the system development cycle risks, system implementation and system design posed the highest risk. In the system development environment risks, work environment and the management process posed the highest risks. Under programmatic environment risks program interface risks and contract risks were considered to pose high risk levels. This analysis indicates that every stage and environment of the SDLC pose significant risk; these should be managed in time to avoid the possibility of them compounding to the entire project and leading to the need for more resources to manage them later.

The study found out that financial institutions used varied standards/leading practices used in developing financial information system risk framework and assessment processes. IRAM – (Information Risk Analysis Methodologies), British Standard BS-6079-3:2000, OCTAVE - Operationally Critical Threat, Asset, and Vulnerability Evaluation being used by more than 14% of the institutions.

The main challenge faced by financial institutions when using the current financial information risk management program according to the study is the inability to utilize all the available data/information and knowledge gathered from system risk framework and assessment process. This was closely followed by the inflexibility in the adoption of the dynamic and rapid rate of change in technology and the overlap between old and new technologies; then the limitation on sharing past and present knowledge/information across the financial sector

The main desired improvements to the current risk management program as indicated by the respondents from the survey was the ability to make predictions/forecasts on risks, alternative causes of action, implication risk management activities on enterprise prior to implementation. This could be made even better if the predictions were assigned weights. The other main improvement indicated was the ability to learn and integrate past and present knowledge and information and a holistic/enterprise wide approach to FISRM.

### 7.2.4 Analysis of the Application of Bayesian Networks in FISRM

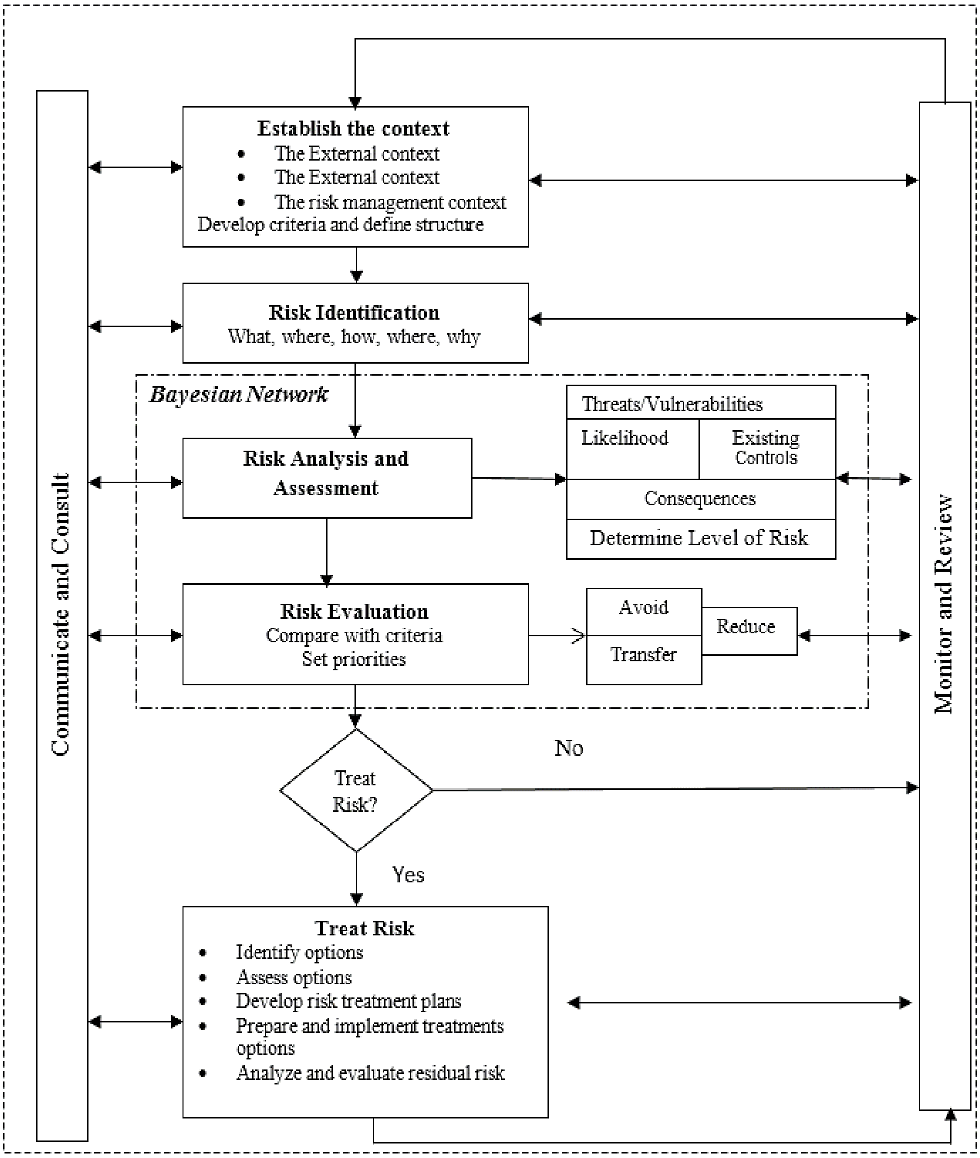
The researcher sought to find out whether the respondents agree that most of the risks faced by financial information systems would be better managed if they were handled during the development phases of the system. The respondents overwhelmingly agreed that this approach would significantly improve FISRM program.

Majority of the respondents based on their knowledge and experience felt that incorporating BN with artificial intelligence capabilities will significantly improve in FISRM program.

## 7.3 The Proposed Generic FRISM Framework

The proposed generic framework is based on recognized standards that will can be tailor made to suit specific financial institutions' needs. The framework proposes that FISRM starts from the first stage of SDLC, additionally Bayesian Networks be incorporated appropriately into the risk management programme. Figure 7.1 illustrates how the proposed financial information system risk management program will take place throughout the SDLC. Bayesian network will be used in risk analysis and evaluation and risk assessment.

*Figure 7.1: Generic risk management framework*



In this framework the researcher proposes the incorporation an AI based BN approach in the dotted area of the generic risk management framework in risk analysis and assessment as indicated in Figure 7.1. As such BN will be used in risk analysis, assessment and evaluation including decision making on risk treatment. Also the researcher proposes that FISRM be initiated from the initial phase of the SDLC.

After risk identification, BN will be used for risk analysis and assessment, here the likelihood of the risks will be modeled against existing controls/mitigation and consequences of risk events; using these factors the level of risk is then determined. The input for this phase will come from software development risk taxonomy.

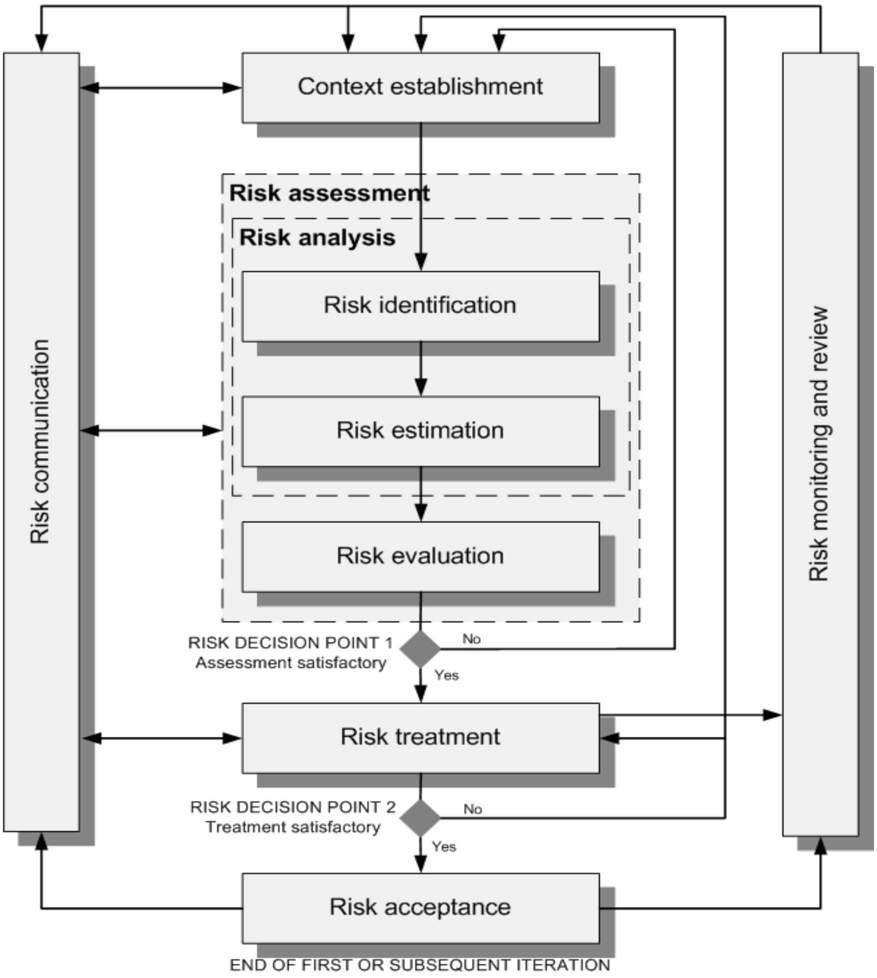
Risk evaluation will then be done by comparing with criteria set priorities; that is comparing the level of risk found during the analysis process with previously established risk criteria against which risks are compared in risk evaluation. Risk evaluation has to consider the big picture including the stakeholder’s objectives and risk tolerability, the degree of control over each risk, cost-benefits analysis and potential opportunities. Based on this evaluation, the organization will determine how to mitigate the risk by either; treating, avoiding, reducing or transferring risks. In the event that the risk is treatable, a criteria for risk treatment initially determined by the organization is followed. Management may also fail to respond by allocating responsibilities in the risk treatment process to third parties with respect to the level of risk. Throughout this entire process monitoring and evaluation has to take place and instituting corrective measures where necessary and ensuring correct documentation.

Within this framework the researcher also propose the inclusion of risk ontology to ensure interoperability, formalization and reuse of already gathered information and a general shared understanding of the risk domain.

The main objective of the BN is to facilitate decision making in ISRM taking advantage of it’s merits including; (i) ability to model direct and casual relationships, (ii) graphical representation for easy understanding and reporting, (iii) ability to model data that is incomplete or uncertainty, support for qualitative and quantitative modeling, (iv) bi-directional inference i.e. inputs can be used to predict outputs and outputs can be used to estimate input requirements, (v) they are powerful decision support tools with ability to learn compound by ability to not only predict but also associate a probability with each prediction.

Figure 7.2 depicts the key decision points of the proposed model as facilitated by BN in the generic RM framework. Note that this process is iterative and the modeler can always go back to the initial stages and make adjustments. Before implementation financial institutions will be able to model and predict the various scenarios and the implications of all alternative courses of action and thereby make an informed decisions.

*Figure 7.2: Key decision points (ISO/IEC 27005)*



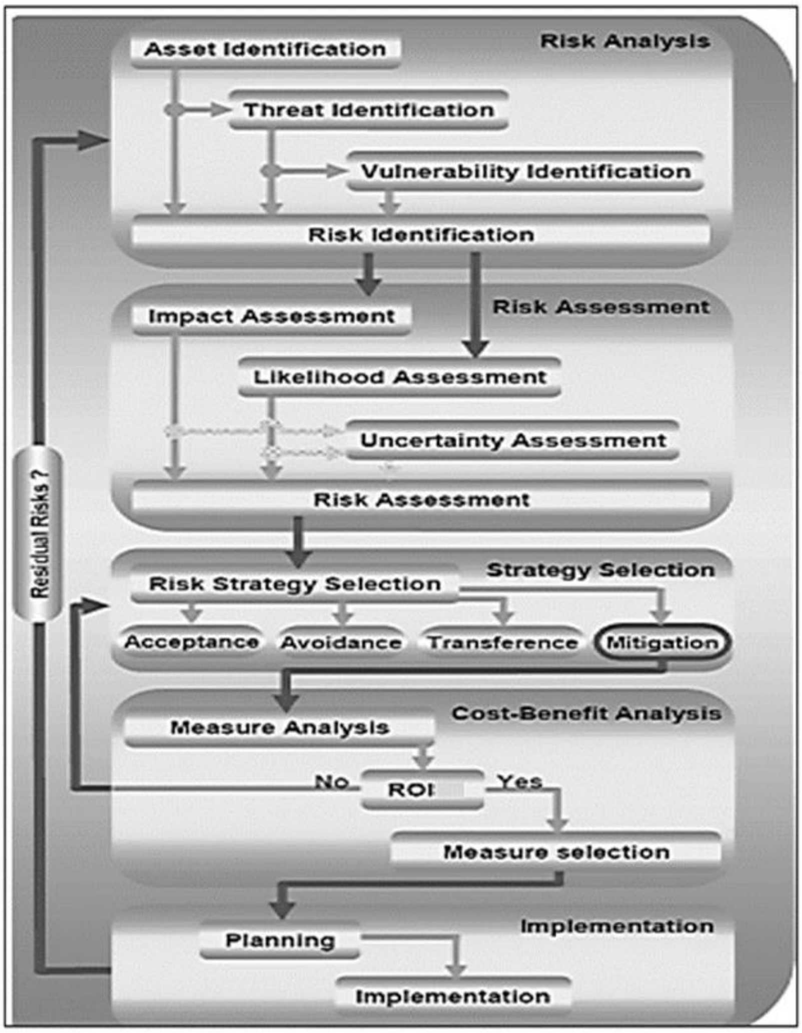
Once context establishment and risk assessment have been conducted, it is necessary to evaluate if sufficient information is available to take decision about risk treatment. If not, a new iteration (maybe partial) with updated context and risk assessment, is conducted.

Otherwise, risk treatment task is performed (Risk Decision Point 1). Several iterations of the risk treatment task could be needed to reach the best state in terms of residual risk and ROI. Moreover, since the effectiveness of the risk treatment depends on the results of the risk assessment, it is possible that no acceptable level of residual risk can be reached. In this case, a revision of the process starting from the context establishment can be necessary to update the different parameters.

After risk treatment, the risk acceptance task has the objective to ensure that residual risks are explicitly accepted by the managers of the organization. Finally, risk communication is a task to be performed throughout the process, to be sure to have all of the relevant information at each task of the process. Thereby, the whole process should be clearly documented.

To further elaborate the framework, let's decompose it as shown in the figure 7.3. After risk analysis and assessment, BN will be extensively used to model (and predict) the effect of a particular risk management strategy on the measure of analysis. These measures will be the indicators for success or failure of strategy selection (for example in this case cost benefit analysis's indicator is return on investment- ROI), another scenario could be for IS defects the indicator could be defects found during operation. The measure of analysis is relative to the objective of the modeler. This projections are valuable tools in facilitating decision making due to its holistic nature that propagates the effects of any choice throughout the entire model.

*Figure 7.3: Risk analysis flow chart*



## 7.4 Model evaluation and Testing

Jakeman et al. (2006) stress that model evaluation should go beyond the traditional attitude of ‘validation’ based only on model accuracy, to also include subjective criteria such as fitness for purpose and transparency of the modelling process. Model evaluation may include: sensitivity of model to plausible parameter changes; critique of assumptions; documentation; critique of the model development process; and ability to perform under a range of conditions including unexpected scenarios.

In an effort to improve the current risk management approach the researcher proposed a generic framework (in the sense that it could be applied across the financial sector) with high ability to be customized to suit the specific needs of a given financial institution. Further, as opposed to the norm, FISRM be initiated in the initial phase of FIS development and continue throughout the cycle; additionally the researcher proposes the adoption of BN for risk management. To test the system the researcher will start by going through the entire FIS development domain and to specific risks from the domain (system defects found in operation) that are encountered after implementation and how the proposed framework comes to play.

### 7.4.1 BBN Parameterization for FIS project Development Domain

The structure of a BBNs can be found from domain knowledge and/or data. It is recommended that the structure of BNs is built based on existing theories, knowledge or hypotheses. BBNs are capable of structural learning from data using a score-based algorithm, which searches for a structure that maximizes the chosen entropy scoring function, or a constraint-based algorithm, which maps out the model structure based on the conditional dependencies found between each pair of variables (Cheng and Greiner 2001, Cansado and Soto 2008).

The model is intended to be used at an overall project level aimed at IS project managers. It is used to predict and assess the overall risk/quality status of a large information system development project. What makes the model so powerful, when compared with traditional software risk management models, is that we can enter observations anywhere to perform not just predictions but also many types of trade-off analysis and risk assessment. The model is based on extensive empirical data drawn from many sources and can be tailored extensively for different classes of projects. The model enables us to predict different aspects of resources and quality while monitoring and mitigating different types of risks.

Prediction and the ability to model the casual relationships between variables are some of the major strengths of BN's approach to risk management. This framework enables financial institutions to manipulate the variables in the nodes which propagates the entire framework; as such risk managers are able to monitor and evaluate the impact of the various risk management courses of action on other factors from a holistic perspective based on their relationships before implementation. This prediction aspect is very important in risk management as it helps managers not only to make better decisions but also in resource allocation.

To illustrate this further consider the “Quality delivered” in the Delivered quality subset in figure 5.4; to manage it risks we must consider all relations “Quality delivered” has on other variables and predict the influence of any RM step to them. One way of doing this by starting from the “Quality delivered” and going backwards against the direction of the arrows; you will notice that the effects goes into other subsets, predicting and evaluating these effects holistically as opposed to locally or linearly is a valuable risk management tool, furthermore this functionality is applicable in evaluation, selection and monitoring of risk mitigants.

*Figure 7.4: Bayesian Belief Network for FIS project development domain*

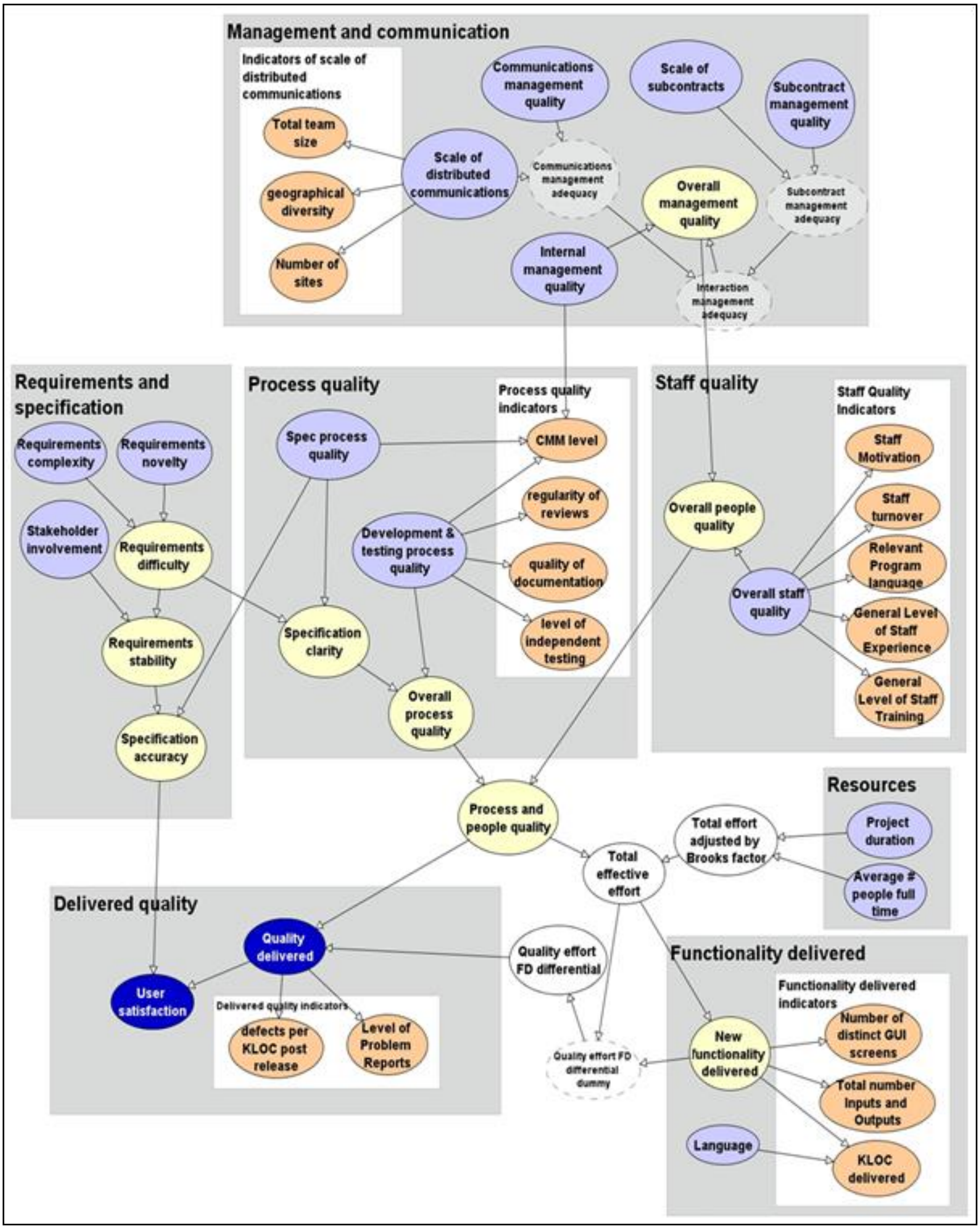
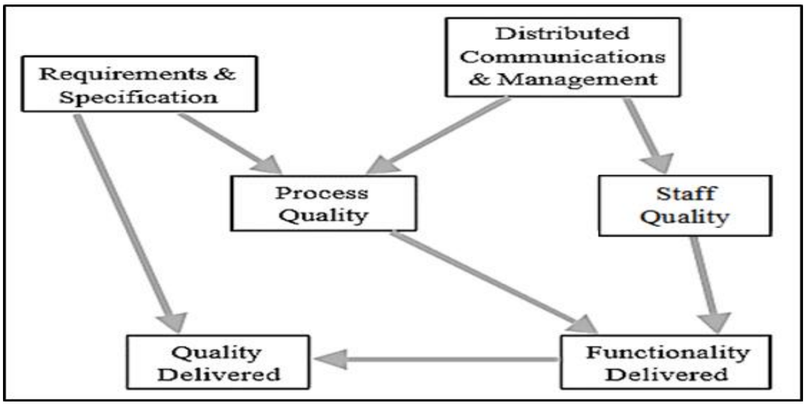

Management and communication
Indicators of scale of distributed communications
Total team size
geographical diversity
Number of sites
Scale of distributed communications
Communications management quality
Communications management adequacy
Scale of subcontracts
Subcontract management quality
Subcontract management adequacy
Overall management quality
Internal management quality
Interaction management adequacy
Requirements and specification
Requirements complexity
Requirements novelty
Stakeholder involvement
Requirements difficulty
Requirements stability
Specification accuracy
Process quality
Process quality indicators
Spec process quality
CMM level
regularity of reviews
Development & testing process quality
quality of documentation
level of independent testing
Specification clarity
Overall process quality
Staff quality
Staff Quality Indicators
Staff Motivation
Staff turnover
Relevant Program language
General Level of Staff Experience
General Level of Staff Training
Overall people quality
Overall staff quality
Resources
Project duration
Average # people full time
Total effort adjusted by Brooks factor
Process and people quality
Total effective effort
Delivered quality
Quality delivered
User satisfaction
Delivered quality indicators
defects per KLOC post release
Level of Problem Reports
Quality effort FD differential
Quality effort FD differential dummy
Functionality delivered
Functionality delivered indicators
New functionality delivered
Number of distinct GUI screens
Total number Inputs and Outputs
Language
KLOC delivered

Figure 7.4 depicts the major risk and risk management areas in the financial institutions information system development cycle domain. The entire model is too complex to be depicted in detail in this study in its entirety; as such it has categorized into subnets for easy understanding and manageability starting from the FIS development risk management. This full model (and its subnets) enables us to cope with variables that cannot be observed directly, this is done by the use of indicators which can be modified to suit the context. Here, the causal link is from the 'quality' to directly observable values for example from 'delivered quality to delivered quality indicators (defects per KLOC post release and level of problem reports). Generally risk management is done after analyzing predictions by manipulating the indicator variables accordingly and propagating the changes, based on this predictions informed decisions are made.

The model enables risk management of FIS in large scale including the ability to model tradeoffs between cost, quality, schedule and functionality. Compared to traditional models, in this BBN observations can be entered anywhere and enable the model to make predictions, analyses and risk assessments on different aspects of resources and quality while monitoring and mitigating different types of risks.

The simplified BBN schematic model in figure 7.5 is used for easy understanding of FISRM

*Figure 7.5: Simplified schematic view of BBN for FIS project development model*



The subnets in the model are:

- Distributed communications and management. Contains variables that capture the nature and scale of the distributed aspects of the project and the extent to which these are well managed.
- Requirements and specification. Contains variables relating to the extent to which the project is likely to produce accurate and clear requirements and specifications.
- Process quality. Contains variables relating to the quality of the development processes used in the project.
- Staff quality. Contains variables relating to the quality of people working on the project.
- Functionality delivered. Contains all relevant variables relating to the amount of new functionality delivered on the project, including the effort assigned to the project.
- Quality delivered. Contains all relevant variables relating to both the final quality of the system delivered and the extent to which it provides user satisfaction (note the clear distinction between the two).

The full model enables us to cope with variables that cannot be observed directly. Instead of making direct observations of the process and people quality, the functionality delivered and the quality delivered, the states of these variables are inferred from their causes and consequences. For example, the process quality is a synthesis of the quality of the different software development processes requirements analysis, design and testing.

The quality of these processes can be inferred from ‘indicators’. Here, the causal link is from the ‘quality’ to directly observable values like the results of project audits and of process assessments.

The strength of this approach is in the fact that it allows the model to be adapted to use whichever indicators are available.

At its heart the model captures the classic trade - offs between: quality, effort, time and functionality.

Quality (where we distinguish and model both user satisfaction – this is the extent to which the system meets the user's true requirements and quality delivered this is the extent to which the final system works well).

Effort (represented by the average number of people full – time who work on the project).

Time (represented by the project duration) and functionality (meaning functionality delivered).

So, for example, if you want a lot of functionality delivered with little effort in a short time then you should not expect high quality. If you need high quality then you will have to be more flexible on at least one of the other factors (i.e. use more effort, use more time or deliver less functionality).

What makes the model so powerful when compared with traditional software cost models is that observations can be entered anywhere in the model to perform not just predictions but also many types of trade - off analysis and risk assessment. So we can enter requirements for quality and functionality and let the model show us the distributions for effort and time. Alternatively we can specify the effort and time we have available and let the model predict the distributions for quality and functionality delivered (measured in function points).

### 7.4.2 Model for Predicting software defects

The implementation of the proposed framework can be used in various scenarios depending on the objective. To further test the proposed framework the researcher will model a prediction for information system defects and reliability (as there is a direct relationship between the system defects and reliability of an information system) (Fenton and Neil, 2013). Further the model cuts across all the phases involved FIS life cycle i.e. from development to implementation and system use. It should be noted that the framework can however be used in risk management in the entire FISRM domain.

Fenton and Neil, (2013) note that software defects found in operation are encountered by users; this means they will be found after system implementation. The proposed model provides a way of handling this throughout the SDLC phases and ensure minimal defects are found in operation hence a reliable system. By modeling from prior knowledge and updating the model in the light of new knowledge and also making adjustments to the model and predicting the outcome; the FIs will be able to make informed decisions without “trial and error” and the cost that comes with it. Backward and forward inference made is also used to determine cause-effect including those of direct and prior relationships.

Defects have a very significant impact on the reliability of the system which in turn poses one of the major risks in any information systems development project, implementation and usage. The defects have an impact on; the final user (efficiency and effectiveness of task performance), cost of fixing the defects, human resource consumption, and delay in delivering the new product or feature to mention but a few and by extension the reliability of the IS.

Reliability; the probability a system will satisfactorily perform the task for which it was designed or intended, for a specified time and in a specified environment determines the overall quality of an IS, and identifies the areas of vulnerability. As such it is one of the critical FISRM areas and for this reason was selected to be used in the proposed model.

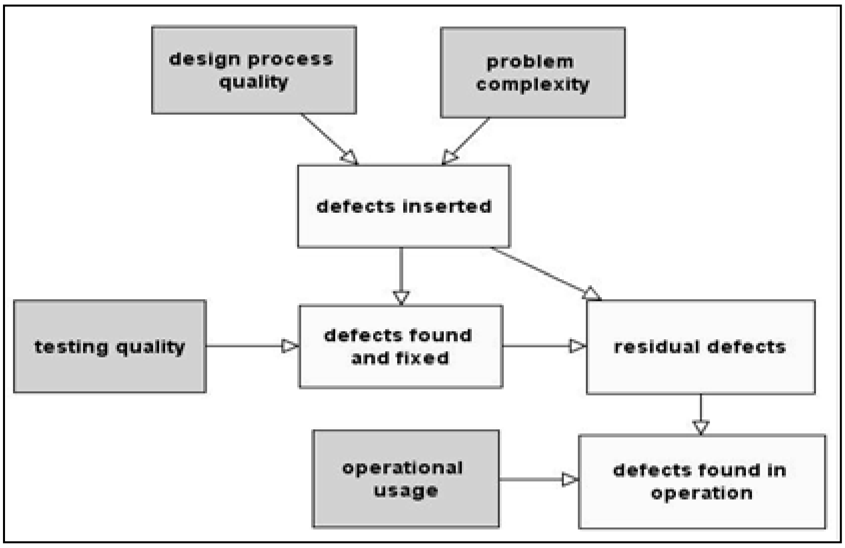
One of the main objectives of a IS development metrics (measure of degree to which a system, component or process possesses a given attribute) program is to achieve process improvement to do this it looked at those projects that, in metrics terms, were considered most successful. These are the projects with especially low rates of customer-reported defects, measured by defects per thousand lines of code (KLOC).

One of the classic weakness of traditional software metrics the omission of sometimes obvious and simple causal factors that can have a major explanatory effect on what is observed and learnt.

Operational defects **-** those found by customers and are dependent on the number of residual defects but also critically dependent on the amount of operational usage.

Residual defects **-** The number of residual defects is determined by the number of defects introduced during development minus the number successfully found and fixed after testing. This is dependent on the number introduced which is influenced by problem complexity and design process quality. Also, the number of defects found is influenced amount of testing effort.

*Figure 7.6: BN causal model for software defect prediction and reliability prediction*



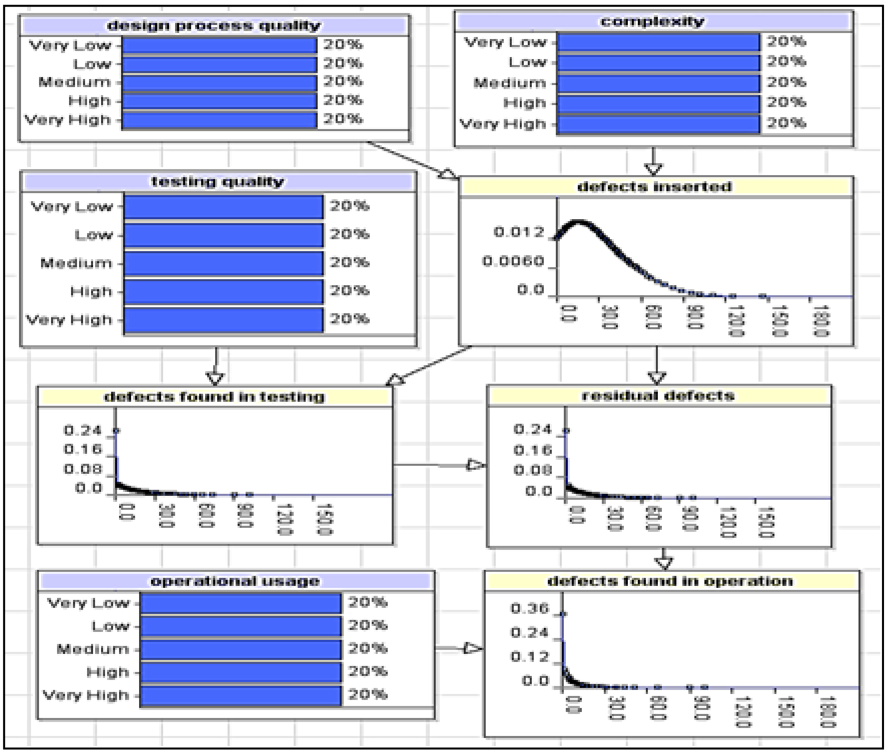
In this casual model the number of operational defects (those found by customers) in an IS module is what we are really interested in predicting. It is not only dependent on the number of residual defects but also critically dependent on the amount of operational usage. If the system is not used no defects will be found irrespective of the number therein (the model allows you to predict this with perfect precision). The number of residual defects is determined by the number introduced during development minus the number successfully found and fixed, of course defects found and fixed is dependent on the number introduced. The number introduced is influenced by problem complexity and design process quality. The better the design the fewer the defects and the less complex the problem the fewer defects. The number of defects found is influenced not just by the number therein but also by the amount of testing effort.

In regards to risk mitigation, this framework uses the casual relationships between the various variables within and outside their subsets which are then manipulated accordingly to mitigate the risks based on risk evaluation (as indicated in parameterization figure 7.4). For instance, considering the model for defect prediction (figure 7.6) to mitigate the risks of defects found in operation; we follow the arrows backwards to the casual relationships (in this case residual defects and operational usage). Residual defects are also influenced by two factors mapping this relationships with a risk ontology and manipulating the variables and predicting the outcome before implementation helps choose the best mitigation option.

*Prior marginal distributions*

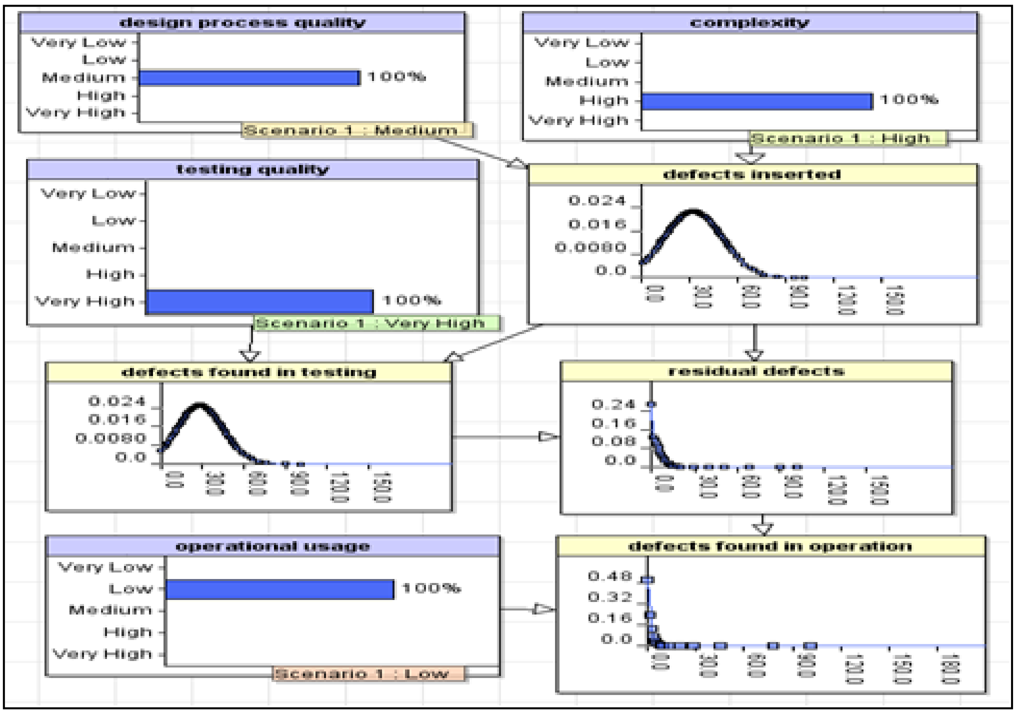
The financial sector risks vary greatly from institution to institutions based on their internal and external environments. To test the model further the researcher considers an environment with even prior distributions for the independent variables each with an equal chance (20%) of occurrence (assuming all the other factors are constant). For instance, for the design process quality of an FIS is just as likely to have very high complexity as very low, and that the number of defects found and fixed in testing is in a wide range. This is especially the case for new organizations or ones that are developing new systems. The model in figure 7.7 indicates uncertainties based on initial knowledge and data from the domain before any observations are made. Prior marginal distributions as indicated in figure 7.7 represents uncertainty model before any specific information is input that is the module i.e. it is a model based on prior data.

*Figure 7.7: Prior distributions*



As we enter observations or in light of new knowledge about the domain the probability distributions are updated and propagated throughout the model as indicated in figure 7.8. Suppose the modeler based on new knowledge determines that the design process quality is medium, high complexity and very high testing quality and low operation usage;

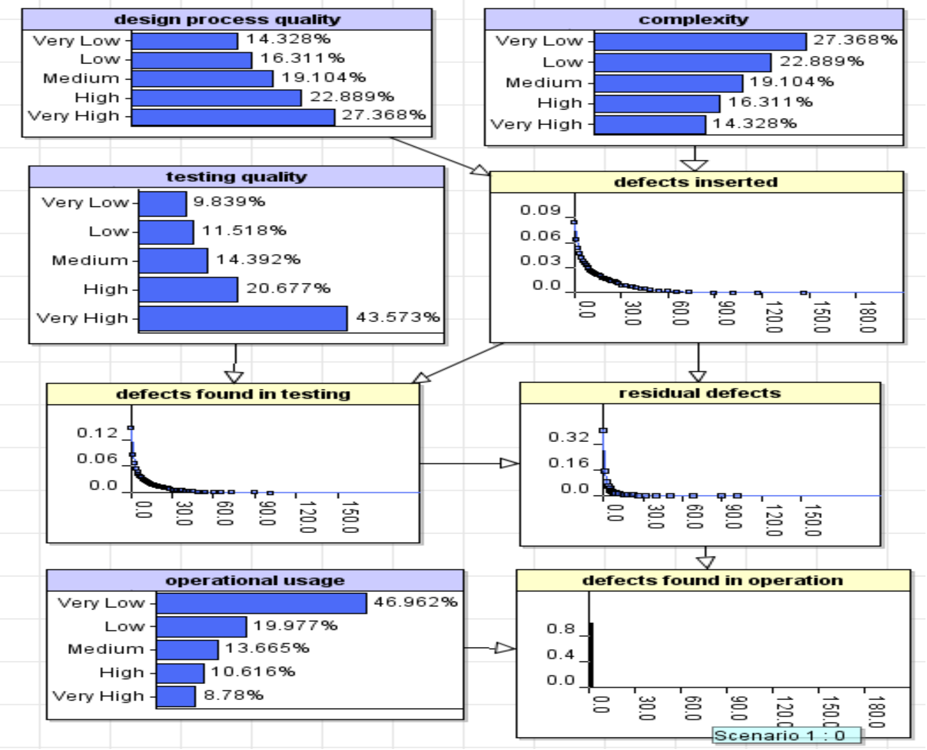
*Figure 7.8: Scenario 1; based on new knowledge/observations*



From the model in figure 7.8 there is a high number of defects found in testing with a mean of 35, when we compute the predicted operational defects is close to a mean of 1. This prediction could be explained by the fact that; due to the very high testing quality most of the defects were found and fixed and also since operational usage is low some of the residual defects would not be found.

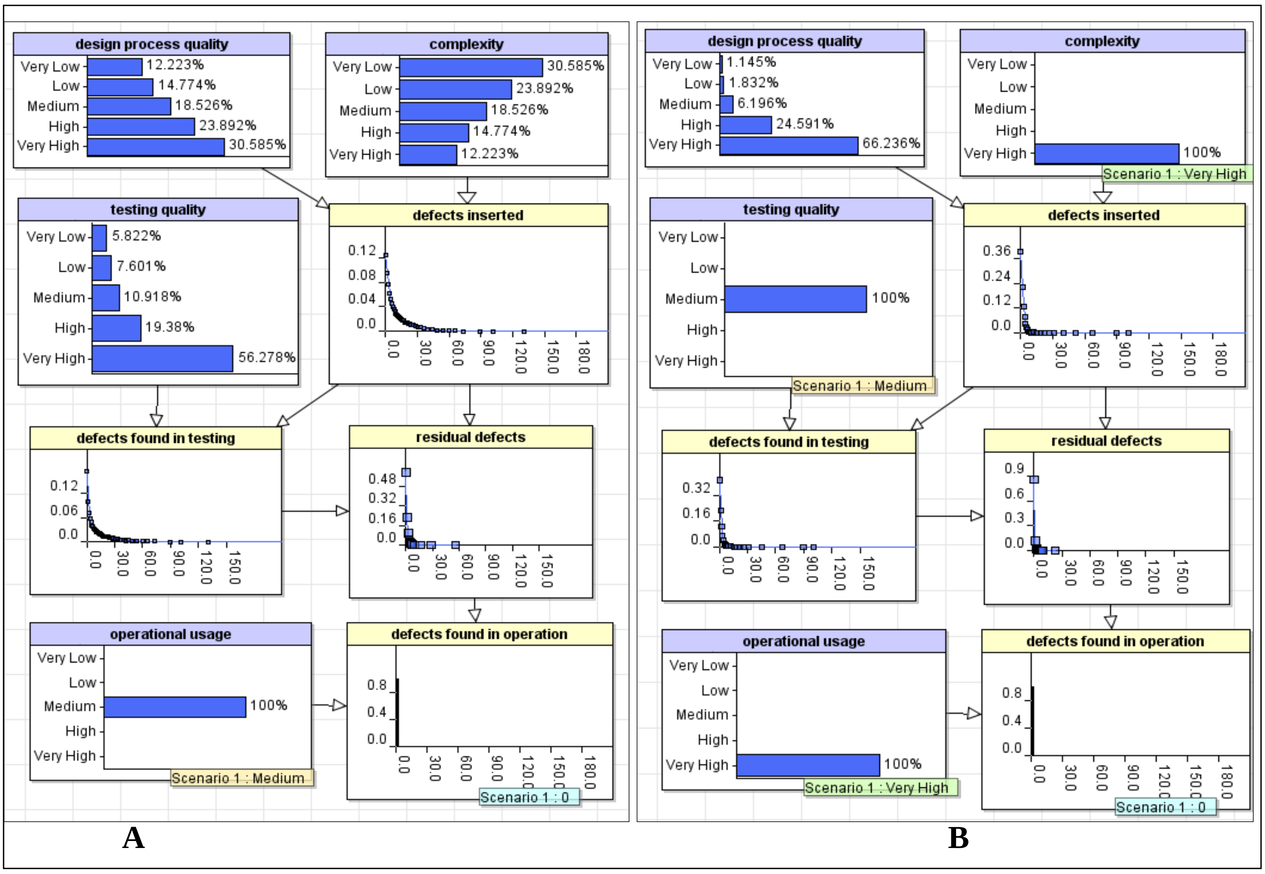
Let’s look at a scenario where no defects were found in operation, upon running this through the model. The most likely explanation to this observation as presented by the model (figure 7.9) is very low operational usage and also that testing and process quality was higher than average and problem complexity was lower than average.

*Figure 7.9: Scenario 2; Zero defects in operation*



Suppose we discover that operational usage is medium, running this through the model we realize that it (the model) is convinced that the explanation is that the testing quality was so good that most of the residual defects were found and fixed, the problem complexity was below average and the design process quality was above average. This is illustrated in figure 7.10 A. Suppose the over a period of time the modeler realizes that the testing quality was actually medium, the problem complexity high, and the operational usage very high. The model believes that very high design process quality is the reason for the result (i.e. zero defects).

*Figure 7.10: Model for multiple observations*

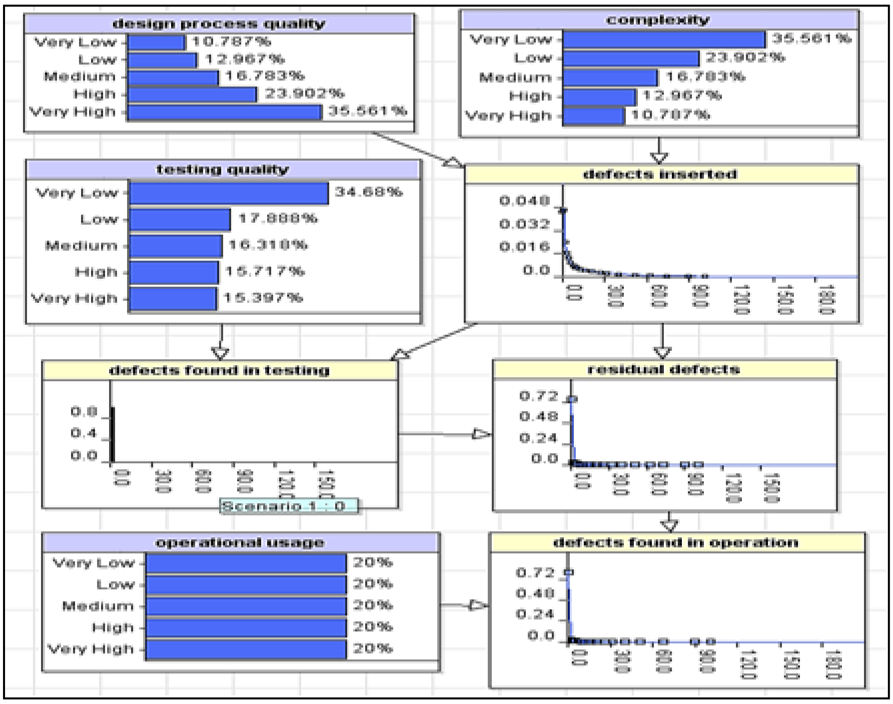


### 7.4.3 Unexpected Scenarios

Literature and prior knowledge has indicated that all factors constant it is highly improbable to find no defects in testing. Suppose however that we find out that this is the case during development of a module. Inputting this scenario into the model we have the model in figure 7.11 that explains this; the design process quality was way above average and the problem complexity must have been very low.

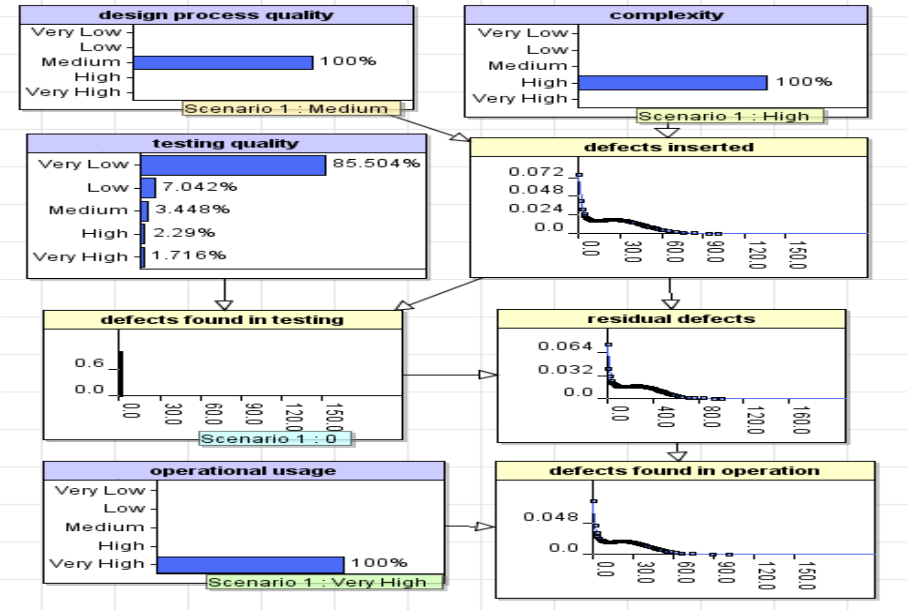
Here the model the complexity was below average, while the design process quality was above average and the testing quality below average as indicated in figure 7.11.

*Figure 7.11: Model for zero defects in testing*



Let us further consider the event that this knowledge is updated such that; defects found in testing = 0, design process quality = 'medium', complexity = 'high', operational usage = 'high' and no observation is entered for testing quality. Upon running the model in figure 7.12 produces the following output.

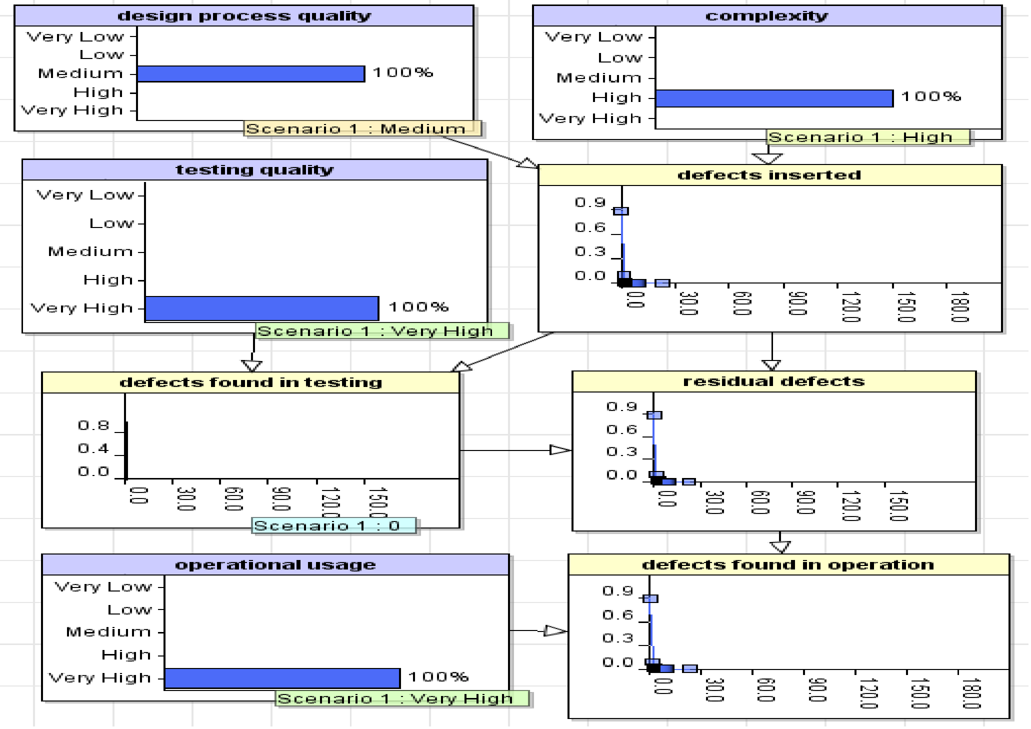
*Figure 7.12: Model for other observations*



The model predicts the number of operational defects to have a mean distribution of 21.693 and that the testing quality must have been very low. As such, most of the defects inserted were not found and fixed and so were passed over to the users of the system who found them during operation.

Suppose the manager notices this in the model and improves the testing effort to very high, upon updating the model the model prediction for operational defects drops dramatically to a mean of 0.126. Critical analysis reveals that given rigorous testing by experts and no errors were found, minimal errors would be found by system users. The beauty of BN comes to light here; notice that by improving the testing quality the defects inserted have also reduced as shown in figure 7.13. This is explained by the casual relationship between staff quality and quality delivered (see figure 7.4). By improving testing quality it means the quality of the testing staff, who technically are the developers (especially for alpha tests) this by extension means that the design process quality will improve explaining the reduction in defects inserted.

*Figure 7.13: Model for multiple observations*

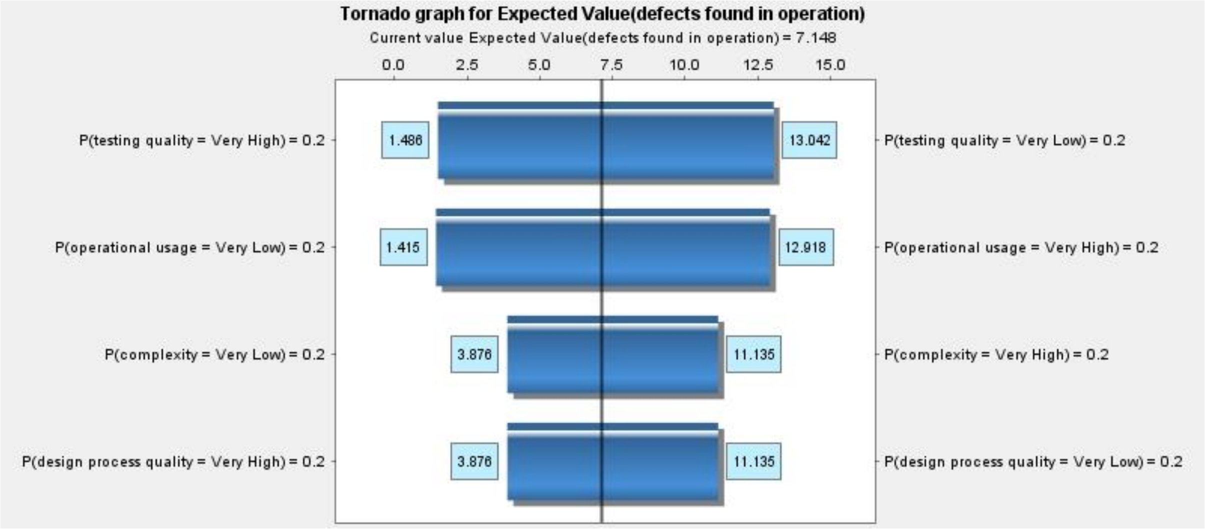


### 7.4.4 Sensitivity Analysis

Generally, sensitivity analysis is defined as the study of how uncertainty in the output of a model can be attributed to different sources of uncertainty in the model input. The sensitive variable is modeled as uncertain value while all other variables are held at baseline values (stable), (Steffes-lai, 2014). In model and risk analysis in general sensitivity analysis is crucial in evaluating risks and potential ways to mitigate them. There is a very wide range of uses to which sensitivity analysis is put including;

- Support decision making or the development of recommendations for decision makers (e.g., testing the robustness of a result).
- Enhance communication from modelers to decision makers (e.g., by making recommendations more credible, understandable, compelling or persuasive).
- Increase understanding or quantification of the system (e.g., understanding relationships between input and output variables).
- Model development (e.g., searching for errors in the model and testing).

*Figure 7.14: Sensitivity analysis of defects found in operation*



The sensitivity analysis indicates that testing quality is the highest determinant of defects found in operation closely followed by operational usage as indicated in the tornado graph in figure 7.14. Below are additional sensitivity analyses of the various variables

*Figure 7.15: Defects found in operation|design process quality*

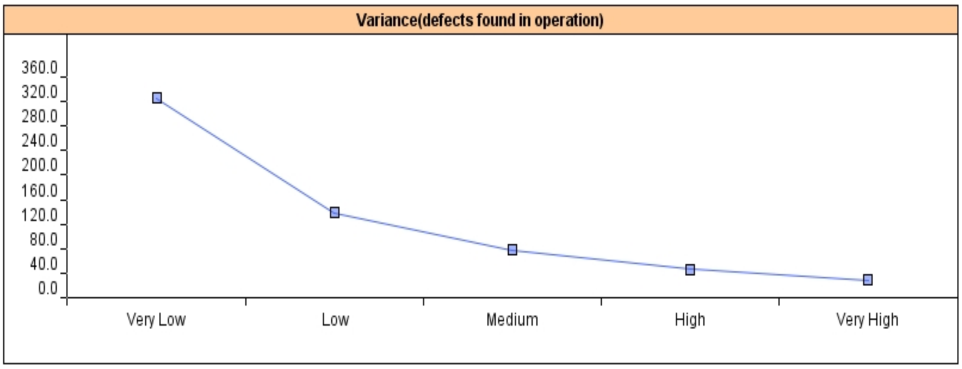
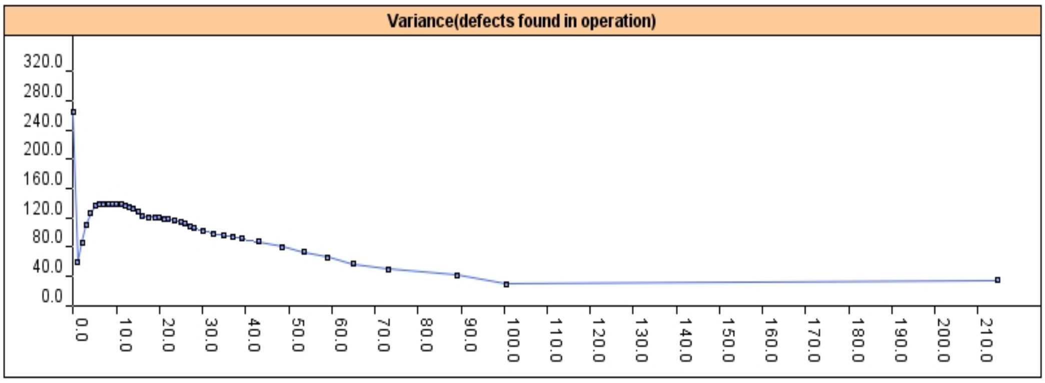


*Figure 7.16: Defects found in operation|defects found in testing*



As the framework proposes once the modeling is done and all possible courses of actions have been factored; the organization makes a decision about the risk management activities to be implemented starting from the initial SDLC phase. For instance after considering all the possible scenarios as far as system defects found in operation are concerned the financial institution then begins the risk management from the design process quality and the problem complexity. Technically, these stages in the SDLC are the system design and system requirement and analysis phases. So as the system development progresses so does risk management; of course this is done inconsideration with other project factors.

## 7.5 Contribution of Research

This research proposes the initialization of FISRM from the initial stages of the SDLC as opposed to the current approach that manages it from the implementation phase. Additionally it proposes the introduction of BN into the risk management program. By introducing BN to replace the frequentist approach, the researcher has resolved the weakness that exist in the current approach which manifests when there is no data or no experience history (N =0), (advent of a new phenomenon such as new technology) in which case the frequentist approach collapses. Other weakness include the inability to use incomplete or disparate data. However BN could use other data such as expert knowledge, intuition knowledge and update this knowledge when new information becomes available.

There has been research on risk management of information systems both locally and internationally. However, these researchers have only addressed the aspect of risk management in regard to already developed systems. Furthermore, locally no research has been done in regards to proposition of a risk management model that could; (1) be highly customized as to present local solution to local risks facing financial institutions (2) with the ability to be used as a central repository (off course with the facilitation and oversight of a regulatory authority) for financial institutions to not only be able to update the database but also use it to make their own risk management strategies (3) use all the available information be it disparate, incomplete and in some instance non-existent; also uses knowledge from experts and past experiences. (4) Ability to update information from anywhere in the node in light of new information and have it propagated throughout the model and make predictions on effects of manipulation of the variables.

The proposed framework unlike the existing system provides a method of integrating all the risk factors, their proposed mitigation techniques and other related factors (such as cost) and the relationship between them and a way of modelling and predicting the possible outcome without actual implementation. This is a valuable tool to aid in decision making.

It also proposes a the use of more than one technique in the approach; that is BN, statistics, AI and the inclusion of all available information (complete, incomplete and disparate) from experts and past experiences. Further it has the ability to learn, if modelled correctly and over time this approach as compared to those currently in use that are based on scientific foundations of statistical decision making is way more exhaustive with prediction capabilities hence economical. This is valuable in the risk management industry.

## 7.6 Conclusion

The research has indicated that majority of the financial institutions develop their FIS internally through their staff or outsourcing this service and are involved in the development of their systems. They also have a formal financial information system risk framework and assessment process in place but lack a common risk language that is broadly used and understood across the financial sector.

In regards to reporting and metrics there exists a defined financial information system risk performance framework and metrics that are monitored and reported to management; additionally IS Managers receive sufficiently regular and robust information on risk from RM teams.

There was a considerable projected increase in financial investment in FIS risk management activities in the next two years which would be mainly allocated in process automation technologies and control self-assessment respectively with the lowest allocation being on staffing. The general perception is that the current FISRM program is not sufficient to ensure efficiency and effectiveness.

The analysis of FIS risks showed that under the development cycle risks the implementation and system design posed the highest risks factors. Under the development environment the work environment and management methods posed highest risk. Programmatic risks were majorly attributed to program interface risks and contract risks. This findings are significant as they not only show how much risk is present throughout the SDLC with a significant number of then from the initial stages of system development phases.

The financial institutions were found to be using a myriad of Standards/leading practices in developing their financial information system risk framework and assessment processes.

Two of the major challenges faced by the current FISRM program are; limitation on sharing past and present knowledge/information across the financial sector and the inability to utilize all the available data/information and knowledge gathered from system risk framework and assessment process. Similarly, three of the major desirable improvements to the current system are: (1) the ability to make predictions/forecasts on risks, (2) alternative causes of action, implication risk management activities on enterprise prior to implementation, ability to learn and integrate past and present knowledge and information and (3) a holistic/enterprise wide approach to FISRM.

On the analysis of the application of Bayesian Networks in FISRM the study found out that FISRM would be better if done throughout the development phases as opposed to the current approach where it's mostly done after implementation. Also majority of the respondents that had knowledge of BN felt that its incorporation into FISRM program had the potential to significantly improve its' efficiency and effectiveness.

The study indicates that financial information systems risk management programs have considerably evolved in the recent past and shown significant resolve to face the challenges with most organizations formalizing their programs and continuing to focus on maturity of their programs and better ways to manage their information systems risks. There is a positive trend seeing as formalization of the various organizations risk management programs are underway; also these organizations are increasing their investments in risk management programs. This indicates that these programs are still developing with few in early maturity. Ideally, spending should level off and may even decrease as a program gains maturity and is optimized and investments typically increase at higher rates when a program is in development. Reporting and metrics are sufficiently done by majority of the organizations however incorporating all this knowledge into the current risk management program seems to pose a challenge.

### 7.7 Recommendations

Organizations involvement in FIRM should not only start after system implementation, FIS are prone to significant risk way before implementation. As such an effective financial information system risk management program should involve should be initiated at the initial phase of the SDLC and span the entire life cycle to ensure maximum benefit to the organization (efficiency and effectiveness). This would also save on costs as a result of compounding of risks that are not handled early. A convergent FISRM framework is needed for the current and future RM program.

The goal of convergence is to design a program, organization, and processes that can better manage risk through adequate measures and monitoring methods on a sustainable, consistent, efficient, and transparent basis. This will result to a mature and effective financial information system risk management program that is flexible, efficient, and sustainable that supports not only today's business requirements, but those of the future. Early/timely initiation of FISRM coupled with the incorporation of BN presents a frame work that that brings all the above to the table; it is also a dynamic and flexible approach that does not work in isolation but in consideration to leading standards and practices of risk management.

BN present tremendous potential to the practice of FISRM, if well implemented. As such implementation should be done with the bigger picture in mind, meaning short term, tactical, and strategic objectives of the organizations should be considered. As with any major technological adoption, the initial investment is expected to be high but will be highly cost effective strategically. Additionally, a repository should be considered with access levels for financial institutions to enable access to already available information. Holistically this could be looked at from a sectorial perspective, that is, the entire financial sector with the facilitation of a regulatory body such as the central bank could form such a frame work with access to all financial institutions in the country. With such a framework in existence all formalization of reporting and metrics, common risk libraries this repository would be a database for the entire FISRM both new and existing and would take RM to a whole level.

There is no question that that the dynamic technological world keeps changing presenting new opportunities and risks in equal measure, financial institutions have no option but to keep up to remain competitive. It is therefore imperative that their RM initiatives do so too, and what better way to do this than the adoption of a framework that learns and evolves over time –BN. This should however be done systematically and requires intense planning and testing. The use of a common risk language and metrics that is understood across the organization (and or sector) is of utmost importance to ensure quality reporting. Documentation, communication and consultation should be done at every stage effectively and efficiently. The details of the proposed framework are presented in detail in this section.

## 7.8 Suggestion for Future Research

This research focused on the financial sector, further research should be made on the applicability of the proposed model on other sectors/industries. The researcher had proposed a consideration into implementing a national (or perhaps regional) repository  that would be used by all the financial institutions and their stakeholders primarily for RM. Research should be done to determine the feasibility of this initiative and the best implementation approach and methodology to ensure success.

# References

Acton A. (2013). *Issues in Industrial, Applied, and Environmental Chemistry*: (2013 Edn). Scholarly Editions. Available at: https://books.google.co.ke/books?id=SiKyrs2JN1oC

Adusei-Poku K. (2005). *Operational Risk management - Implementing a Bayesian Network for Foreign Exchange and Money Market Settlement*. University of GÄottingen. Kumasi, Ghana.

Australian Council on Healthcare Standards (ACHS) (2013). *Risk Management and Quality Improvement Handbook*. Sydney Australia; ACHS;

AIRMIC, Alarm, IRM: (2010). *A structured approach to Enterprise Risk Management (ERM) and the requirements of ISO 31000*. Available at: https://www.theirm.org/media/886062/ISO3100_doc.pdf

AS/NZS ISO (2009). *Risk management: Principles and guidelines* (AS/NZS ISO 31000:2009). Sydney, NSW 2001, Standards Australia and Wellington 6140 Standards New Zealand.

Barber D. (2012). *Bayesian Reasoning and Machine Learning*. Cambridge. Cambridge University Press

Barton N., Kuikka S. Varis O., Uusitalo L., Farmani R. and Johnson S. (2008). Bayesian belief networks as a meta-modeling tool in integrated river basin management Pros and cons in evaluating nutrient abatement decisions under uncertainty in a Norwegian river basin. *Ecological Economics*, 66, 91-104.

Bashar A., Parr G., McClean S., Scotney B. and Nauck D. (2010). Knowledge Discovery Using Bayesian Network Framework for Intelligent Telecommunication Network Management. *Research and Technology*. British Telecom, Adastral Park, Ipswich IP5 3RE, UK

Berg H., P. (2010). Risk management: procedures, methods and experiences. In *proceedings of the international symposium SYMORG*. RT&A # 2(17). (Vol.1).

Bessis J. (2010). *Risk Management in Banking*. Crosspoint Blvd., Indianapolis, IN 46256, USA. Wiley Publishers.

Bessiere P. Mazer E. Ahuactzin J.M. and Mekhnacha K. (2013). *Bayesian Programming*. Chapman & Hall/CRC machine learning & pattern recognition series. CRC Press

Ben-Gal I., (2007). Bayesian Networks, in Ruggeri F., Faltin F. & Kenett R., *Encyclopedia of Statistics in Quality & Reliability*, Wiley & Sons

Beynon-Davies D. (2013). *Business Information Systems*. Basingstoke UK, Palgrave Macmillan

Borsuk H. et al. (2004) A Bayesian network of eutrophication models for synthesis, prediction, and uncertainty analysis. *Ecological Modelling*, 173, 219-239.

Borek A., Parlikad K. A., Webb J., Woodall P. (2013). *Total Information Risk Management: Maximizing the Value of Data and Information Assets*. Newnes. Elsevier Science

Broad, J. (2013). *Risk Management Framework: A Lab-Based Approach to Securing Information Systems*. Newness. Elsevier Science

Carl L. (2014). *Risk Management: Concepts and Guidance*, (5th Ed.). FL., USA. Taylor & Francis

Cheng J., and Greiner R., (2001). *Learning Bayesian Belief Network Classifiers: Algorithms and System*. Available at https://webdocs.cs.ualberta.ca/~greiner/PAPERS/cheng-CSCSI01.pdf

Cortez A. (2011). Winning at Risk: Strategies to Go Beyond Basel: Volume 638 of *Wiley Finance.* Hoboken, NJ; John Wiley and Sons

Cansado, A., Soto, A. (2008) Unsupervised anomaly detection in large databases using bayesian networks. *Applied Artificial Intelligence* 22(4), 309–330 5.

Cooper R. and Schindler S.  (2011). Business Research Methods. McGraw-Hill/Irwin Series Operations and Decision Sciences. *The McGraw-Hill/Irwin series operations and decision sciences.* Business statistics. McGraw-Hill/Irwin

Creswell, W. and Clark L. (2011). *Designing and Conducting Mixed Methods Research,* Thousand Oaks, CA: Sage Publications

Chornous G. and Ursulenko G. (2013). Risk management in banks: new approaches to risk assessment and information support. ISSN 1392-1258. *EKONOMIKA* 2013 Vol. 92(1)

Daly R. et al, (2011). Learning Bayesian networks: approaches and issues. *The Knowledge Engineering Review*, Vol. 26:2, 99–157. & Cambridge University Press.

Davendranath G. H. (2013). *Computer concepts and management information systems*. Patparganj, New Delhi. PHI Learning Pvt. Ltd

Deloitte L. and Touche P. (2012). Risk assessment in practice. *The Committee of Sponsoring Organizations of the Treadway Commission (COSO).* AICPA, Leigh Farm Rd., Durham

Easterby-Smith M., Thorpe R. and Jackson P. (2012). *Management Research.* SAGE Publications. Available at: http://www.lums.lancs.ac.uk/dml/profiles/64/

Emmel N., (2013). *Sampling and Choosing Cases in Qualitative Research*: A Realist Approach. Thousand Oaks, CA. Sage Publications.

Farmani R. et al (2009). An evolutionary Bayesian belief network methodology for optimum management of groundwater contamination. *Environmental Modelling & Software,* 24, 303-310.

Feng N. and Xie J. (2011). A Bayesian networks-based security risk analysis model for information systems integrating the observed cases with expert experience. *Scientific Research and Essays* Vol. 7 (10), pp. 1103-1112. College of Management and Economics, Tianjin University, 300072 Tianjin, China.

Fenton and Neil (2011). *The use of Bayes and causal modelling in decision making, uncertainty and risk*. Risk and Information Management Research Group. Queen Mary University of London.

Fenton N., and Neil M. (2013). *Risk Assessment and Decision Analysis with Bayesian Networks*. CRC Press. Available at: https://books.google.co.ke/books?id=Wg7LBQAAQBAJ

Fenz S. (2011). From the resource to the process risk level. *Proceedings of the South African Information Security Multi-Conference*: Port Elizabeth, South Africa. Available at: http://Lulu.com, 2011

Gaidow S. and Boey S. (2009). *Australian Defence Risk Management Framework: A Comparative Study*. Commonwealth of Australia, DSTO Systems Sciences Laboratory. Edinburgh South Australia 5111 Australia

Gaol F., Mars W. and Saragih H., (2014). *Management and Technology in Knowledge, Service, Tourism & Hospitality*. FL., USA Taylor & Francis

Gay R. and Peter A. (2012). *Selecting Samples*. Prentice Hall. Upper Saddle River, New Jersey, Columbus Ohio.

Grover J. (2012). Strategic Economic Decision-Making: Using Bayesian Belief Networks to Solve Complex Problems. *SpringerBriefs in Statistics*, Volume 9 SpringerLink : Bücher. Springer Science & Business Media,

Hardcastle E., (2008). *Business Information Systems*. Elizabeth Hardcastle & Ventus Publishing ApS.

Hart T. and Pollino A. (2009). *Bayesian modelling for risk-based environmental water allocation.* Canberra: National Water Commission.

Heidrich, J., Oivo, M., Jedlitschka, A., and Baldassarre, M.T. (2013). Product-Focused Software Process Improvement: *14th International Conference*, *PROFES* 2013, Paphos, Cyprus, June 12-14, 2013, Proceedings. Springer Berlin Heidelberg.

Ho V. (2008). Bayes theorem and its applications in quantitative risk assessment. Retrieved from www.hkrams.org

Holmes D. (2010). *Innovations in Bayesian Networks: Theory and Applications*. Berlin Heidelberg: Springer

Hughes C. (eds.) (2012). *Valuing People and Technology in the Workplace: A Competitive Advantage Framework*. Hershey, Pennsylvania, IGI Global.

ISO 27000 Directory. *The ISO27001 Certification Process*. Retrieved April 10, 2010, from http://www.27000.org/ismsprocess.htm.

ISO, (2012). *ISO 31000 - Risk management*. Available at http://www.iso.org/iso/home/standards/iso31000.htm.

Jesionek R. (2008). *Introduction to IT Governance* (In Polish), CEO – Top Managers magazine, 05.04.

Johanson H. (2012). Determining Project Requirements, (2nd Ed.): Mastering the BABOK® and the CBAP® Exam. *ESI International Project Management Series*. CRC Press.

King, S. (2012). *Research Methods for Information Systems*. Mercury Learning & Information Availlable at: https://books.google.co.ke/books?id=ZvVWuQAACAAJ

Kim A. et al. (2012). Compliance Risk Assessment Measures of Financial Information Security using System Dynamics. *International Journal of Security and Its Applications* Vol. 6, No. 4. Korea University, Financial Security Agency, Korea

Koski and Noble, (2012). A Review of Bayesian Networks and Structure Learning. *Mathematica Applicanda* Vol. 40(1) 2012, p. 53–103.

Kouns and Minoli (2011). Information Technology Risk Management in Enterprise Environments: *A Review of Industry Practices and a Practical Guide to Risk Management Teams*. 111 River Street Hoboken NJ, 07030. United States. John Wiley & Sons.

Larrañaga et al, (2013). *A review on evolutionary algorithms in Bayesian network learning and inference tasks.* Information Sciences. Elsevier Inc.

Lang M. (2011). *IT Architecture and Risk Management*. Munich, Germany. GRIN Verlag.

Lazaros S. and Prodromos (2011). Software Development Project Risk Management: A New Conceptual Framework. *Journal of Software Engineering and Applications*, 2011, 4, 293-305

Marcot B. et al., (2006). Guidelines for developing and updating Bayesian belief networks applied to ecological modeling and conservation. *Canadian Journal of Forest Research*, 36, 3063.

MAS. (2013). *Technology risk management guidelines*; Monetary Authority of Singapore. Singapore.

Marshall D., and Brainerd G., (2010). *Just Enough Software Architecture: A Risk-Driven Approach. Available at:* https://books.google.co.ke/books?id=ITsWdAAzVYMC

Madill K. (2003). *AS/NZS 4360:1999 Risk management.* Standards Australia. Australia.

McManus J. (2012). *Risk Management in Software Development Projects*. FL., USA Taylor & Francis

Manikandan A., Anbuoli P., and Saikishore E. (2011). International Conference on Computer Applications - *Computer Applications I.* Research Pub. Services.

McKay R. (2006). *Systems Development Life Cycle Framework*. National Institutes of Health; Rockledge Dr. MSC7740, Bethesda.

Menezes J. (2013). Defining Indicators for Risk Assessment in Software Development Projects. *Clei electronic journal*, volume 16, number 1.

Poole D. and Ramon M (eds) (2014). Uncertainty in Artificial Intelligence: *Proceedings of the Tenth Conference on Uncertainty in Artificial Intelligence*. University of Washington, Seattle, July 29-31, 1994. Elsevier Science

Neuman L. (2012). *Basics of Social Research: Qualitative and Quantitative Approaches*. New Jersey,USA: Pearson Education.

Nielsen T., and JENSEN F. (2013). *Bayesian Networks and Decision Graphs*. New York. Springer Science & Business Media.

NIST, (2010). *Guide for Assessing Security Controls in Federal Information Systems and Organizations*. NIST Special Publication 800-53 A. Gaithersburg, MD 20899-8930. National Institute of Standards and Technology. Available at: http://csrc.nist.gov/publications/nistpubs/800-53A-rev1/sp800-53A-rev1-final.pdf

OCC (2012*). Supervisory guidance on model risk management*. Available at: http://www.occ.treas.gov/news-issuances/bulletins/2011/bulletin-2011-12.html

O' Mahony, A. and Doran, J. (2008). The Changing Role of Management Accountants; Evidence from the Implementation of ERP Systems in Large Organisations'. *International Journal of Business and Management,* Vol. 3, No. 8, pp. 109-115

Orzechowski .R. *Effective application of IT in the enterprise,* (In Polish) E-mentor no 3(20)/2007, June 2007

Patrick D. Howard (2011). *FISMA Principles and Best Practices: Beyond Compliance*. 6000 Broken Sound Parkway NW. CRC Press.

Pandey K. and Mustafa K. (2012). A Comparative Study of Risk Assessment Methodologies for Information Systems. *Bulletin of Electrical Engineering and Informatics*. Vol.1, No.2. New Delhi.

Parsons J. and Oja D. (2013). *New Perspectives on Computer Concepts 2014: Comprehensive*. Boston, MA. Cengage Learning

Purtell .T. (2008). New View on IT Risk: Building a successful Information Technology risk management program*, The RMA Journal*. Philadelphia, March 2008

Pollino A, and Henderson H. (2010). *Bayesian networks:* A guide for their application in natural resource management and policy. Technical Report no. 14. CERF Hub: Landscape Logic

Pieplow B. (2012). *Project Risk Management Handbook: A Scalable Approach.* Risk management task group. Caltans

PwC, (2013). *Financial Institution Technology Risk Management*. US, PricewaterhouseCoopers LLP. Available at http://www.pwc.com/fsi

Rasmussen S. et al (2013). *Bayesian network as a modelling tool for risk management in agriculture.* Department of Food and Resource Economics (IFRO) University of Copenhagen Rolighedsvej 25. DK 1958 Frederiksberg Denmark.

Rollason V. and Haines P. (2012). *Outcomes from the application of ISO 31000:2009 risk management principles to coastal zone management*. Available at: https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=1&cad=rja&uact=8&ved=0CB8QFjAA&url=http%3A%2F%2Fwww.coastalconference.com%2F2011%2Fpapers2011%2FVerity%2520Rollason%2520Full%2520Paper.pdf

SAI Global, (2004). *Risk management*. AS/NZS 4360. SAI Global, 2004.

Saunders K., Lewis P. and Thornhill A. (2012). *Research Methods for Business Students*. Pearson Education Limited, Available at: https://books.google.co.ke/books?id=u4ybBgAAQBAJ

Schou C. and Hernandez S. (2014*). Information Assurance Handbook: Effective Computer Security and Risk Management Strategies*. McGraw Hill Professional.

Scutaria M., and Nagarajanb R. (2013*). Identifying Significant Edges in Graphical Models of Molecular Networks.* Available at: http://arxiv.org/pdf/1104.0896.pdf

Sennewald, C. A. (2003). *Effective Security Management* (4th ed.). Burlington: Elsevier Science.

Sekaran, U & Bougie, R.(2010). *Business Research Methods.* 5th Edition. Wiley India

Standards Association of Australia and Standards New Zealand (2013). *Handbook: Risk management guidelines-companion to AS/NZS ISO 31000:2009*. Volume 436, Issue 2013 of SAA HB

Steffes-lai D. (2014). *Approximation Methods for High Dimensional Simulation Results - Parameter Sensitivity Analysis and Propagation of Variations for Process Chains*. Logos Verlag Berlin GmbH

Stoneburner, G. et al. (2009). *Risk Management Guide for Information Technology Systems: Recommendations of the National Institute of Standards and Technology*. Retrieved on 15/1/2014, from http://csrc.nist. gov/publications/nistpubs/800-30/sp800-30.pdf

Siayor A. (2010). *Risk Management and Internal Control Systems in the Financial Sector of the Norwegian Economy*: A case study of DnB NOR ASA. Tromsø University Business School University of Tromsø.

Stern R and Arias J. (2011). Review of risk management methods. *Business Intelligence Journal.* Vol.4 No.1.

Ticehurst L. et al (2008) *Integration modeling and decision support:* A case study of the Coastal Lake Assessment and Management (CLAM) tool. Mathematics and Computers in Simulation, 78, 435-449.

Trček, D. (2006). *Managing Information Systems Security and Privacy*. Berlin: Springer.

Wang, S. and Wang, H. (2012). *Information Systems Analysis and Design*. Water Circle Boca Raton, FL. Universal-Publishers

Whitman, M. E, & Mattord, H. J. (2008). *Management of Information Security*. Boston: Course Technology Cengage Learning.

Zikmund W., Barry J., Carr C., and Griffin M. (2013). *Business Research Methods* (9th edn). Cengage Learning, South-Western.

# Appendix I: Questionnaire

## *A SURVEY ON RISK MANAGEMENT OF FINANCIAL INFORMATION SYSTEMS USING BAYESIAN NETWORKS*

I' am Ann Kibe a Doctoral student at Jomo Kenyatta University of Agriculture and technology undertaking a PhD. My Dissertation is on *Risk management of Financial Information Systems using Bayesian Networks.*

I kindly invite you to participate in this research by filling this questionnaire based on the simple instructions provided. This research is purely academic, and will be treated with utmost confidentiality.

### **Section A: General Information** (please indicate with an x or √ where it applies)

1. In which financial industry sector is your company?

| Sector | |
|---|---|
| Banking | |
| Housing finance | |
| Capital markets | |
| SACCO | |
| Microfinance | |
| Insurance | |
| Others | |

2. Which of the following best describes your title?

| Title | |
|---|---|
| CIO/Technology Director | |
| Chief Risk Officer | |
| Operational Risk Officer | |
| Information Technology Risk Officer | |
| Chief Information Security Officer | |
| Information Systems Manager | |
| Other | |

3. When in need of Financial Information systems, my company ……

| | |
|---|---|
| Tailor makes the system (internally or outsources) | |
| Acquires already developed systems | |
| Don't know | |

4. In the case of tailor made systems, is your organization actively involved in every stage of the information system development?

| | |
|---|---|
| Yes | |
| No | |
| Don't know | |

**<u>Section B:</u> The Current State of Financial Information System Risk Management** (please indicate with an x or √ where it applies)

5. Does your organization have a formal financial information system risk framework and assessment process in place?

| | |
|---|---|
| Yes | |
| No | |
| Don't know | |

6. Please indicate the extent of your agreement with the following statement: There is a common risk language that is broadly used and understood throughout and across my sector (e.g. banking, insurance).

| strongly agree | agree | Neither agree nor disagree | disagree | strongly disagree |
|---|---|---|---|---|
| | | | | |

7. Systems Managers within my organization receive sufficiently regular and robust information on risk from the company's information technology risk management teams.

| strongly agree | agree | Neither agree nor disagree | disagree | strongly disagree |
|---|---|---|---|---|
| | | | | |

8. Does your company have defined financial information system risk performance framework and metrics that are monitored and reported to management on a regular basis?

| | |
|---|---|
| Yes | |
| No | |
| Don't know | |

9. By what percentage estimate do you think your organizations' financial investment in financial information systems risk management activities will change over the next two years?

| | |
|---|---|
| Increase by less than 5% | |
| Increase by 5-25% | |
| Increase by more than 25% | |
| Decrease by less than 5% | |
| Decrease by 5% to 25% | |
| Decrease by more than 25% | |

10. How will your organizations' investment in information system risk management be allocated toward the following initiatives?

| | Strong investment | Moderate investment | Low investment | Very low investment | No investment | Don't know |
|---|---|---|---|---|---|---|
| Staffing | | | | | | |
| New technologies | | | | | | |
| Process automation | | | | | | |
| Control self-assessment | | | | | | |
| Reporting & monitoring | | | | | | |
| Alignment with corporate risk management | | | | | | |

11. Do you agree that the current information system risk management framework in financial institutions is sufficient?

| strongly agree | agree | Neither agree nor disagree | disagree | strongly disagree |
|---|---|---|---|---|
| | | | | |

## SECTION C: Analysis of Financial Information System Risks and Risk Management Programs Used by Financial Institutions.

(Please indicate with an x or √ where it applies)

12. Based on your experience and professional knowledge please indicate the risk levels in the following information systems development environments.

| **Development Cycle Risks** | **Risk level** | | | | |
|---|---|---|---|---|---|
| | Very high | high | neutral | low | very low |
| System Requirements | | | | | |
| System Design | | | | | |
| Implementation | | | | | |
| Test and Evaluation | | | | | |
| **Development Environment Risks** | | | | | |
| | Very high | high | neutral | low | very low |
| Development Process | | | | | |
| Development System | | | | | |
| Management Process | | | | | |
| Management Methods | | | | | |
| Work Environment | | | | | |
| **Programmatic Risks** | | | | | |
| | Very high | high | neutral | low | very low |
| Resources Risks | | | | | |
| Contract Risks | | | | | |
| Program Interface Risks | | | | | |

13. Which of the following standards or leading practices have you used in developing your financial information system risk framework and assessment processes? Select all that apply.

| **Standards/Practices** | |
|---|---|
| COSO ERM - Committee of Sponsoring Organizations of the Treadway Commission -Enterprise Risk Management | |
| British Standard BS-6079-3:2000 | |
| AS/NZS 3000 Risk Management Standard -Australia and New Zealand Standard | |
| COBRA - Consultative, Objective, and Bi-functional Risk Analysis | |
| IRAM - Information Risk Analysis Methodologies | |
| Canadian Risk Management Guideline CAN/CSA-Q850-97 | |
| CRAMM - CCTA Risk Analysis and Management Method | |
| NIST SP800-30 | |
| OCTAVE - Operationally Critical Threat, Asset, and Vulnerability Evaluation | |
| Other | |

14. What FIVE main challenges is your organization facing when using the current financial information risk management program?

………………………………………………………………………………
………………………………………………………………………………
………………………………………………………………………………
………………………………………………………………………………
………………………………………………………………………………
………………………………………………………………………………
……………………………

15. In your opinion what FIVE specific desirable features would you recommend to be made to the current FISRM framework to ensure it's more effective and efficient? Please list them in order of priority

………………………………………………………………………………
………………………………………………………………………………
………………………………………………………………………………
………………………………………………………………………………
………………………………………………………………………………
………………………

## **SECTION D: Application of Bayesian Networks in Financial Information System Risk Management.** (please indicate with an x or √ where it applies)

16. Do you agree that most of the risks faced by financial information systems would be better managed if they were handled throughout the SDLC?

| strongly agree | agree | Neither agree nor disagree | disagree | strongly disagree |
|---|---|---|---|---|
|  |  |  |  |  |

17. Are you aware of Bayesian Networks and their application? (If not please don't respond to the rest of the questions)

| | |
|---|---|
| Yes |  |
| No |  |
| Don't know |  |

18. Based on your knowledge and experience do you think that incorporating BN has the potential to significantly improve in FISRM program?

| Highly significant | significant | Neutral | Not significant | Very insignificant |
|---|---|---|---|---|
|  |  |  |  |  |

**Thank you for your valuable contribution!**

## Appendix II: Financial Institutions Listing

| Type Of Institution | Name Of Institution |
|---|---|
| Commercial Banks | Barclays Bank of Kenya Ltd<br>Consolidated Bank of Kenya Ltd.<br>Charterhouse Bank Ltd<br>Chase Bank (K) Ltd.<br>Citibank N.A Kenya<br>Commercial Bank of Africa Ltd.<br>Credit Bank Ltd.<br>Co-operative Bank of Kenya Ltd.<br>Development Bank of Kenya Ltd<br>Diamond Trust Bank Kenya Ltd.<br>Dubai Bank Kenya Ltd.<br>Equity Bank Ltd.<br>Ecobank Kenya Ltd<br>Equatorial Commercial Bank Ltd.<br>Family Bank Limited<br>Fidelity Commercial Bank Ltd<br>Fina Bank Ltd<br>First community Bank Limited<br>Giro Commercial Bank Ltd.<br>Guardian Bank Ltd<br>Gulf African Bank Limited<br>Habib Bank A.G Zurich<br>Habib Bank Ltd.<br>Imperial Bank Ltd<br>I & M Bank Ltd<br>Jamii Bora Bank Limited.<br>Kenya Commercial Bank Ltd<br>K-Rep Bank Ltd<br>Middle East Bank (K) Ltd<br>National Bank of Kenya Ltd<br>NIC Bank Ltd<br>Oriental Commercial Bank Ltd<br>Paramount Universal Bank Ltd<br>Prime Bank Ltd<br>Standard Chartered Bank Kenya Ltd<br>Trans-National Bank Ltd<br>UBA Kenya Bank Limited<br>Victoria Commercial Bank Ltd<br>Kenya Commercial Bank Ltd |
| Housing Finance Institutions | Housing Finance<br>National Housing Corporation |
| Mortgage finance companies | Housing Finance Ltd |

| | |
|---|---|
| SACCOs | Stima Sacco<br>KUSCCO<br>Afyasacco<br>Tembo Sacco<br>Ken tours Sacco<br>Utumishi Sacco<br>Mhasibu Sacco<br>Kenya Bankers Sacco |
| Licensed Deposit Taking Microfinance Institutions | Faulu Kenya DTM Limited<br>Kenya Women Finance Trust DTM Limited<br>SMEP Deposit Taking Microfinance Limited<br>Remu DTM Limited<br>Rafiki Deposit Taking Microfinance<br>UWEZO Deposit Taking Microfinance Limited<br>Century Deposit Taking Microfinance Limited<br>SUMAC DTM Limited<br>U&I Deposit Taking Microfinance Limited |

Table III: Financial institutions listing