LUND UNIVERSITY
School of Economics and Management

Knowledge Discovery in Government Decision Making Process

-A study of Swedish government agencies and publicly owned companies-

---

Master Thesis, 15 Credits, Department of Informatics
Presented: June, 2009

Authors: Imad Bani-Hani
Kelvin Kiritta

Supervisor: Odd T. Steen
Examiners: Agneta Olerup, Eric Wallin

Lund University
Informatics

Knowledge Discovery in Government Decision Making Process

-A study of Swedish government agencies and publicly owned companies-

Master thesis, presented June 2009
Size: 83 pages
Supervisor: Odd T. Steen

Abstract

The computerization of government transactions results in dramatic increase in growth of data across government agencies. This data is presented, generated, maintained and used independently in each government department and agencies which leads to poor decisions and isolated planning. This environment hinders effective access to data, reuse and aggregation, so that decision makers can get useful information at the right time and in the right format, which can guide them in decision making process. Given the turbulent environment and complex decision situations that the government decision is characterized, there is need to extract useful knowledge from these data to enable decision makers to access useful and sufficient knowledge. Therefore, there is need for effective techniques and tools to integrate the data from different sources into a consistent format, which permit the decision maker to access a cleansed and consistent data, and also derives useful knowledge. Knowledge discovery in databases (KDD) is such techniques that extracts and identifies useful Knowledge from huge data sets; hence assist decision makers in the process of effective decision making. This study explore the perceived usefulness of the knowledge discovered through the knowledge discovery in database (KDD) in the decision making process of government agencies. The empirical findings collected from Sweden Government Agencies have attempted to explore the perceived usefulness of knowledge on the decision making process therefore add to the understanding of knowledge discovery and use of such knowledge by decision makers in the government.

Keywords:
KDD, Data Mining, Data Warehouse, Decision Making, Government, eGovernment, Perceived Usefulness

ACKNOWLEDGMENT

We are very thankful to God for his mercy, support and opportunity to excel in our education and knowledge during the study period.

We are very grateful to our family who supported us during this period. We extend our deepest appreciation to our Supervisor Dr. Odd Steen for his invaluable contribution and guidance which enable us to carry out this study.

We feel greatly indebted to our four interviewees for their valuable insights and contribution to this study.

Thank you.

Imad Bani-Hani & Kelvin Kiritta

Table of Contents

List of Figures:

List of Tables:

1. Introduction

1.1 Background

The decision-making process in governmental organizations reveals the complexity of many of the decisions being made (Winterman et al 1998). The types of decision made in government agencies include those associated to management, research, funding, and policy making and advisory.

The government decision making as it relates to rapid changing and turbulent environment is often characterized by uncertainty. Moreover, the decision situations are complex and ill structured (uncertainty) which cannot be treated by normal procedures (Bots et al 2000). The insufficient information available to address these decision situations inhibit decision makers to formulate an appropriate understanding deem necessary for complex situations (Radford et al 1993).

The views put forth by Simon (1977 according to Mintzberg 1977) on the decision making process in complex organization includes government and business organization suggested a model which is based on concepts of bounded rationality and satisfaction. The model was convenient to study the decisions which are made under pressure and turbulent environment. Simon's model emphasized the need to use intuitive, experience and systematic procedures when dealing with non programmed and ill defined decisions (Daft 2004; Vasu et al 1998). Simon (1977 according to Mintzberg 1977) stated that the decision making process consists of four stages: intelligence, design, choice and implementation which is considered the most general and yet complete model for rational decision making. The flow between these stages is interchangeable, depending on the decision maker satisfaction (Turban et.al 2007). This four step model is highly influenced by the data and information used in each phase. The overall process of decision making depends much on the quality of available data or information. However, in government there are still problems in identifying the correct data which assist the decision making process and as result of this many decisions are based on poor data.

Hoss (2001) argued that the Government today is inducted by rapidly increasing volumes of data. The sources of this data explosion lie to the increasingly data sources largely caused by the adoption of government initiatives towards eGovernment (Norris et al 2001). The electronic government refers to the effective use of the information and communication technology by government agencies in order to increase their working effectiveness and efficiency in aspects of service provision to the citizens, business and between the government agencies (Chen 2008).

It has further been pointed out that every large organization, in business or government, has large quantities of data that have been collected over a period of years. There have been initiatives from organization to make the historical data useful, this includes starting off projects to develop and implement data warehouse to facilitate the massive storage of data (historical and new data) and above all provide access to the data at a right time and right format thereby improving decision making process (Bieber 1998).

Harper (2004) considers data warehousing as architecture which organize data from different sources into a single repository of information. The data warehouse integration characteristic ensure the consistency and quality of information, this environment provides better means to

access the information which can be in form of reports and queries. Government can discover useful insight and trends which can help to improve policies decisions and service delivery. Also, the Data warehouse ability to handle historical data provides opportunities to access the variation of data which helps to detect trends and guide forecasting and planning activities. Harper (2004) emphasized that the data warehouse architecture leads to potential information which improves the quality of government services thus complement to the vision of eGovernment.

Furthermore, the Data warehouse represents an enterprise wide data collection, which is central and defines a common basis for several enterprise systems accessing it. From the stored data new knowledge can be derived or discovered using technologies such as Knowledge Discovery in Databases (KDD). This technology consists of several steps starting from data set selection from the data sources to the new knowledge creation. Accordingly, Fayyad et al (1996a) describes Knowledge Discovery in Databases (KDD) as "the nontrivial Process of identifying valid, novel, potentially useful, and ultimately understandable patterns in data". The non-trivial process mentioned in the definition implies that, all processes involved in discovering knowledge are not directly computed therefore search or inference procedures should be used. The knowledge discovered should be novel and valid which lead to some benefits to the user (Fayyad et al 1996a). This knowledge discovered derived from government data sources is crucial for the decision makers as it assists them in better understanding various decision situation thereby achieving improved decisions.

Therefore, this thesis is interested to gain understanding of the knowledge discovery process through KDD and how such knowledge is used to assist the decision making process in the government agencies.

1.2 Problem Area

There have been dramatic increases in growth of data across various areas including business and government (Fayyad et al 1996a; Singh 1998). In the government, the source of this data is mainly associated with human resources, projects, plans, decisions, reports. Moreover, the computerization of government transactions also contributes to the vast amount of data in government agencies, these data exists in different formats and in various systems, spread across a number of agencies (Hovy 2008). Furthermore, this data is presented, generated, maintained and used independently in each government department and agencies. This leads to poor decisions and isolated planning (Prabhu 2006). This environment hinder effective access to data, reuse and aggregation, so that decision makers can get useful information at the right time and in the right format, which can guide them in decision making process (Turban et al 2007).

From these data sets, the organization can extract useful information, which facilitates the making of appropriate decisions (Ogut et al 2008). Therefore, there is a need for effective techniques and tools to integrate the data from different sources into a consistent format, which permit the decision maker to access a cleansed and consistent data, and also derives useful knowledge. Knowledge Discovery in Databases (KDD) is such techniques that extracts and identifies useful Knowledge from huge data sets; hence assist decision makers in the process of effective decision making (Fayyad et al 1996a).

The Knowledge Discovery in Database (KDD) can intelligently and automatically transform the data and information into useful knowledge. This knowledge and information created is considered the foundation and basis for the process of decision making (Sharma 2004), therefore KDD empowers the decision makers, with information which assists them in increasing the effectiveness of the decisions made.

Moreover, the government decision making in rapidly changing and turbulent environment is often characterized by uncertainty, due to the lack of useful and sufficient information, this leads to poor decision making. Thus, this research aims to explore the usefulness of the knowledge discovery through the KDD process in assisting the decision making in the governmental agencies.

1.3 Research Questions

In this research study, we will answer the following question;

- What is the perceived usefulness of KDD in the Government decision making process?

1.4 Research Purpose

The purpose of the study is to describe the perceived usefulness of the knowledge discovered through the knowledge discovery in database (KDD) in the decision making process of government agencies. This study explore the effect of knowledge on the decision making process therefore add to the understanding of knowledge discovery and use of such knowledge by decision makers.

1.5 Delimitation

Knowledge can be extracted through different types of tools and techniques such as Online Analytical Processing tools (OLAP), Executive Information Systems (EIS), Reporting and Querying tools and KDD. Also, decision making follows different models which include pure rationality model, incremental model, bounded rationality model (Simon model). This research tends to explore the use of KDD as a knowledge discovery tool to support government decision making process which follows Simon models. This research does not explore the technical aspect of the KDD knowledge extraction processes such as data-set selection methods or data mining algorithms but it focuses on the use of the knowledge discovered through the KDD process to aid in the Government decision making process.

2. Government Decision Making

2.1 Overview of Decision

There are number of definitions of the term decision from various scholars. Simon (1960 cited in Holsapple 2008) defines decision as being a choice where a choice presents a course of action to be taken for a given situation. (Fishburn 1964 cited in Holsapple 2008) consider decision as the choice of a strategy for action. Decision can also be defined as a choice leading to a certain objective (Churchman 1968 cited in Holsapple 2008).

The above definitions suggest a relationship between decision and course of action in a process to achieve a certain objective. This observed relationship can be referred to as decision making. Knowledge is fundamental to decision making as people use knowledge available to them to make decisions about actions that shape themselves, organizations in which they participate, and the world in which they live (Holsapple, 2008). Knowledge enables people to choose the best course of actions among available alternatives in accomplishing various decision tasks thereby facilitating decisions.

2.2 Definition of Decision Making

A basic part of the organization survival is planning; Simon (1977 according to Mintzberg 1977) stated that the decision making and planning are similar to the process of management to some extent. Planning consists of a series of decisions that addresses “What should be done, When, Where, Why, How, by Whom”. Turban et.al (2007) defined decision making as the process of choosing between different courses of actions in order to attain a specific goal.

We can notice that both scholars are describing the decision making as a process that includes different steps to reach a final decision. These decisions are governed by different characteristics, frequent changes that appends in the decision-making environment leading to uncertain state of the decision made which affect the decision quality and impose pressure on the decision maker. Moreover the information used in the decision making could vary from insufficient information to too much information (information overload) which makes it difficult to know when to stop collecting information (Turban et.al, 2007).

2.3 Government Decisions

The decision-making in government organizations reveals the complexity of many of the decisions being made (Winterman et al 1998). While the decision making process framework is applied to all agencies, the way the decisions are implemented can vary significantly based on the primary function of the agency and the decision nature.

2.3.1 Decision Type

Decision Types made in government agencies include those associated to management, research, funding, and implementation of policy and advisory.

The *management* decisions are the decisions undertaken by managers/executive in an attempt to accomplish the organization objectives. This involves decisions which are based on organization strategic responsibilities, controlling and allocation of organization resources (Turban et al, 2007). Management decisions in government may include project scheduling, budget analysis and preparation, investment decisions, negotiating recruitment issues and make or buy decisions.

The government also involves with the *research and development* (R&D) decisions, mostly this kind of decisions deal with what to research on relevant areas in the organizations so as to improve productivity and efficiency. The aim of research in government is to improve current operations hence improve future performance. R&D decision may include the study on a specific organization process or segment; revise the current organization process, commissioning of research, acquisition to get technology and market access. Winterman et al (1998) claimed that research decisions in the government are often associated with policy of funding. The study carried out by Matheson (1998) reveals that decisions which involve Research and Development are difficult as they are usually made in the faces of many uncertainties. He further justified his argument by stating that the R&D process is inherently uncertain meaning that without uncertainty there would be no R&D, this implies that no one knows when R&D will succeed and the level of that success.

Funding is one of the decision types being made in the government, funding decisions presents choice of actions of injecting funds (money) in a project, investment and /or research initiatives. Those funds can be allocated for either short term or long term purposes. In government agencies, to make financial decisions over a certain value may differ according to the nature of the investment, the overall project environment, or political sensitivity. Winterman et al (1998) stated that, the grade of the decision maker is not necessarily the main indicator of financial decision-making authority, and financial value alone cannot be taken as an indication of the significance of the decision (ibid.). Therefore, government agencies need to make good funding decisions about whom and what to fund in order to achieve the best value of it.

Policy making and implementation is among the roles of government agencies (Verschuere 2009; Winterman 1999). To develop and implement policies, the agencies has to follow a policy making process. Nabukenya et al (2009) regard policy as a proposed course of action of a person, group, or government within a given environment providing obstacles and opportunities which the policy was proposed to utilize and overcome in an effort to reach a goal or realize an objective or a purpose. Thus the definition of policy is oriented toward accomplishment of some purpose or goal. Also, Sabatier (1999) describes the process of policy making to include the manner in which problems get conceptualized and are brought to a governing body in order to be resolved.

Government agencies are also charged with the responsibility of *advisory* to government on the specific areas on which its functions are derived Winterman (1999). The government agencies may advise the governments (ministers) on any matter relating to its functions, powers, and duties. This might involves the formulation of new regulations, laws or procedures. Usually, series of decisions are involved in order to fulfill the advisory function that underpins the decision making process in all the activities mentioned.

2.3.2 Decision Situations

Simon (1977 according to Mintzberg 1977) argues that decision making process involves evaluating and comparing alternatives along with the prediction of the future outcome of every proposed alternative. Also, Radford et al (1993) described that the most essential part of decision making is in the formulation of alternative courses of action to meet the situation under consideration and making the choices among selected best alternatives after an evaluation of their effectiveness in achieving the decision makers' objectives.

Moreover, every decision situation exists in an environment. This environment consists of a set of circumstances and conditions that affect the manner in which the decision making problem can be resolved (Hipel at al 1993). Radford et al (1993) classified decision situations into programmed (well-structured) and ill-structured (complex decision situations).

Programmed or well-structured decision situations can be approached effectively by following rules and patterns of behavior that have been established as a result of previous experience. Decision makers under this category assume that complete knowledge is available for them to identify the outcome of each course of action (Turban et al 2007).

Ill-structured or complex decision situations are new and unique to decision makers in one or more of their aspects. These situations cannot be treated by a well-established procedure. Decision makers involved in complex decision situation can undertake lists of objectives at the same time and those objectives may be directly or indirectly related to a particular situation.

Some or all of the objectives of one participant may be in conflict with those of one or more of the others. It is often that the information available to the decision makers is insufficient to allow formulating a complete and exhaustive description of a complex situation. Thus decision making under complex decision situation is more difficult because there is insufficient information which may guide decision makers to base their appreciation of the circumstances (Turban et al 2007). As a result of this each decision maker have a different outcome for the same course of action. Resolving the complex decision situations requires systematic discussion between the decision makers. This process involves negotiation and bargaining where each decision makers giving his/her decision on the given decision problem, final decision outcome is achieved on consensus basis (Radford et al 1993). In addition, decision makers attempt to obtain adequate information so that the problem can be treated as decision under certainty (Turban et al 2007).

The ill structured problem needs more effort than the structured ones. Different models have been introduced by scholars to address this kind of situation.

2.3.3 Characteristics of Government Decision Making

The major role of government is decision making and service delivery (Misra 2007). Nutt (2006) differentiate the decision making of private and public organization based on the role that both play in the society. In private organizations decision activities aim to get profit and satisfy stakeholders, while in public organizations decision activities aim to deliver the best service to the public (Nutt 2006; Vasu et al 1998). As a consequence of these differences, the following characteristics of government decisions are depicted:

- Decision involves many people with diverse interests, decision makers can collaborate with other organization in attempt to accomplish a task which aims to provide service to public. There is transparency of how they execute decision as they are not working on competition environment (Nutt, 2006; Bots et al 2000).
- Many decisions are made through consensus due to the fact that decision makers have limited mandate thus cannot take individual decisions (Nutt 2006).
- The need of relevant data and profound analysis is crucial for a good decision (Turban et al, 2007), since most of the decision situations in a turbulent and changing environment are ill-structured which needs special treatment.
- Complex decision situation needs computerized systems that have the ability to access the data and perform the analysis to assist the decision maker (Turban et al 2007). These decisions usually include many different interest of the society, and since it is hard to include all the society in the decision therefore conflict are likely to appear. Also setting criterions for evaluating the decisions is hard since it has a large variety of qualitative and quantitative criterions which exhibit difficult values to specify such as quality of life and safety (Bots et al 2000).
- Previous experiences and problem results cannot always predict future results. As a matter of fact a considerable number of decisions are concerned with future planning for several decades, such decisions includes the infrastructures decisions (road and transport) (Bots et al 2000).
- Thinking about the problem leads to request data and information to help in modeling the problem, which is considered a part of the decision making process.

2.4 Decisions Making Process

As previously mentioned several decision making processes had been introduced to assist the decision maker in understanding the road-map of solving a complex decision situations. Pure rationality model, incremental model, bounded rationality model had been introduced by scholars to address the steps employed to reach a decision for a certain situation. In the next section we will enlighten the bounded rationality model which had been introduced by Simon (1977 according to Mintzberg 1977) and had been accepted by different scholars Vasu et al (1998).

2.4.1 Simon's Decision Making Process/Model

The decision making process consists of several phases; each phase contains internal processes (Simon 1977 according to Mintzberg 1977). These phases are intelligence, design, choice and implementation. The phases start with the intelligence and end with the implementation, with the possibility to return to any previous step. It is important to consider the complexity of the decisions situation which is governed by the variables that affect the decision situation therefore the modeling of the processes is an important and essential part. The decisions are influenced by different variables such as the decision, uncontrollable, intermediate and result variables.

The *Decision Variables* are controlled by the decision makers which are used to describe the different course of action, for instance, the number of people employed in a project or the amount of money invested in a project. Moreover, the *Uncontrollable Variables* are the variables

that affect the decision making and the final outcome but they are not controlled by the decision maker, for instance the economical status of the country or even the unexpected new regulation from the tax authority. The *Intermediate Variables* are important in the final outcome because of the indirect impact they have which can be controlled somehow, for instance the salary of the employees define their satisfaction hence affect the productivity. Finally the *Outcome Variables* are the variables that measure the effectiveness of the output by comparing it to the defined goals. (Turban et al 2007).

These variables combined together shape the decision situation which lead in having a simple or complex decision situation hence affect the decision making phases which will be described next.

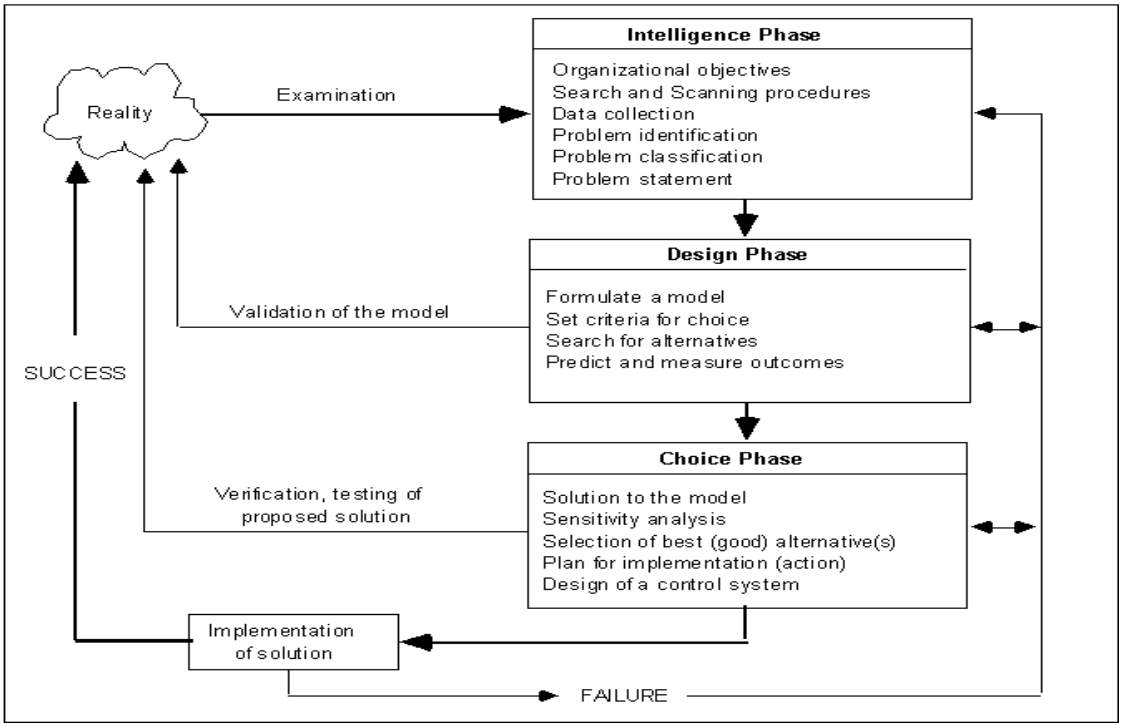

Figure 2.1: *The decision making/modeling process (adopted from Turban et al. 2007)*

❖ *The intelligence phase*

This phase is concerned with identifying the problems or opportunities available and scanning the environment for relevant knowledge from internal and external sources (Turban et al. 2007; Holsapple 2008). The decision maker uses information systems to aid him in this phase. KDD is been used to discover relationship between different variables and factors (Turban et al 2007) hence assist in the problem/opportunity identification.

The decision maker begins with checking whether the organizational goals are met or not and what is the actual situation. The dissatisfaction of the actual state compared to the desired goals is the problem identification and also marks the beginning of the intelligence phase.

Furthermore, the decision maker explores the problem to identify its magnitude, complexity and symptoms. After identifying the problem, the decision maker collects relevant data for further analysis to assist him/her in exploring the details of the problem such as the location, severity and significance (Turban et al 2007) which permit the decision maker to allocate resources for further exploration of the problem/opportunity (Pyle 1999 cited in Ogut et al 2008).

During the data collection, the decision maker could face several problems such as the cost of collecting the needed data, too much data or even there is no data to address the problem situation.

Simon (1977 according to Mintzberg 1977) reiterated that three types of problems situations exist; structured, semi-structured and unstructured. The structured problems are the problems that repeat themselves and have a standard solution; however the unstructured problems have a fuzzy nature and complex structure, hence have no defined solution which requires Simon's Decision making model to address the problem. Semi-structured problem are a mixture of structured and unstructured problems (Simon 1977 according to Mintzberg 1977).

- ***The design phase***

The design phase is concerned with the development and the analysis of possible courses of actions to solve the problem (Turban et al. 2007; Holsapple 2008), to do so the decision maker has to design a model which is based on the decision variables that affect the decision situation then create a criterion which is also called principle of choice, to evaluate and describe the acceptability of the solution approach in respect to the decision situation. The KDD enables the decision maker to understand hidden relationship between variables which is used to develop the models (Turban et al 2007) hence lead in developing and forecasting different courses of actions.

Different principle of choices had been introduced by scholars however the most commonly used are *Normative* and *Descriptive* Models. The normative models aims to identify the best course of action possible by examining all available courses of action and illustrating the reason of choice; on the other hand the descriptive models are used to describe the situation as they believe to be, this model is used to explore a set of alternatives but not all of them therefore the result is somehow satisfactory and not optimal (Turban et al 2007).

- ***The choice phase***

The decision on a course of action is made in this phase; it involves the search for the appropriate course of action, followed by the evaluation and recommendation of the defined course of action which is considered the solution of the problem. The decision maker exercises his authority and decides to choose a solution based on the acquired knowledge (Holsapple 2008).

In some cases the choice is complicated and needs to return to a previous phase, for instance if none of the alternatives are palatable, the decision maker return to the design phase, or even the decision maker needs to return to the intelligence phase due to the fast changing environment and the need to acquire extra knowledge (Holsapple 2008).

❖ ***The implementation phase***

As a final stage of solving a problem, the course of action recommended by the decision maker is put in action. The implementation of a solution is a complicated phase because it contains different issues that affect its success such as the degree of support of top management, the resistance to change and many other factors that have to be considered (Turban et al 2007).

2.4.2 Decisions Making Process in Government

The decision making process in government can be explained through the Simon's general decision making model. This model describes the complex situations faced in an organization through the decision making framework (Vasu et al 1998). The government decision making complexity resides in rapid changing and turbulent environment often characterized by uncertainty (Winterman et al 1999; Khorshid 2004; Turban et al 2007). Therefore, Simon's model is appropriate to reflect the complex decision making process in the government agencies.

As previously discussed, the decision making process begins with the problem identification and understanding by scanning the environment for relevant information. The government agencies have a potential source of data and information constituted of the integration, interconnection of the agencies and the eGovernment services provided to the citizen and business partners through the internet platform. In figure 2.2 Misra (2007) introduced a common sense model that illustrates the entities of government decision making process and their relationship.

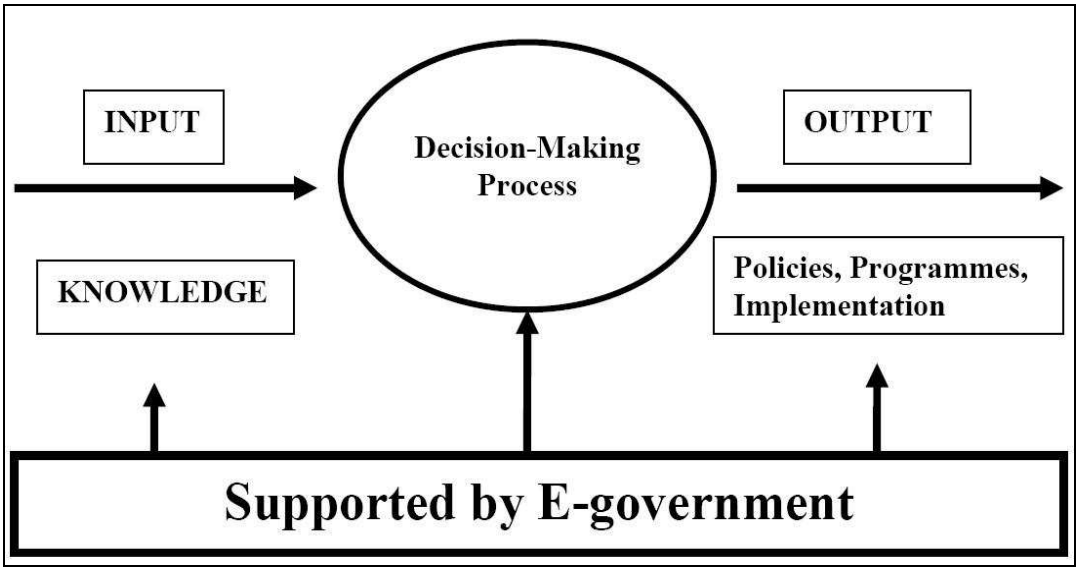

Figure 2.2: *The decision-making process in government supported by eGovernment (adopted from Misra, 2007)*

The decision maker start by exploring the available information and knowledge to understand the problem situation, this task is mainly performed with the help of the data provided through the

eGovernment infrastructure and the available analysis tools. The eGovernment infrastructure provides an important support to the decision making process in the government.

Literatures reveal that the major function of government is decision making which guide them to provide better and improved services to citizens (Misra 2007). However, for the government to achieve an effective decision making, it requires that the decision makers get access to the appropriate and useful knowledge to guide them to evaluate and implement appropriate decisions (Ogut et al 2008). In this regards, eGovernment infrastructure provides the best available technologies and tools which enable government to better create, manage and leverage knowledge which is crucial for efficient decision making (Misra 2007). These technologies help the government agencies to cope with the problem associated with data explosion and integration thereby offer the mechanism to provide knowledge at a right time and right format to facilitate timely and better decisions.

Policy making, management, research and development, funding have been identified as decision situations which provides a base for decision making in government (Nabukenya et al 2009; Winterman 1999). In order to achieve the best decision making, the mentioned decision situations has to be examined and evaluated accordingly to see if they meet organization objectives and goals (Turban et al 2007).

Most of these situations are complex and ill structured (uncertainty) which cannot be treated by normal procedures (Bots et al 2000).

2.4.3 Example of Decision Making Process in the Government

This section illustrate how a policy making is being formulated based on Simon's model for decision making process; Dunn (1994) presented a policy decision making process which contains five stages:

In the stage of *Intelligence*, *Agenda setting* establishes priorities among the issues of public concern that require policy action or the change of a previous policy (ibid).
In the agenda setting, the decision maker first formulates a procedure to identify the information sources (which could be from other agencies) which will be used for scanning and search the data to identify the issues of public concern. By using these procedures, the data collection will be performed and an adequate knowledge will be available to classify, decompose and identify the problem owners.

The classification is important to organize the issues identified based on their structurdness, in this step the decision maker categorize the issues of public concern in structured, semi-structured and unstructured problems. The structured issues have a systematic solution therefore the solution will not be complex and will be solved, however the semi-structured and unstructured problems are more complicated and need extra steps and efforts.

Having ill structured situations often require decision makers to decompose the problems. The decomposition is useful in semi-structured issues since it involves structured and unstructured problem; therefore decomposing the problem can produce structured problems which will be easily solved and unstructured which needs further process therefore the decomposition can ease the understanding and identification of solutions (Turban et al 2007) .

Also, it is important to identify the problem owner in order to assign the responsibilities to the right department to deal with the problem, in other words assigning authority for the problem solver (Turban et al 2007).

Finally after performing all the steps of the intelligence phase, a considerable amount of knowledge will be gathered, the issue of the public concern will be well understood, hence lead to formulate the problem statement.

The second decision phase, *design*, involves policy analysis which aims at better understanding of the public issue (problem statement) on the agenda; the problem is formulated and alternative policies are created to solve it (ibid). To do this, the facts are clarified and the interests and objectives of citizens and stakeholders are considered. The process for formulating alternatives is guided by selected choice which acts as benchmarking criteria for desired policy

This phase deliberately studies the problems and opportunities that the desired policy brings to the organization. The evaluation of the problems and opportunities must ensure that good policies alternatives are produced to support organization planned objectives. Yager (2008) affirmed that a fundamental difficulty that arises when making decisions involving alternatives with uncertain outcomes; is the comparison of the alternatives. This is due to the fact that the diversity and complexity of these alternatives makes their direct comparison almost impossible thus making it difficult to predict and measure outcomes for the alternatives. Yager (2008) suggested a risk modeling solution to policy decision making which models the uncertainty associated with a course of action.

Turban et al (2007) advocated for the need to measure the outcomes for each alternative against the goals. This measures the degree of success for the proposed solutions, in other words it measures the degree of resemblance between the policies formulated and the required ones.

We can learn that the need for modeling the decision situations comes as result of the inadequate knowledge of decision makers which could help them to manage the complexity of unstructured challenges. Nabukenya et al (2009) emphasized those actors in policy making need to have adequate knowledge to understand the dynamics of a particular problem and develop options for action.

The *choice* phase is concerned with the *Policy decision* which rely on the previous analysis, hence a final decision is made and the chosen policy is fully specified (Dunn, 1994).

After the analysis of the policy formulation from previous phase the decision maker exercise his authority and chose to implement the new policy based on the his satisfaction of the decision variables . It is possible that the previous phase did not provide enough and convincible choices or unexpected changes appeared which lead to return to any previous phases (Holsapple 2008).

At the *implementation* stage, the policy chosen from the previous phase is ready to be implemented and put in practice, therefore it is essential to use necessary public resources and regulations or even create them to make the policy operative (Dunn 1994).

Monitoring is the final stage, *Policy assessment*, which the responsible people track and check if the policy developed is being adopted by the targeted agencies (parliament, governmental agencies or even the court) (Dunn, 1994).

2.5 Quality of Decision Making (efficiency, effectiveness and legitimacy)

To analyze the measures which access the decision making process in the public, the quality of decision made has been identified as important measure (Nabukenya et al 2009; Bots et al 2000). Accordingly, various criteria have been derived from literature which measures the quality of public decision making; this includes effectiveness, efficiency, and legitimacy (Dror 1997; Bots et al 2000).

2.5.1 Efficiency

Bots et al (2000) defined efficiency as the ratio of outcome over effort. In the process of decision making measuring efficiency, effort is considered as time spent on decision making. Time in question can be time period utilized to reach a decision or the total number of hours spent of the decision process thus this dimension (time) of decision making quality can be measured (Bots et al 2000).

2.5.2 Effectiveness

Effectiveness is the measure of goal attainment. (Veld 1987 cited in Nabukenya 2009) defines effectiveness as the real result compared to the intended result, specified in the design. This definition as it relates to the decision making realm implies the ability of resulting decisions to address the defined decision situation or problems. This also indicates that to achieve the effective decision making, the result decisions should be able to meet the stated decision goals (Nabukenya et al 2009). In addition, Huber (1986) stated that the timeliness of organization decisions should also be accounted for effective criterion to maximize the quality of decision making process. Timeliness in this context refers to the ability of the organization to provide the decision responses at a given time based on the assigned decision task or situation. The quality of goals achieved is however a subjective concept, which indicates that decision actors might have varying meaning to the effectiveness of the goals achieved, attributed to their varying interest and perception on intended goals which necessitate for the need to define clearly the goals for decision making process in order to able to measure the effectiveness of decision making (Nabukenya et al 2009; Bots et al 2000).

2.5.3 Legitimacy

Legitimacy is also considered the measure for the quality of decision making in public. Bots et al (2000) relates legitimacy to the effectiveness and efficiency which also forms the basis for his definition. He further clarified legitimacy into judicial and administrative interpretations. Judicial interpretation involves examination of the decision process against the provisioned laws and regulations, while administrative interpretation involves examining the public consideration and support for decision outcome (Bots et al 2000). In the view of complexity, legitimacy has seen not only determinant for quality but also for the effectiveness.

The quality of decisions is attributed to the effective understanding and use of the information available. Because of the actual explosion of digital data caused by the intense use of the ICT which affect the decisions made in the government agencies, it is important to explore the problem dimensions and identify possible solution.

2.6 Data Explosion in Government

There has been increasing growth of data and information in the government. The advancement and development of ICT allows the collection of data from different sources leading to enormous accumulation of data across government agencies (Alshawi et al, 2003; Hovy 2008; Fayyad et al 1996a). As consequences of this, the government has employed ICT to automate their transaction in order to provide efficient service and support the decision making process (Hovy, 2008).

The whole process in which government employ ICT to automate their transaction is much connected to the eGovernment initiatives. Literatures confirm that adoption of government initiatives towards eGovernment leads to the explosion of data (Norris et al 2001). In eGovernment, the governments have deployed web-portals, wiki and online information systems which allow them to provide better service to citizens and also enable them to interact with other stakeholders, all of these ICT technologies have resulted to the increasingly electronic transactions (Layne & Lee 2001).

Fachausschuss (2000, p.3 cited in Codagnone et al 2007) defined eGovernment as "the implementation of processes of public participation, decision-making, and service provision in politics, government and administration with an intense usage of ICT". This definition contains several numbers of areas that the eGovernment addresses in order to improve the quality of services provided and the management of the agencies. One of the important aspects of the eGovernment is the decision making, by interconnecting the agencies together the decision maker has the ability of using information from different agencies to support the decision situation encountered.

Also McClure (2000) defined eGovernment as the usage of technology in government activities and specially the web-based internet applications to effectively deliver services to citizens, business partners and other governmental agencies hence create a better interaction with the citizen, increase the performance of services and decrease cost of activities. Both definitions agree on the importance of ICT in governmental activities to efficiently utilize the capabilities of service provision of the government. Moreover McClure (2000) stressed the need to utilize the web-based application to fulfill the goals, as the internet is actually the communication media of different agencies, citizen and private organization. In addition both scholars pointed out that the ICT supports different activities of the government which includes social, political and public administration which constitutes a focus of several science disciplines.

Therefore Wimmer (2002) stated that eGovernment is an interdisciplinary concept that integrate several disciplines such as Social and human sciences, Political, strategic, democracy and legal sciences, Information and knowledge research sciences, Organizational and economic sciences, Computer sciences. The integration of all the mentioned science disciplines which are supported by the ICT infrastructure caused an explosion of data generation.

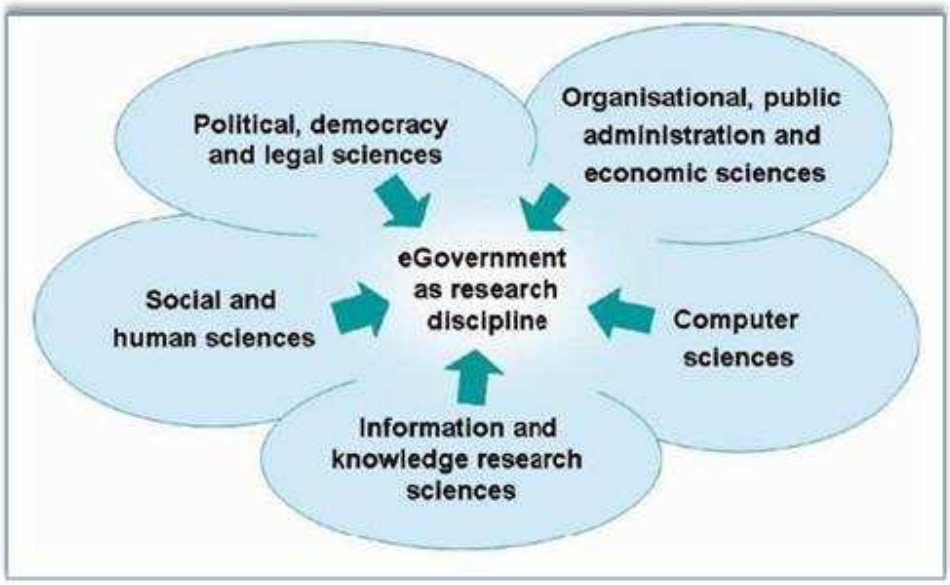

Figure 2.3: *Multidisciplinary and integrative role of eGovernment as a research discipline (adopted from Wimmer (2002)*

Due to the large amount of information that is collected from the public agencies activities (Hovy 2008), the need to research an effective way to manage this information is important. The field of artificial intelligence and computer science developed some efficient techniques to address the problem of information over-load which affect significantly the governmental agency activities including the decision making. Some of these techniques address the need of intelligently and efficiently retrieving information from data storage such as intelligent search engine, other techniques uses special data mining (KDD) technologies to extract knowledge from large data repository such as data warehouse (Wimmer 2007). These techniques assist humans to overcome their cognitive limits by extracting useful information from a large amount of raw data. The knowledge which is created through KDD provides useful information to the decision makers which guide them through evaluation and selection of appropriate actions and alternatives thus improving decision making process. In the next chapter we explain the KDD process and how such technology assists the decision maker in his decisions.

3. Knowledge Discovery in Database (KDD)

3.1 Knowledge Discovery

The advancement of the technology and the fast development of the information age have a significant effect on the amount of data collected (Alshawi et al 2003).

The fierce competition in the market demand that the strategies employed to be agile and meet the external changes effectively, to do so the analysis of the data gathered is crucial (Turban et al 2007). Monthly, Quarterly and annual reports are generated to be used in shaping the organizational strategies however with the large amount of data available this process of data analysis and report generation is a burden and create an information overload (Turban et al., 2007). It is important to use Information system equipped with tools that reduce the amount of information presented to assist the decision maker in his decisions, and even reveal and discover information that the normal data analysts could not find due to the cognitive limitation capabilities of humans (Fayyad et al 1996a).

3.2 Knowledge Discovery in Database (KDD)

The KDD is the outcome of the union between different research fields such as pattern recognition, statistics, expert system, etc., hence the general idea about KDD is the processing of information in order to produce high level knowledge from low level of data. Knowledge Discovery in Databases can be defined as "the nontrivial Process of identifying valid, novel, potentially useful, and ultimately understandable patterns in data" (Fayyad et al 1996a).

The non-trivial process mentioned in the definition implies that, all process involved in discovering knowledge are not directly computed therefore search or inference procedures should be used. The knowledge discovered should be novel and valid which lead to some benefits to the user (Fayyad et al 1996a). Moreover, Fayyad's definition point to four characteristics of the KDD output which are illustrated below:

1. Valid, refers to the usefulness or relation of the knowledge discover for the task in hand.
2. Novel, refers to the non-obvious and hidden knowledge that the human cannot discover manually.
3. Understandable, refers to the clarity of the knowledge discovered that a human can understand.
4. Potentially useful, refers to that the knowledge discovered is useful somehow to the user.

3.3 KDD process

The process of extracting knowledge from low level data source includes several iterative steps together with the user interaction (Fayyad et al 1996b).

Before starting the KDD process it is important to develop an understanding about the application of the targeted domain and the knowledge available to have a clear view of the KDD process goals, and this is considered step *one*.

The *second* step is concerned with the data-set selection from the sources where the data mining algorithm will be employed. This data-set could be chosen from a normal database or a data warehouse. Scholars agreed that the use of data warehouse increase the chance of having efficient outcome since the data mining will be applied on cleansed and integrated data (Fayyad et al 1996b).

Third, is to pre-process and clean the data-set from noises or even identifying a strategy to fill empty fields.

Fourth is the Data reduction and projection which is concerned with defining useful features in the dataset that represent the goal of the KDD process, usually for the data reduction; dimensionality reduction and transformation method are used to reduce the number of variables (Fayyad et al 1996b).

The *fifth* step which is the mining part of the KDD process includes three sub-steps which had been identified by Fayyad as fifth, sixth and seventh steps. First sub-step is deciding on a data mining method based on the understanding of the application domain and business needs. Several data mining methods are used in the KDD process such as summarization, regression, classification and clustering (Fayyad et al 1996b) which will be illustrated in the next section of this chapter. The *second* sub-step is concerned with the choice of the data mining algorithm and method of selection to identify the patterns which includes two parts a) identifying the model and the parameters used in the pattern identification, b) matching the method of data mining to the general criteria of the KDD process. In the last sub-step, the data mining algorithm search for pattern of interest in the processed dataset using some well known algorithms such as classification rule, artificial neural network and clustering. This sub-step is considered the heart and the core process of the KDD since the discovery is performed in this step.

Sixth, after identifying patterns from the previous step the result will be interpreted with the possibility of going back to any previous phase; also it could include a visual interpretation of the extracted patterns (Fayyad et al 1996b). The interpretation and evaluation process of the knowledge extracted by data mining involves domain experts; they are responsible to evaluate and determine whether the knowledge extracted is useful or not (Mitra et al, 2003)

Finally, the patterns discovered are considered new knowledge and are ready to be used whether in a report or to be integrated in another system or even to be stored in knowledge management system (Fayyad et al 1996b).

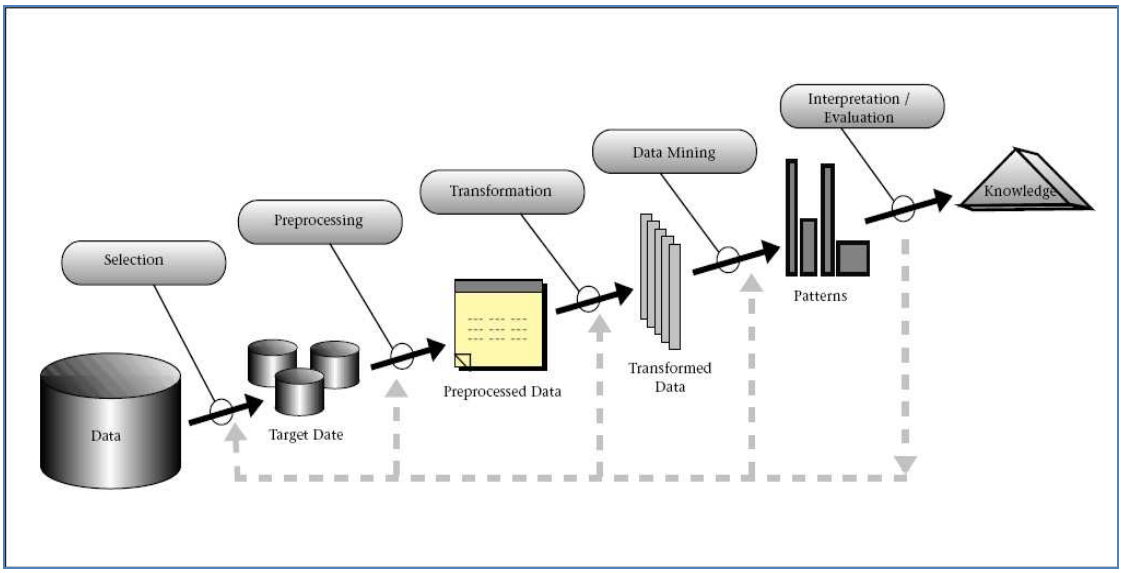

Figure 3.1: *An overview of the steps that compose the KDD process (adopted from Fayyad et al., 1996b).*

All these steps are important and highly iterative depending on the need of the application domain however the most important part and which is considered the heart of the KDD is the seventh step which include applying the data mining algorithms (Fayyad et al 1996b).

Since the effective KDD process depends much on the quality of data it extracts, in the next section we explain the concept of data warehouse as it is applies to the process of KDD. Data warehouse provides the staging area for effective KDD process and offer environment for KDD to extract useful information as it ensures the quality of data in the repository. Moreover, the data warehouse ensures the proper integration of huge amount of information or data, organize, cleanse and present them in a unified manner thus increases the efficiency and effectiveness of KDD output which enables decision makers to make efficient and effective decisions (Ogut et al 2008; Chandury et al 1997; Hovy 2008).

3.4 Definition of data warehouse

There are various definitions of data warehouse, as such many scholars have defined the data warehouses differently but all yield the same meaning. Chen (2008) define data warehouse as a theme-oriented, integrated, persistent dynamic data structure to collect data and to support decisions by retrieving, converting, cleaning, and rebuilding external data in the conventional OLTP and other types of databases.

Inmon (1993) define a data warehouse as subject-oriented, integrated, time-variant and non-volatile collection of data in support of management's decision making process". In context of this definition, data warehouse enable the decision makers to access the right data in a right format which facilitates analysis thereby achieving improved decisions.

Ang et al (2000) also define data warehouse as a repository of summarized data (current as well as historical data) assembled in a simplified format tailored for easy end-user access. These data are collected from existing operational systems and are structured to be available in a form ready for analysis and decision making.

The above definitions of data warehouse underscore the primary goal of integrating various sources of data into single format. The users make use of this data for analytical processing activities include KDD methods, querying, reporting and other decision support activities. During this process of user operations, the end user applications only need to search the data warehouse instead of the source databases.

3.4.1 Data warehouse and KDD

The relationship between the activity of KDD and the data warehouse leads to an architectural foundation of decision support systems. Inmon (1996) argues that data warehouse is used to set the stage for effective KDD. Data warehouse is used to set the stage for KDD through two major roles which are data cleaning and data access (Fayyad et al 1996b). The process of data cleaning is associated with the role that DW enables the transformation of data into single unified format which presents single version of truth, in this process noise, errors and missing data are identified and eliminated. Also, data warehouse provides access to data which was difficult to obtain, such as archive and offline data (Fayyad et al 1996b). Although KDD can be employed on relational database, however the data warehouse greatly improves the chances of success in KDD (Inmon 1996). The characteristics of data warehouse enhance KDD process and prospects for success. Inmon (1996) described the use of the data warehouse characteristics in the KDD activities which are illustrated below.

The *integration of the data* permits the data miner to easily explore a vast amount of data. The miner usually spends his time in cleansing and preparing the data to be ready for the mining process in an effective manner. The data preparation includes reconstruction of keys, modifying integration rules, standardizing the data structure and the translation of encoded values. The data warehousing process includes the mention task for preparing the data, therefore Data Warehouse includes a ready data for exploration hence the data miner can perform the mining directly without wasting time**.**

The data in a data warehouse is *detailed* and *summarized*, the degree of details permits the miner to examine and explore the data in a granular form and perform a drill-down analysis, and also the low level of details provided can contain some important patterns that need to be carefully scrutinized. Furthermore, the summarized data allow the user to rely and build their work on others work, therefore there is no need to repeat the work someone else did which save huge amount of time.

Mining *historical data* is important, it reveal pieces of information that help to understand the seasonality of the organization activities and businesses. Mining a recent or current information could never lead to discover trends and a behavior pattern in a long term period.

The miner needs to understand the data before performing the mining, since it is very difficult for the data miner to work with vague and unexplained data. In data warehousing concept the

metadata aims to explain both the content and context which create a perfect platform to perform mining and extract patterns.

3.4.2 Data Warehouse Architecture and KDD

The data warehouse architecture presents the techniques and tools involves in collecting data from various governmental sources, processing of data and role that these techniques and tools play in extracting knowledge from Data warehouse. Chaudhuri & Dayal (1997) presents the architecture (see figure 5) which is divided into two parts, back end and front end application. Back-end application concerns with application that collects data from various sources and processing of data, front end application involves tools which extract knowledge from data warehouse which are OLAP, Reporting and Querying tools, EIS and Data mining. In this study, attention is given to Knowledge Discovery of Database (Data Mining). The following subsections briefly describe the components of the data warehouse architecture.

The *sources of data* are various; it could be the World Wide Web or even the personal notes. The government data sources consist of internal and external sources. The internal sources includes internal memos, reports, databases (OLTP, legacy systems), colleagues, etc (Winterman et al 1998), generally speaking the internal sources are the internal activities and interaction that happened inside governmental agency. The external sources represent the data acquired from outside the agency such as internet (eServices), other governmental agencies and information service centers (Winterman et al 1998).

The *ETL (Extraction, transformation and Load)* is the process of extracting the data from the data sources to cleanse and integrate it then load it in data storage (Enterprise data warehouse or Data marts), the importance of the ETL is in the role it plays in loading the data storage with integrated and cleansed data for further applications Turban et.al (2007).

The data after has been extracted, transformed and cleansed, are *stored* in data warehouse. The organization may choose to employ integrated enterprise data warehouse which collect the information on all the subjects across all department in the organization or data marts which collect the information for specific subject or particular department (Chaudhuri & Dayal 1997; Turban et al 2007).

Literatures indicate that due to the limitation imposed by high cost of data warehouse implementation, many organization favor the use of data marts which is easy to build and also enable to achieve the data quality efficiency as it employ consistent data model which ensures that same version of data for particular department to be accessed and viewed by all end users across organization (Chaudhuri & Dayal 1997; Turban et al. 2007). Since data marts are designed for specific interest for particular department, it captures and provides only the relevant data which is required.

However, there are problems associated with data marts implementation which includes the limitation of number of users it can support, the integration problem may also present if the business model is not well defined (Chaudhuri & Dayal 1997; Watson et al. 2001; Turban et al 2007).

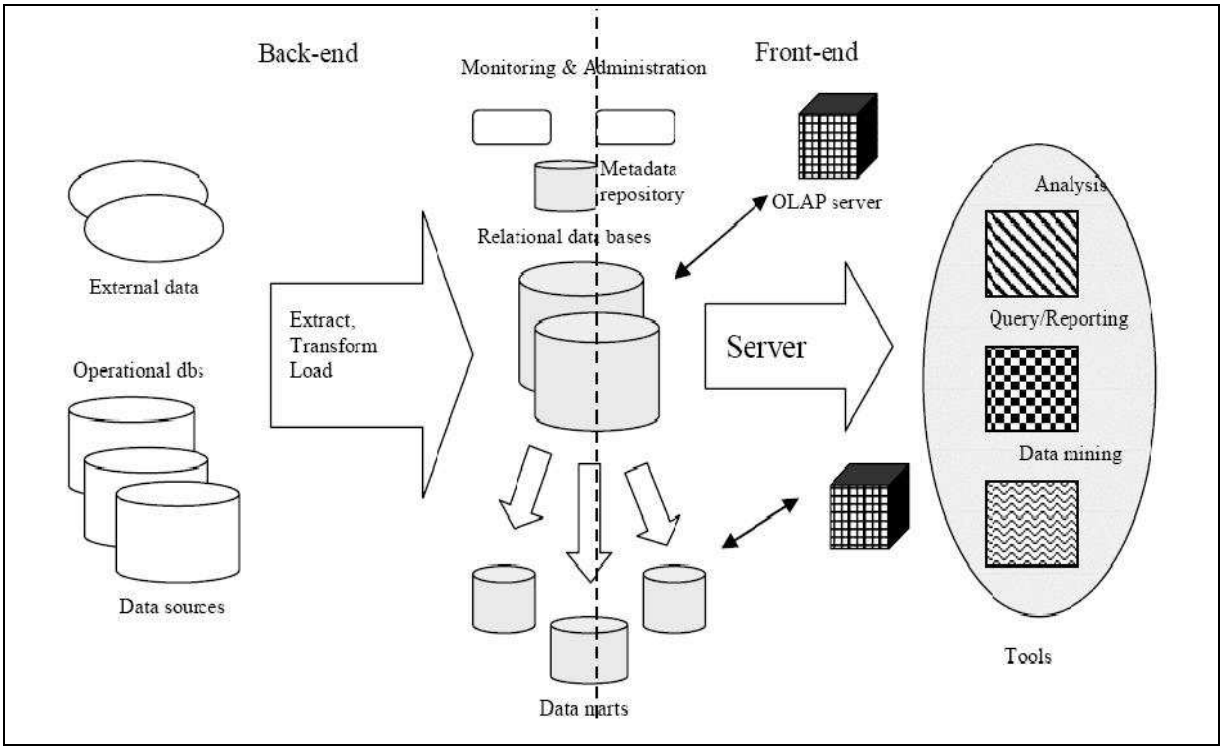

Figure 3.2: *Example of an architecture of a data warehouse (adopted from Chaudhuri & Dayal 1997)*

There are number of front-end *applications* that enable user access to the data stored in the data warehouse (Turban et al 2007; Chaudhuri & Dayal 1997). These tools can be categorized into the following as follows.

1. Reporting and Querying tools: These tools are used to generate operational reporting and also allow querying (use SQL query) data warehouse for any answers (Turban et al 2007).

2. Application development tools: These tools facilitate the creation of in house-application and services, developed application can also be intergraded with other application such as OLAP (Conolly 2004).

3. Executive Information system tools (EIS): These are tools designed to support the decision making at all levels in an organization. They allows users to develop customized graphical decision applications tailored to address some decision situations thereby improving decision making problems (Conolly 2004). In addition, they offer capabilities such as drill –down analysis and exceptional reporting which saves to monitor and guide the organization performance (Turban et al 2007).

4. Online Analytical Processing (OLAP) tools: These tools enable users to analyze data using complex and multidimensional views. Implementing these tools assumes that the multidimensional conceptual mode is implemented in data warehouse which means that data stored in data warehouse should be organized in multidimensional view (Chaudhuri & Dayal 1997; Connolly 2004; Turban et al 2007). OLAP tools allow examining of huge

data items in complex relationships which leads to the identification of patterns, trends and exceptions.

5. Data Mining: These techniques are used to identify meaningful hidden trends of data and unexpected relationships in huge sets of data in data warehouse which queries and report cannot effectively disclose (Chaudhuri & Dayal 1997; Connolly, 2004). Connolly (2004, pg. 1161) argued that data mining capabilities are potential than OLAP tools due to its ability to build predictive models rather than retrospective models. This argument is also supported from other school of thought who regarded data mining as appropriate tools to support decision making due to its abilities to analyze complex relationships of data in broader and concise way (Singh 1998; Connolly 2004).

3.5 Data Mining as a Step of KDD Process

Data Mining is the important and key step of KDD process as it is the step in which discovery of knowledge occurs. The data mining involves the application of algorithms and methods to the data sets to identify the relationships between data-sets data and reveal trends or pattern of data which offer understandable and useful information to decision makers (Ogut et al, 2008).

3.5.1 Definition of Data Mining

Data mining is the core process employed in the Knowledge Discovery in Database. Fayyad et al (1996b) defined data mining as "a step in the KDD process that consists of applying data analysis and discovery algorithms that produce a particular enumeration of patterns (or models) over the data".

Also, Hand et.al, 2001 defines data mining as analysis of large observational data sets to find unsuspected relationships and to summarize the data in novel ways that are both understandable and useful to the data owner (organizations). The definitions given also complement by Turban et al 2007 who consider data mining as a process that uses statistical, mathematical, artificial intelligence techniques to extract and identify useful information and subsequent knowledge from large database

From these definitions, we can draw the position of data mining as a sub- process within overall process of discovering knowledge from large data sets. Particularly, data mining algorithms are employed to extract pattern from large data sets which can also serve to give a difference with the KDD process which is overall process of discovering useful knowledge from data.

Having understood the definition of data mining and pointed its distinction with the KDD process, the explanation is now given to describe the use of Data Mining in the KDD process.

3.5.2 Data Mining in the KDD Process

The data mining in the KDD process involves the iterative application of data mining method and algorithms to extract and enumerate pattern from data sets (Fayyad et al 1996b; Sardieh et al 2008). To better understand the Data Mining process in KDD, the goals of knowledge discovery

need to be explored, the methods and algorithms which use these methods also need to be described.

Fayyad et al (1996b) distinguished data mining goal into two types of goal; verification and discovery, later concerned with verifying users hypothesis while former concerned with system autonomously way of find new patterns. These goals further subdivided into prediction and descriptions. The prediction goals is aimed to enable the system to extract patterns for predicting the future behaviors of some entities while with description, the goal is to make the system discover patterns and present them to a user in a format that human can understand. These two goals are considered in practice to be the major goals of data mining Fayyad et al (1996b).

The usefulness of data mining depends mainly on the application area, for instance the prediction and forecasting is widely used in marketing and sales activities. The mostly used data mining method for this kind of activities is called classification. *Classification* is a learning function that maps (classifies) a data item into one of several predefined classes; Examples of classification methods in knowledge discovery applications include the classification of trends in financial markets (Fayyad et al. 1996b). Also *Clustering* method is common in this application area since it seeks to identify a finite set of categories or clusters to describe the data. Examples of clustering applications in a knowledge discovery realm include discovering homogeneous subpopulations for consumers in marketing databases (Fayyad et al 1996b)

Moreover, the data mining is employed to discover relation between different variable hidden in a huge database which could assist the decision maker in his activities. For instance, *Dependency modeling* method deals with finding a model that explain the significant dependencies between variables, and *Summarization* method involves methods for finding a compact description for a subset of data. Summarization application can be employed to interactive exploratory data analysis and automated report generation.

Having described and illustrated the use of data mining as core process in KDD, the next section examines the use of KDD techniques in support various Government organization activities. The Government major activities include decision making and service provision. To accomplish these tasks Government requires considerable amount of useful information to enable decision makers to identify and understand decision situations thus make timely and improved decisions (Ogut et al 2008). Literatures confirm that Government has accomplished the above mentioned activities by employing KDD applications which offer useful information to aid in decision making process.

3.6 KDD in Government

The KDD applications are being used in the various areas in public sectors (Government), this includes finance and economy, market, healthcare, criminal justice and defense, transport (Bach 2003; Fayyad et al 1996). The KDD helps to support this organization activities as it identify and extract valuable information which enable decision makers in the government to take appropriate and better decision hence improve efficiency to provide services.

The motivation of using KDD in government could be attributed to the fact that decision makers in government are confronted by turbulent decision environment which consist of ill-structured

decision problems and poor information quality which inhibits them to make appropriate decisions (Ogut et al 2008). Given this situation, the KDD plays important role in the decision making process at it provides useful and quality information assists decision makers in achieving better decisions. There are number of application areas pointed above which KDD helps, we briefly elaborate the use of KDD in decision making context with the focus on the decision surface those areas.

The US government tax agencies have employed KDD application Clementine to foster collection of tax and audit and reduce fraud (Bach 2003). In this context KDD is responsible to identify the problems and/or opportunities such as good or bad tax players and also pin point the tax evaders. The identified problem/opportunities revealed by KDD thus enable the decision makers to understand the problem or opportunity and allows them to make appropriate decisions to address the problem or opportunities.

Fayyad et al (1996b) also mentioned that US treasury financial crimes enforcement network which have employed KDD to the identify suspicions financial transactions which might indicate that there is money laundering activities.

Furthermore, Vector et al (2000) illustrates the use of KDD in support the decision making process for South African Department of Arts, Culture, Science and Technology (DACST). KDD project was initiated by DACST with objectives to construct a knowledge base to support the government decision makers when they formulate the Knowledge and Technology Policy Framework. Vector et al (2000) further clarifies that KDD was used to assure that the data in the data warehouse support the findings in the Knowledge synthesis report, and also KDD was used to generate new insights which can helps to test the suitability of the policies included in the final Knowledge and Technology Policy Framework. The derived usefulness of KDD in this case includes its ability to allows the decision makers to acquire knowledge directly from domain data thus facilitates timely decisions.

Literatures also indicate that KDD has been using to support decision making process at criminal justice departments in Government. In this context, the KDD is used to evaluate and generate the crime patterns and provides insights for crime situations thus enable the decision makers to plan for the resources allocation in order to prevent crimes from occurring (Bach 2003).

3.7 Summary of Literature Review

The established theoretical frameworks described in the previous chapter, underline the increasing data explosion in the government. Theories further indicates that, the reasons for such enormous increasing of data in the government is attributed to the use of ICT by the government in an effort to provide best service delivery to citizens and support their decision making process. The use of government agencies of ICT to automate their transaction contributes much to this problem.

This dramatic increasing of data leads to the scattered and non-unified data as different systems are used to capture and process data. Moreover, the data are presented, generated, maintained and used independently in each government department and agencies. This environment leads poor decisions and also inhibits the smooth processing of data into useful knowledge which can aid decision makers in the Government in support of decision activities. Given the turbulent and complex environment that the Government is experiencing, there is a need of effective tools to extract more useful information to assist the decision makers to better understand and identify the decision situation and be able to respond to various challenges facing them and hence make efficient and effective decisions.

There are number of available techniques and tools identified that can address the problem of data explosion and derives useful knowledge to assist the decision making process. The Knowledge discovery in database (KDD) is among the available techniques that the Government agencies and departments have employed. KDD consists of eight process starts from data selection to the evaluation process of identified knowledge. Data warehouse has been employed in data selection and it is used to set the stage for KDD by offering data cleaning and data access thus enhance the efficiency of KDD process. Importantly also is the Data Mining which is the steps are important as well as they involves the evaluation of discovered knowledge to determine the relevance of such knowledge before is put into use.

4. Research Methodology

4.1 Research Approach

This study adopts qualitative research as a research approach. The qualitative research permits us to explore the research topic and subjects in a detailed view so as to get the deeper understanding of phenomenon, in this context the usefulness of knowledge discovered in database (KDD) in the decision making process in the government agencies (Creswell 2007). Also, it is important to explore the decision makers' activities in their natural settings and their use of knowledge provided by KDD in treating decision in complex and rapid changing environment to get better understanding of the usefulness of KDD in the process of decision making.

4.2 Research Design

The selection of an appropriate research strategy which is useful for carrying out the research is a critical decision and requires considerable attention. Yin (2003) discussed three different conditions that help social researchers to determine what and when to use a particular research strategy, he mentioned the conditions as: the types of research questions, the ability of the researcher to control events and access available resources and the degree of focus on contemporary as opposed to historical events.

Our research question is based on "what" questions. The central question is "*What are the perceived usefulness of the KDD in the government decision making process?*" .Yin (2003) discussed the two types of the 'what' question, First, exploratory 'what' question which justifies the rationale for conducting an exploratory study. The second type of 'what' question is actually in a form of 'how many' and 'how much' line of inquiry. The use of what question as exploratory permit any of the strategies to be used (Yin 2003). Thus our research study adopts an exploratory approach.

Also, this study investigates the perceived use of knowledge discovery in databases (KDD) to support decision making process in the government agencies which is an emerging field and can be related to contemporary phenomenon. Thus through this study, we will be able to get detailed description of the phenomenon which will help to describe how the knowledge discovery in databases support the decision making process in government agencies. Our study involves government agencies in Sweden that employ KDD as a knowledge discovery technique and make use of the knowledge in the decision making. The selection of Swedish government agencies as our research study is attributed to the advancement of ICT infrastructures in Sweden Government (UN eGovernment Survey, 2008). This advancement of ICT leads to the automation of the Government transactions which results to an explosion of data in government agencies thus provide a case to explore how the KDD can extract useful knowledge to be used in the decision making process.

4.3 Data Collection

The data collection process involves the gathering of research evidences. Yin (2003) and Creswell (2007) states that data collection phase is an extensive process of collecting evidence from different sources of data. Yin (2003) further categorizes six sources of evidence used to collect the data in case studies which includes documents, archival records, interviews, direct observations, participant observation and physical artefacts. Given the limitation of time for conducting this research project, it will not be possible to make available all the sources for analysis. Thus our research study will involve interviews as a source of evidence.

4.3.1 Interview

In this research, the interview is employed to collect useful information which enables us to understand the research topic from the subject's point of view, Yin (2003). Furthermore, the interview gives the opportunities to share the government agencies' understanding, experiences and perspectives in the process of knowledge creation and the use of such knowledge in assisting the government decision making process. This useful knowledge received from the interview results is valuable for the analysis which helps us to address the research questions.

Since our research is focused on the use of KDD in Government decision making process, we initiated the selection process of appropriate interviews by searching through the internet for Swedish Government agencies or organisation which has employed KDD techniques. After we found the agencies or organisation which implemented KDD, we further looked for the personnel who are involved with the technical process of KDD and those who are responsible to take decision based on the knowledge provided by KDD.

Our research interviews was conducted with the following persons from government agencies in order to collect the empirical data that address the research question, they are; *Prof. Lennart Waldenlind* the head of signal group in Swedish Medical Product Agency, *Dr. Niklas Noren* Acting Manager of Research & Development at the Uppsala Monitoring Centre, X is a Business Intelligence consultant at Y company and *Dr. Anette Hulth* a researcher in KDD field in the Swedish Institute for Infectious Disease Control. The interviewees have a significant educational degree which proofs their academic knowledge and backgrounds as well as they occupy important and critical position in their respective agencies. They demonstrated that they use KDD methods in their activities and daily work to support their decision making or even to provide knowledge that the decision maker use. However, all interviewees demonstrate strong knowledge and experience within KDD use in decision making context

Additionally, semi-structured and open ended questions are adopted so as to allow us to ask the interviewee about their opinions about the events (Yin, 2003). Furthermore, it allows to adapt to the situation and ask further questions if necessary which also attract rich discussion on the matter.

We have tried to ask questions that are not biased and which also serve our line of inquiry. This is achieved through posing open ended questions and avoiding threatening questions (Yin 2003). In this regards we did not put "why" questions as would creates defensiveness in the informant

party. The interview put forth the environment that interact with subjects as informants rather than respondents by asking open ended questions and involve them in a dynamic discussion to exchange ideas and opinions (Yin 2003; Kvale 1996).

Interview guide indicates the purposes and topics to be covered in interview with a detailed sequence of questions (Kvale 1996). These questions cover thematic (knowledge production) and dynamic dimensions. Moreover, thematic is related to the relevance of the research topic, theoretical conceptions and subsequent analysis. Thematic are structured to obtain categorized data. The dynamic dimension contains questions that promote a positive interaction, keep the flow of the conversion going and motivate subjects to talk about their experiences and feelings.

Our interviews used two guides; each guide contains one theme. The first theme addresses the technological part such as Knowledge Discovery in Database (KDD), Data Warehouse and discovered knowledge; this theme is used to interview the subjects from the ICT division of the government (public) agencies. The technical theme aims to capture the technical process of the data collection and the knowledge creation using data warehouse and KDD. For instance, we have asked "What are the sources of data in the data warehouse?", "Does the KDD techniques identify or discover knowledge" and "How accurate and relevant the discovered knowledge is". These three questions try to capture the varieties of the sources that feed the data warehouse which provides appropriate platform for KDD, as this extract useful knowledge which was addressed in the second question. Hence, evaluates the quality and relevance of the knowledge discovered which also reflects the goal of the last question.

The second theme address the usefulness of knowledge discovered in the decision making process of the government agencies; this theme is used to interview subjects who are involved in taking decisions in the government agencies. In this theme, we posed the following questions, thus follows: "How do you use the KDD knowledge to identify problem or opportunities in your department or organisation?" and "How the knowledge discovered assist you in developing and exploring different course of actions for your decision problems"? As a consequence, these two questions aim to capture the use of KDD knowledge in the decision making phases (see appendix A).

4.4 Data analysis

This stage involves examining, categorizing and testing the empirical evidence in the study (Yin 2003). For this purpose, this research analyse interviews which is our sources of empirical data. The analysis is presented as follows:

4.4.1 Interviews

The process of analyzing the interviews follows the transcription phase. This research employs interview notes taking and tape recording as a method to facilitate the transformation of oral interviews into written texts for analysis. The major strength of tape recorder is to capture all the empirical data accurately and offer a reliable method to cross check and reuse them (Yin 2003; Kvale 1996). However, the use of tape recorder depends on the interviewee approval and consent. Moreover, the use of recorder allows us to focus on the interview subject and gives

control of the interview process. To attain the reliability of transcribed interview, two authors are involved in transforming the taped interview into two written transcripts which had been compared to ensure that both have higher consistency.

Furthermore, the interview notes assists in coding the answers and key points given by respondent during interview. The note taking shall also capture the reactions of respondents as they are reacting towards the interview questions. Reactions can be behaviours, body language and gestures which along can give insight and deep understanding of some phenomenon regarding the topic. This method requires active listening and awareness during the interview (Kvale 1996).

❖ *Method of Analysis*

Interviews were analysed through ad hoc meaning generation technique which use multiple methods to organise the interview texts and condense the meanings into the form that can help to draw conclusion to the study (Kvale 1996). With the use of the meaning condensation methods, the interview data were condensed into relevant data which address the purpose of the study to answer our research question.

Also, meaning categorization is employed in order to code interviews into categories. Categorization allows long statements to be reduced into predefined categories, and thus reduce and structure a large text into a few tables and figures (Kvale 1996). The analysis contains seven categories (decision making, data, knowledge, intelligence, design, choice and implementation).The method strongly provides an overview over the meanings experienced by the respondent (Kvale 1996).

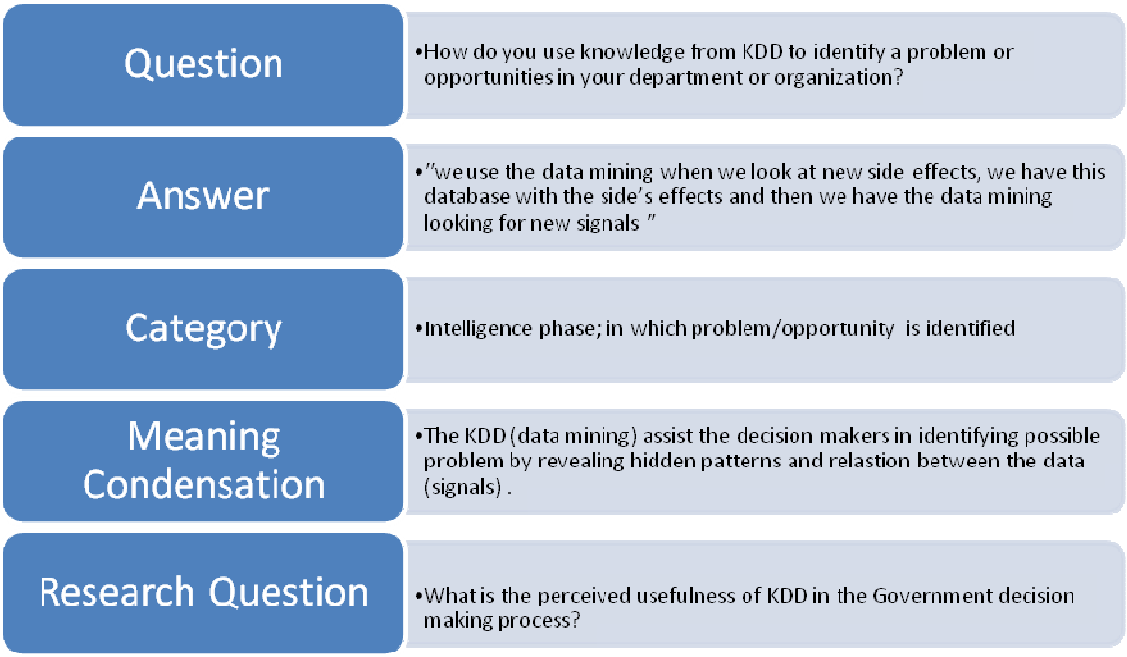

Figure 4.1: *Illustration of the process of analyzing the interviews meaning*

Moreover, our research employed cross-interview synthesis as analytical technique to analysis our interviews (Yin 2003). This technique permits us to aggregate our result across each interview. Moreover, word tables are employed to present the result for usefulness of KDD in each decision making phases namely intelligence, design, choice and implementation.

The result of each table helped to pinpoint the similarities and difference of the usefulness of the KDD for each category as compared to the usefulness of KDD identified in literatures. This also helped us to draw our conclusion and generalize our findings hence show the contribution of this study.

Moreover, the evaluation of the perceived usefulness of KDD are based on a criterion which has been derived from Fayyad et al. (1996a) definition of KDD as "the nontrivial process of identifying valid, novel, potentially useful, and ultimately understandable patterns in data" (see section 3.2). This criteria consists of four entities; meaningfulness, goal relevance, operational validity and innovativeness (Menon et al 1992).

The meaningfulness is how much the knowledge is relevant and makes sense for the decision maker. Goal relevance indicates whether the knowledge discovered is potential usefulness to the current task in hand or not. The operational validity points to the actual use of the discovered knowledge. The last entity is innovativeness which is a characteristic of the KDD output which provides non-obvious knowledge to solve practical problem.

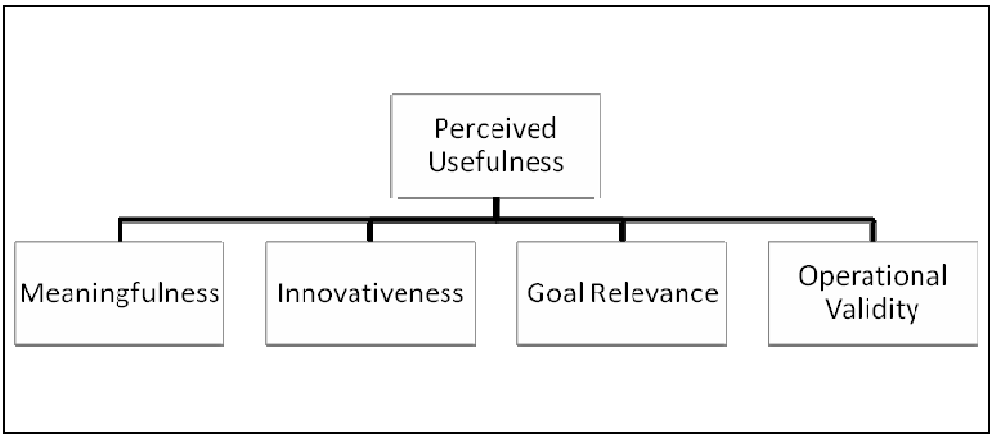

Figure 4.2: *Criteria for the perceived usefulness (adopted from Menon et al 1992)*

4.5 Research Quality

In this chapter, we are addressing the measures to ensure good quality of research. Lincoln & Guba (1985 cited in Seale 1999) mentioned validity and reliability as the criteria that can measure the trustworthiness of the research. This notion of trustworthy consists of four elements credibility, transferability, dependability and conformability which are analogous to the concepts

of internal validity, external validity, reliability and objectivity (Seale 1999). The below is an illustration on the measures that we adopted to achieve the quality in our research.

Table 4.1: *Lincoln and Guba's translation of terms (adopted from Seale, 1999)*

Conventional Inquiry	**Naturalistic Inquiry**	**How to achieve**
Truth value (internal validity)	Credibility	peer review of reports, member checks
Applicability (external validity)	Transferability	Detailed and rich descriptions of the settings
Consistency (reliability)	Dependability	Cross checking the different interpretations of data collected (auditing)

1. *Internal Validity*: This ensures that the research process guarantee that result and final conclusions really correspond to the purpose and the research questions. This relates to the truth and confidence of the research findings Lincoln & Guba (1985 cited in Seale 1999). In our research this has been achieved through peer review of reports and member validation.

2. *External Validity*: In this process, the major aim is that purpose, research questions, result and conclusions should correspond to issues and interests of general importance (in both practice and theory). Achieving external validity allows us to generalize or transfer the results of qualitative research from other contexts or setting as the data will be collected from different settings (government agencies) Lincoln & Guba (1985 cited in Seale 1999) These have been achieved through detail studying of each agency and generalize the results from one to another.

3. *Reliability*: the degree of consistency of results over time, it seeks to ensure that different observer make the same explanation of a particular object to answer the need of having a single version of research therefore different interpretation are unacceptable Seale (1999). In our research this is achieved through cross checking the different interpretations of data collected and assuring that they are similar.

The *member validation* is concerned in presenting a convincing account using the opinion of people on whom the research has been done, to check that the account has correctly incorporated different perspective Seale (1999). Bloor (1997) identified three different types of member validation; the validation of the member taxonomy, the validation of the researchers' analysis by the demonstrated ability of the researcher to pass as a member and the validation of the researcher's analysis by asking collectivity members to judge the adequacy of the researcher's analysis.

However, we validated the analysis by asking the members of the study to evaluate and validate the accuracy of the interview transcript, the description of the participants in their settings and the full final report of the study with monitoring their response. The members of this research agreed on the interpretation of the interviews conducted and in the context it was used, hence improve the credibility of the research.

4.6 Research Ethics

Ethical conduct has increasingly become crucial in social sciences research method. Researchers found themselves in the dilemma between clear adherences to ethic conduct procedures while complying with regulatory regimes requirements from ethic committees. Beauchamp and Childress (1994 cited in Seale 1999) defined ethic as a generic term for various ways of understanding and examining the moral life'. It is concerned with perspectives on right and proper conduct. The increasingly research activities have great impact on the society thus there is a potential need to monitor and ensure that ethics are taken seriously and embraced within the research process. The need to incorporate ethic is to enhance quality and produce what is right, good and virtuous as by product of research works. It is important that research integrity be incorporated in order to assist in the validation of the research and enforces researchers to behave ethically. There is a great need to avoid and eliminate scientific misconduct, corruption so as to produce more good to the people and hence minimize harm to them. Israel and Hay (2006) argued that research need to develop better understandings of the politics and contexts within which ethics are regulated. In our research, we have considered and complied with important ethical issues such as informed consent, confidentiality in order to avoid harm and doing good.

4.6.1 Informed Consent

It is important that the research participants understand exactly their involvement in a research and what they have authorized (Seale 1999).

In order to assure an ethical conduct, we explained to the subjects the purpose of this research, the part of the research that this interview will be used for and the potential risks that they could face .Also, we informed the subjects that the interview will be recorded for further analysis and interpretation. All the mentioned information revealed to the participant is considered a part of the informed consent which aims to minimize the possible harm and risk, and increase the trust between the participants and us, hence protect participants and the agencies from unpleasant consequences.

Furthermore the participation in our research was completely voluntary and lacks from any kind of pressure or influence performed from superior employees or managers.

The anonymity of the participants is guaranteed in case they asked for that. This is due to the lack of direct identification information. We had also enforced the anonymity by ensuring that the data collected would not be cross-checked with any other source of information that could reveal the identities of the participants.

4.6.2 Confidentiality

In this research, we consider the duty of confidentiality as crucial and we take necessary procedures to ensure that it is complied in our study.

The study protected the confidentiality of research participants so that their private data will not be reported however to reveal their personal identifiable information, there shall be a formal agreement which gives the approval (Israel et al 2006; Singer et al 2002).

Protection of data collected in a research is an important step in achieving confidentiality. As we are anticipating the data gathering from interviews which will concern the interviewees and their corresponding government agencies, there would be a formal written agreement with the subjects in order to enforce the efforts to maintain the confidentiality of all information collected. Furthermore, the data collected is being stored in the computer protected passwords so as to restrict the access to this sensitive information (Israel et al 2006).

In an effort not to disclose the information of our subjects, the names and details of one government agency involved in this research will not be made available to anybody except to the authors of the thesis. As a result of this, the interview transcript will not appear in the appendix of our thesis.

4.6.3 Avoiding Harm and Doing Good

Harm could be as physical or social and its degree depends mostly on the risk and consequences encountered in revealing the information for that reason debriefing the interview is an essential part in our ethical consideration which will create an awareness of the data exposed. Therefore the balance between harm and the benefit of the sensitive information is a key issue (Israel et a 2006).

Our research strives to balance between doing harm and maximize the benefits to the society .We look forward that the research outputs will have significant contribution to the understanding of the use of new knowledge discovered in assisting the decision makers to acquire the appropriate information that aid them to make better and improved decisions hence contribute to the society well being. Also, the result contributes to the knowledge base in this area of study.

5. Empirical Findings

In this chapter, we present findings and analysis of the interviews performed. The analysis of the interviews involves illustrating interviewee's natural meaning along with their interpretations. The following categories describes and analyze the interviews held with the respective organizations, these are; decision making, data sources, knowledge, intelligence phase, design phase, choice phase and implementation phase. This structure eases the understanding of the interviews content and address the research question efficiently.

The four interviews were conducted; two of them involved personnel (UMC and public owned company X) who are involved with the technical process of the knowledge creation through KDD which had ability to answers some decision making questions. The other two interviews involved the decision makers (MPA and SMI) who use the KDD's knowledge to aid them in the decision making process who could also answers some few technical questions related to KDD.
The Following table present a thick description of the agencies interviewed as well as the interviewee background, expertise and education level.

Table 5.1: *Interviewee and agency descriptions*

Interview	Type of Interview	Agency	Agency Description	Interviewee Name	Interviewee Expertise
Interview A **See Appendix C1**	Telephone	Uppsala Monitoring Centre (UMC)	Uppsala Monitoring Centre (UMC) is a Collaborating Centre for International Drug Monitoring founded based on an agreement from 1978 between the World Health Organization and the Swedish Government which is also updated 2001. The centre is responsible for the collection of data about adverse drug reactions from around the world and particularly from countries that are members of the WHO in order to explore and extract useful knowledge related to drug reactions using knowledge discovery tools such as KDD. The knowledge found is used to identify potential risks which could affect human life hence take action in this respect.	Dr. Niklas Norén	He is a senior Statistician and acting as research manager in UMC. He is responsible of research and development of new methods for knowledge discovery in adverse drug reaction surveillance. He also provides professional support on data analysis methodology.
Interview B **See Appendix C2**	Face-to-face	Swedish Medical Product agency (MPA)	MPA is a government agency responsible for regulation and surveillance of the development, manufacturing and sale of drugs and other medicinal products. The agency is located in Uppsala. The MPA is employing Knowledge discovery in Database (KDD) to extract knowledge presented as signals that helps in discovering trends about the side effects of the drugs approved in Sweden, Europe and other countries markets. These signals helps MPA in taking appropriate decisions regarding the approval for the drug quality, safety and efficacy	Prof. Lennart Waldenlind	He is the Head of Signals Group within Swedish Medical product agency (MPA). Lennart is in charge of the signal group which is division of the drug safety department. He is responsible for detecting and evaluating signals extracted from data warehouse through KDD.
Interview C **See Appendix C3**	Face-to-face	Agency X	The company X is a Swedish- owned limited liability Company. The company has employed data analysis technique which includes KDD, OLAP, Balanced Scorecards and dashboards. Company X has been using these techniques to monitor and measures performances in financial, economy, statistics and market departments.	Person Y	Person Y is a Business Intelligence consultant in company X. Persons Y is responsible for consulting on data analysis technique and Business intelligence solutions. Y worked with company X since 2002, prior to that Y worked with other companies on the same area of Business intelligence
Interview D **See Appendix C4**	Telephone	Swedish Institute for Infectious Disease Control (SMI)	SMI is government expert agency responsible for monitoring and surveillance of the epidemiological situation for infectious diseases in humans and promoting protection against such diseases. As a part of surveillance activities, SMI has developed computer supported outbreak detection system which employs Knowledge discovery in Data base (KDD) to produce automatic alarms when the level of any of notifiable disease in Sweden has reached the level that might indicate that there is outbreak of disease.	Dr. Anette Hulth	She is involved in the system of computer supported outbreak detection in SMI, responsible for detecting and evaluating alarms generated from KDD. She has been working at SMI for more than 2 years; she has background in computer and systems science. Annette has conducted numerous researches and published articles on Automatic keyword extraction using Natural Language Processing and Machine Learning for Information Retrieval.

5.1 Decision Making

5.1.1 Interview A

The UMC decision making structure is organized through the joint administration of Sweden Government appointed officials and WHO appointed officials. The centre is involved with decisions concerns the evaluation and recommendation over the adverse drugs reactions. The centre is therefore, strives to make sure that it take effectively and timely decision to maintain that they discover potential hazards coming from drugs and recommend to countries that are members of the WHO Program on International Drug Monitoring the appropriate measure in order to prevent side effects of medicines to humans . As a consequence of this decision environment, it is clear that UMC operates in a rapidly changing environment which requires them to be agile in order to respond to the challenges brought by adverse drug reaction. In an effort to make and recommend appropriate decisions, the UMC has employed Knowledge Discovery techniques to identify signal (adverse drugs reactions). Based on the identified adverse drug a reaction, the UMC is mandated to evaluate and communicating their findings in order to prevent from side effect from such drug reactions.

5.1.2 Interview B

The interviewee B provided us with the understanding of decision making process in MPA. The interviewee mentioned that MPA is involved with three major decisions, which include decisions regarding the approval for the quality of drugs, Efficacy and the safety of the product, it was also mentioned these decisions differ considerably and complex. The complexity of decisions is attributed to the fact the decisions made relates to the health of human beings. Interviewee said that *"so, we have three major decisions, and they are quite different"*. Moreover, the diversity of decision indicates that there is a need of appropriate and useful information to enable decision makers to understand their decision situation thus is able to make efficient and effective decision. Interviewee confirmed that some of the decisions are difficult thus requires more useful information

> *"Approval on the safety of product is the most difficult, safety begins with studies on animals … then you continue with patients, then small group of patients where you study the drugs intensively, the study is supported by reporting of adverse drugs reaction" (See Appendix C2).*

The interviewee has also indicated that there are a number of factors which influence the decision making process in MPA. This includes political pressure which forces the quick approval of drugs, the introduction of new drugs for serious (pandemic) disease and social consideration to lower the cost for drugs which seems unaffordable to some countries. He mentioned that "*… so political pressure contributes in the early approval of the drug …"(See Appendix C2).*

To support the decision activities, MPA needs information to facilitate the decision making process. In light of this, KDD has been used in identifying and extracting the signals for side effects of drugs. The information obtained from discovered drugs side effects enable the SMI to take appropriate measures to address the problem. Moreover, Interviewee also indicated that after identified the side effect of drugs, they evaluated such results using risk benefits criteria

which leads to the following decisions; modifying the drug prescription, using the drug just for severely ill patients and redraw the drug from the market.

Furthermore, interviewee stated that after having discovered side effects for the drugs, MPA discuss the result in groups to evaluate its significance because they are so many factors that can influence the interpretation of the findings. He said that

> *"…first we identify the signals which usually is a statistical signal, and then we take that for a group decision every week and then we discuss according to the signal table what could be the reason for this…"(See Appendix C2).*

According to the interviewee, the decision making process in MPA involves number of decision makers within the agency. Also, for the matter which involves common objectives with other European Counties, there is also group decision making across the European Medical Agencies.

5.1.3 Interview C

The decision making structure in the public owned company X is organized through Financial, economy, marketing and statistics department managed by CEO. According to the interviewee, the major types of decision include budget, policy making and forecasting. The interviewee confirmed that, business pressure and changing environment is driving the company to make quickly and timely decisions.

> *"Well, for instance we have budget, policy making, forecasting. I mean different kind of decision; it depends on how you control your daily business. For examples if we see some businesses are unprofitable then we have to make the decision to close such business now" (See Appendix C3).*

Most of the decision made under such business pressures are new and not structured such decisions creates a need for adequate knowledge to enable decision makers to make better decisions. In an effort to respond to these types of decision situations, organisation X has implemented KDD process to guide them with market decision activities. In this case, the KDD has been used to identify and generate market trends which provide information on customers, purchasing behaviour, and position of markets segments. This information guides the decision makers in achieving efficient and effective market decisions.

5.1.4 Interview D

The decision making structure in SMI is organized through seven departments and three sections headed by the Director General. The departments include Administration, Bacteriology, Epidemiology, Immunology and Vaccinology, Parasitology, Mycology and Environmental Microbiology, Virology and Microbiological Preparedness (KCB). The interviewee mentioned that SMI main decisions task is the ones involved with the control of disease, other decision includes those support main decision such as budget, investment and priorities in the work environment. Interviewee further clarified that SMI is more an expert agency as they charged with providing expert advice to the Government, She said that:

> *"…our main task is to surveillance, to keep track of the status for the infections disease in the country and also suggest what could be done by the decision makers, so SMI is an expert agent…"(See Appendix C4).*

The decision making process at SMI involves activities which aim at identifying the outbreak of diseases and recommend appropriate clinical measures. The decision making process includes the analysis of the reports of suspected notifiable diseases from medical doctors and also from laboratories, this is important as the medical doctors need to confirm if there is a real a outbreak of disease. The decision situations involved with this process are complex and requires more useful knowledge to assist them to better understand their decision situations in order to provide effective and effective decisions.

The interviewee illustrated that, SMI has employed KDD methods to the data sets (reports) in order to extracts alerts which might indicate if there is an outbreak of disease. Moreover, experts (epidemiologists) engage in group decision making to evaluate the result which in this case the outbreak of disease discovered in order to formulate appropriate actions.

5.2 Data sources

5.2.1 Interview A

The data warehouse of UMC is being fed by different external sources, the data is a form of individual case safety reports that contains suspected adverse drug reaction of real patients which have been treated with a specific drug and experience some side effects caused by the drug. These reports are being collected from national centres of pharmacovigilance across the world such as food and drug administration in USA and medical product agency in Sweden *"The sources of data in our database (VigiBase) include individual case safety reports." (See Appendix C1).*

The sources of data are from different government agencies and independent centers across the world which exhibit different data formats therefore the integration of these data is crucial in order to provide consistent and single version of data for further analysis and decision activities which raise the need to implement a data warehouse. Since the relevance of the new discovered knowledge (patterns) has a direct relation to the quality of data used in the KDD process, this justifies the importance of the data warehouse to improve the KDD output *"…it's also related to the data quality issues …"* . Also the data warehouse is used to store historic data which permit the KDD to discover trends in the data *"… the oldest report on database comes from 1968 … the most recent report would be from 2009" (See Appendix C1).*

5.2.2 Interview B

The scattered data across different databases was very problematic for decision making for MPA since a single version of data is important to have a true insight on the actual situation. Before 1995, the Swedish Medical Agency did not have any effective data communication with other European countries because of the lack of a centralized database which lead to act independently without any external contribution. After 1995 there have been important initiatives to collect and integrate data from different European countries in order to enhance the collaboration and increase the knowledge dimension which lead to better drug control and quality in aspect of risks and benefits *"…it's been developed since 1995, but before that the country was reacting more separately, …we have this common database (DW) for all the European countries …" (See Appendix C2).*

The European Medicines Agency (EMEA) eudravigilance data warehouse (DW) is a centralized data repository that stores integrated and cleansed data. Moreover this data store is used in decision making and knowledge extraction using intelligent tools such as data mining. The data warehouse is being feed with data from different sources such as the European countries and the other countries where the drugs are being sold, which create an important data platform to apply the KDD to discover useful and meaningful patterns *"…it is from all Europe … and sold in other countries then all serious reports are coming in our data warehouse then we apply data mining on the data to discover signals" (See Appendix C2).*

5.2.3 Interview C

The problem with data explosion has also been the case of company X. The company collects a lot of information from internal and external sources. Internally, they collect sales figures, products and financial data from internal legacy systems and also collect the same data externally as electronic transactions coming from the business branches. The data explosion in company X is a result of automation of their business transactions. However, there is a need to properly process and integrated the data in a consistent format which can facilitate access and easy analysis. In line with this concept, they employed data warehouse as single repository which collects and integrates all the internal and external data from different sources *"We have a lot of data...all the sales figures, the products and financial data fed into our data warehouse" (See Appendix C3).*

The data warehouse enables company X to have fast access to the quality data thereby facilitates quick and timely decisions. In addition, the company has realized the need to have proper tools/techniques that can assist them to access the data in the data warehouse and extract useful information. Among other applications, company X has been using data mining and OLAP tools *"… and data mining that is used mostly in marketing department" (See Appendix C3).*

5.2.4 Interview D

In contrast to the other interviewed agencies, SMI does not have a centralized data warehouse to store the incoming data. They use Relational Database Management System (RDBMS) to store the data originated from different sources. As a use of the ICT infrastructure, SMI collect data from clinical reports which are coming from medical doctors and contain information on the patient with suspected infectious disease symptoms from any of notifiable disease. They also receive the medical reports from laboratories which contain the test results from cases reported of having been infected with a notifiable disease and in some cases they require information from other agencies such as Swedish Statistics and Sweden National boards of health and welfare *"…the reports that we get from medical doctors ... and reports from laboratories … data from Swedish statistics from time to time … other agencies for example National boards of health and welfare…".* As the other interviews, SMI affirmed that the Database contains historic data which permit to discover trends that help the decision making *"…I believe we have data from the sixties. It also depends on the disease…" (See Appendix C4).*

5.3 Knowledge

5.3.1 Interview A

Knowledge is the final output of the KDD process. KDD involves applying data mining methods and algorithm that screens the data-set in order to extract knowledge. Interviewee A mentioned that they employ KDD techniques to access and extract patterns (knowledge) from data warehouse contains data regarding the individual case safety reports. He said that *"methods developed specifically for the purpose of analyzing this type of data…looking for patterns in the database" (See Appendix C1).*

So the Knowledge Discovery in Database (Data Mining) has been employed to extract useful hidden information (pattern) to reveal relation between groups of drugs and syndrome of various medical events, in other words KDD (Data Mining reveals the unseen relationship between variables of situations which a normal human cannot identify due to the limited cognition hence used this information in effective way. The KDD helps the domain expert in identifying patterns that they could not identify in manual review of the databases (data warehouse).

The knowledge discovered which presents pattern extracted from data sets, need to be verified by experts in the application domain to determine if this knowledge is relevant, in this case clinical experts are being involved with this process of evaluation. He confirms that *"After discovering this kind of data (new knowledge, patterns) we need our domain experts. In this case, domain experts (clinical expert) go through the findings." (See Appendix C1)* .

The evaluation process carried out by clinical experts establishes the usefulness of discovered knowledge and also establish if the discovered knowledge correspond to new knowledge. The resulted evaluated knowledge is put into use to assist the UMC with decision making activities.

5.3.2 Interview B

Interviewee B also provided the opinions on the use of KDD to discovery knowledge. Also in this case, it was iterated that as due to the large amount of data collected, it is important to employ KDD to overcome the human cognitive limits and capabilities and make the data collected useful in a way that improve the agency activities and decisions. Interviewee B said that they use KDD to extract signals (side effects) from which contribute significantly in the decision making. He said that*'…we have this database which we apply data mining for looking to, new signals…'…its help us with 30%-50% in the decision making…' (See Appendix C2).*

Since data mining is a tool that employs algorithms to screen the data for significant patterns, the user interaction is very important. The interaction in the Swedish Medical Product Agency is summarized in two key points. First the sensitivity of the KDD process (data mining) which controls the number of signal discovered and the second one is specificity which is concerned with the data to be mined. The first point is a challenge for the user of data mining since setting a high sensitivity will produce a large amount of patterns which many of them are irrelevant and setting it low, it will produce fewer patterns which could lead to lose some important information.

> *"…the problem is also or what we call sensitivity and specificity, if the data mining is very sensitive you will get a lot of signals and many of them are not real and that is a problem and if it has less sensitivity you don't get so many signals and maybe lose some of them so it is a balance…"(See Appendix C2).*

This challenge in setting the data mining variables leads to produce knowledge which has to be evaluated by domain experts (pharmacologists) to identify the relevance of the discovered patterns, hence taking the right decision.

5.3.3 Interview C

In this case, it has been observed that the agency X collects large amount of data which are fed into data warehouse. The data consist of sales figures, products and financial accounts. The agency realised that could make benefits out of this data and had decided to employ KDD process to extract useful relationship or trends that can help them in decision making activities . Interviewee C mentioned that, the KDD has been applied to the data sets to extract and identify market trends which help the company to prepare and design market and sales campaigns. Furthermore, Interviewee clarified that, the process of data mining is outsourced to another company. *"So we extract the data set from our data warehouse and send it to another company to perform the data mining, then the marketing department use it in their activities" (See Appendix C3).*

As from KDD process, the interviewee described that the knowledge discovered needs to be evaluated by market specialists to determine its relevance based on the available market information. The evaluation process identifies the contribution and usefulness of such knowledge thus enables them to take appropriate courses action to design market campaigns.

5.3.4 Interview D

Interviewee D provides her opinions on the knowledge as it apply to the KDD in the SMI. As explained at previous data theme section, the SMI receives reports from medical doctors and a laboratory, reports contains information on patients with suspected infectious disease of notifiable disease. These reports feature some of important patterns of information which cannot be retrieved through manual procedures. Interviewee agreed that KDD methods have been employed to identify relationship between data and generate alarms when the level of any of notifiable disease in Sweden has reached the level that might indicate that there is outbreak is going on. "*we are applying data mining methods to try to find variation to find something is not normal ." (See Appendix C4).*

This interview also show that the KDD process has been fully implemented, as it is observed that the generated alarms were going through evaluation process by domain experts (epidemiologists) . Epidemiologists has been involved in assess the alarms generated in order to confirm if there is really an outbreak thus recommends appropriate measures to treat and prevent the disease.
"…*manual work is mainly used to validate and evaluate the knowledge discovered by the data mining…" (See Appendix C4).*

5.4 Intelligence phase

5.4.1 Interview A

As the intelligence phase is concerned with identifying problems or opportunities, the KDD play an important role in mining the data in an efficient way to extract hidden information that the normal human cannot reveal and considered problems/opportunities *"...to identify patterns which may be missed in manual review…"*. As KDD points out to a certain patterns discovered from the data warehouse, it limit the number of issues that needs to be investigated for more information gathering, hence save time and decrease the information overload that the domain experts usually face when evaluating a certain problem *"...through limiting the number issues that…" (See Appendix C1).*

As a consequence of discovering knowledge about new suspected drug negative effect, other information is attributed to it, the population proportion (problem owner) with the drug side effect could be specified by KDD *"…in the data, and it says they are all report for young children…" (See Appendix C1).*
Furthermore, the KDD plays important role in using historic data, it scan the data warehouse and retrieve patterns and relationship between the old and new data which could discover information that affect directly the decisions to be made "…*in long term it could of course … discover a new suspected side effect,… it could affect that drug stays on the market or not*." *(See Appendix C1).*

5.4.2 Interview B

The interview gave us an insight on the intelligence phase activities involved in the decision making. He described some of these activities which the MPA engages in identifying problems and opportunities.
The interview showed us that, they employ KDD to search for the information in the process of identifying problem. He said that *"we use the data mining when we look at new signals (side effects)"*.The KDD extract signals from the MPA data warehouse in a timely manner, these signals corresponds to side effects of the drugs. The discovered signals (side effects) enable them to determine the magnitude and significant of the situation. Based on the significance level, the situation can be a problem or opportunity. For MPA it is always an opportunity hence it could lead to improve the health of people however it is considered as a problem for the manufacturer since the drug could be redrawn from the market *"…it is always positive for people but it is negative for the manufacturer… " (See Appendix C2).*

Also, KDD has been used in problem/opportunity identification where there is a need to investigate a huge population with unknown effects. Moreover, the interviewee realizes the importance of using the KDD in this phase, and he confirmed relying on it *"…it is very useful, its help us with 30%-50% in the decision …" (See Appendix C2).*

5.4.3 Interview C

The public owned company X faces a number of decision problems in executing day to day activities at operation, administration and strategic tasks which involves market and sales activities. In order to identify the problem, KDD is used to search for information and provide

knowledge which gives insight on market trends. The market trends information includes the information about the customers, their purchasing behaviour and position of markets segments. It also gives the factors that affect market campaigns *"Data mining gives a lot of information that market department use to identify the market trends and also identify the factors that affect the campaign" (See Appendix C3).*

Therefore, the knowledge extracted by KDD form the basis of the problem or opportunity that the public owned company X might have. The positive market trends could suggest that there is good market opportunities and therefore alert the decision makers to increase the production. The negative market trends could suggest the drop on the sales of the products, therefore alert the companies to launch effective market campaigns that meet the needs of the buyers hence boost the sales. According to interview C, the KDD provides this crucial knowledge (trends) in a fast and good format which helps the decision makers to make timely and better decisions thereby reducing the time and effort could be wasted in manually searching for problems or opportunities.

5.4.4 Interview D

The main objective of SMI is to detect outbreaks of diseases by scanning the database with data mining algorithms. The output of this KDD process is knowledge and hidden information or patterns that has to be evaluated by the epidemiologist in order to consider it to be an outbreak or not *"…KDD (data mining) helps in outbreak detection and which is the mission of SMI*.".Using the mining algorithms, SMI have the ability to analysis the data available from different sources and identify possible outbreaks in a certain population which is considered an opportunity to improve people health. As a consequence of the KDD process in outbreak identification, the time needed to reach a useful knowledge or information has been reduced since the human capabilities are limited when dealing with a huge amount of data *"…KDD (data mining) helps us in directing the resources in a timely manner." (See. Appendix C4).*

Since the SMI has a large quantity of data, it is impossible that the epidemiologists scan the data available in an efficient way to extract information that could lead to an outbreak. Therefore the KDD plays an important role in identifying the needed information hence provides a focus point that the epidemiologists for further investigation which is considered a kind of starting point that is impossible to identify normally, hence decrease the information overload *"…making them focus more on what is probably relevant…it gives us a starting point that the normal scan could not identify." (See. Appendix C4).*

5.4.5 Summary of Intelligence Phase

The following table cross-check the usefulness derived from all the interviews in the Intelligence phase. Based on the below table we notice that *Identify pattern/trends, Time reduction* and *Identify opportunity/problem* had been agreed on by all the interviews however the *Overcoming the human cognitive limits* is agreed on by three of the interviews**.** Moreover, two of the interviews agreed that the KDD *Decrease information overload* and *Identify problem owner***.** Finally, one interview stated that the KDD provides information about the *problem attributes.*

Table 5.2: *Usefulness of KDD in intelligence phase from four interviews*

	Usefulness					
	Identify pattern/trends/relationship between data	**Overcome cognitive limits**	**Decrease information overload**	**Identify problem owner**	**Identify opportunity or problem**	**Time reduction**
Interview A	Yes	Yes	Yes	Yes	Yes	Yes
Interview B	Yes	Yes			Yes	Yes
Interview C	Yes				Yes	Yes
Interview D	Yes	Yes	Yes	Yes	Yes	Yes
Agreed on	4	3	2	2	4	4

5.5 Design phase

5.5.1 Interview A

The interviewee A has provided the usefulness of KDD in the design phase of decision making as it apply to the identification of signal for adverse drugs reactions. In this context, the KDD has enabled them to direct resources as it provides the reports (signal) pattern which indicates where there is a problem (adverse drug reactions). Since the KDD reports highlight the problem, it helps the decision makers to better understand the problem and focus on such problem thus enable them to allocate necessary resources such as clinical reviewers to undertake and evaluate best measures to address the problem. It should be noted that in this case, KDD does not develop courses of actions but it gives the required information to assist decision makers in developing course of actions.

The findings show that, if the KDD would not have been employed, clinical experts would not have appropriate knowledge to develop appropriate measures or course of actions. This is also attributed to the fact that clinical experts would have supposed to screen manually a lot of information in order to come out with problem and appropriate course of actions *"firstly is to screen the database for the interesting report pattern to help direct the resources of the clinical reviewers so that we can use our domain expert as effectively as possible, so that they can focus on the right issues" (see. Appendix C1).*

5.5.2 Interview B

These courses of action are designed based on the information available on the specific side effect, therefore the knowledge is crucial to treat complex situation which contains different variables. Moreover, the data mining is effective in a complex situation more than it helps in a turbulent environment which needs a timely decision in MPA decision situations *"The more severe the problem is the more rapid …"*.In a complex decision situation which is not considered a serious case the data mining could be applied on the data after six months since the information on the specific side effect is limited and the average time to gather adequate size of data is half a year. Hence the KDD discover signals that assist the decision maker in designing the solutions *"…we need to wait in order to have more knowledge … reports on the problem which is extracted by the data mining…" (See. Appendix C2).*

Furthermore, KDD has been extensively used to focus their attention to specific areas during the problem identification *"…this is how data mining could lead to some actions in resources direction and decisions…".* KDD is also effective in forecasting the unforeseen situation; it could provide knowledge regarding expected problems based on the historic data available concerning a specific case. The patterns discovered from previous drug used to treat some diseases, can create a picture about the expected side effects for a new drug, however it does not always helpful since the problems appears suddenly *"… we try to foresee things drawn from the earlier data mining problems…".*The KDD it is used to identify possible solution for a certain problems however in MPA case it is used together with other techniques such as trial test and animal experiments because of the sensitivity of the situation which deals with the human life *"…it is one of the methods… liver problems in animal studies…"(see. Appendix C2).*

5.5.3 Interview C

Having identified the problem or opportunity, the decision makers in marketing department engaged in a discussion to find a way to address the problem situation. The discussion involves identifying the possible courses of action for each problem or opportunity identified. The interviewee demonstrated that KDD assist them to predict and forecast a number of possible solutions for the problems or opportunities. This is performed by identifying the customer behaviour using the trends discovered in the intelligence phase to predict the possible purchasing behaviour which are possible solution for the problem. The generated courses of action are evaluated through criteria set by marketing department *"… then they collect all the information (market trends) … then they take decision based on the knowledge available to launch new campaigns in the future." (See. Appendix C3).*

5.5.4 Interview D

The KDD has the ability to discover the unseen, in SMI case the KDD knowledge is used to evaluate the situation hence support one of the usual courses of actions employed. However when a new situation raise the KDD help the decision maker in directing the resources available for more exploration ***"…****it is a tool that we need … the KDD (data mining) helps us in the directing the resources.****".*** Furthermore, the interview attests that the KDD also assist the decision maker in developing the courses of actions in an indirect way by providing the needed knowledge ***"…****evaluate the outbreak discovered in order to formulate appropriate actions…" (See. Appendix C4).*

5.5.5 Summary of Design Phase

The cross table (see table 4) presents the summary of Usefulness of KDD in Design Phase from all interviewees. All interviewees agreed and attest the usefulness of KDD in assisting to develop the appropriate course of actions, and that it provides knowledge which helps them to better understand the problem situation and hence led to development of course of actions. However, only three (3) interviewees agreed on the use of KDD in assist them to direct the resources.

Table 5.3: *Usefulness of KDD in design phase from four interviews*

	Usefulness	
	Assist in developing the appropriate courses of action	**Direct the resources**
Interview A	Yes	Yes
Interview B	Yes	Yes
Interview C	Yes	-
Interview D	Yes	Yes
Agreed on	4	3

5.6 Choice phase

5.6.1 Interview A

The interviewee A described the use of KDD in the choice phase of the decision making process. He provided his opinions based on the roles that KDD plays in developing different courses of action in the previous phase (intelligence and design), the decision maker need now to choose one and considers it as a solution for the problem. In this case, KDD has been used to extract knowledge and present it to the decision maker to assist him in choosing the best course of action available. He mentioned that KDD discover knowledge about negative effect of drug on a long term which alert decision makers to take appropriate decision regarding the production of this drugs *"…if we discover a new suspected drug negative effect not directly but in long term, it could affect that the drug stays on the market or not… " (See. Appendix C1).*
The discovered knowledge which in this case the result of adverse drug reaction which brings negative effect (harm) to human, helps the decision makers to decide as to whether to withdraw such identified drug from the market or not depending on impact (harm) which such drug brings to the human being.

5.6.2 Interview B

The interviewee B also indicated that the KDD assist the decision maker in evaluating the possible courses of actions by extracting signals (drug side effects) from the data warehouse that will enforce one of the solutions available which add more understanding to the decision maker about the problem situation therefore help the decision maker in using the risk benefit for the appropriate course of action. In this interview B, the central concern is the risk benefit which they use to decide on course of action to be adopted*"…liver problems for example and they are not serious, we apply data mining to identify signals related to this side effect, if the signals have the same results (not serious) then it is maybe enough to add this to the prescribing information…" (See. Appendix C2).*

As the findings indicate, the KDD provides the useful information that guides the decision makers to the make appropriate decision thereby achieving efficiently and effectively decisions. The findings from interviewee C and D have indicated that the KDD has not been exploited to guide decision makers to choose among selected course of actions. The interview shown that, the data mining was not used to guide the decision makers in choosing the best course of action among the possible identified solution.

5.6.3 Summary of Choice Phase

The cross-table below presents the summary of findings from all interviewees in context of choice phase. The first two interviews agreed and attested that KDD assists decision makers in choosing the best course of action and evaluating the possible courses of actions.

Table 5.4: *Usefulness of KDD in choice phase from four interviews*

	Usefulness	
	Assist decision makers in choosing the best course of action	**evaluating the possible courses of actions**
Interview A	Yes	Yes
Interview B	Yes	Yes
Interview C	-	-
Interview D	-	-
Agreed on	2	2

5.7 Implementation phase

Based on the empirical findings and the data collected, the KDD did not provide use or assistance in putting the solutions into actions.

6. Discussion

In this chapter, we present answer to our research question. The main themes used in our empirical findings that target the decision making process are discussed and will permit us to address the research question.

6.1 Usefulness of KDD in the Intelligence phase

Literatures have indicated that, the intelligence phase of decision making is attributed with different activities starting with problem or opportunity identification (Turban et al 2007, Holsapple 2008). In this phase, the problem is classified and problem owner is identified which lead to create a problem statement. The complex and turbulent environment that the government is facing requires appropriate techniques that make use of the huge amount of data available to extract useful knowledge which helps the decision makers in addressing the attributes mentioned (Ogut et al, 2008).

Different scholars agreed that KDD is been useful in identifying the problems and/or opportunities and discovering relationship between different data (Turban et al.2005; Fayyad et al 1996; Bach 2003). This view has been supported by all the interviews in the empirical findings (see Table 5.2). The usefulness of the KDD in this context is due to the innovativeness, meaningfulness and goal relevance of the knowledge discovered since it assist in discovering adverse drug reaction, drug side effects, market trends and outbreak of disease as evidenced in all interviews respectively.

In spite of supporting the usefulness of KDD from literatures, our empirical findings pointed out to other usefulness which has been agreed on by most of the interviews. Our empirical findings indicate that KDD has a potential use in this phase as it provides the decision makers with useful knowledge to make decisions. As per all the interviews, KDD reduce the time of collecting useful information and knowledge which is used in the intelligence phase (see Table 5.2). The agreement of all interviews on the usefulness of KDD in time reduction is mapped from the meaningfulness and goal relevance of the knowledge discovered. The meaningfulness and goal relevance of KDD knowledge enables the decision makers to gain a clear and precise understanding of the problem and/or opportunities faced in a timely manner since the four interviews agreed that it would be very time consuming and approximately impossible to scan the data manually in an efficient and effective way to extract the required knowledge.

Moreover, three of the interviews (A,B and D) affirmed that the KDD helps in overcoming the human cognitive limits as it permit to scan and extract meaningful knowledge and relationship from a large amount of data which is used in the agency operations (see Table 5.2).

As KDD extract summarized meaningful and innovative knowledge from large amount of data which is agreed on by the interview A and D, it is useful in decreasing the information overload faced in the actual complex decision making environment. Also, interview A and D found that KDD identify the problem owner in some cases since it identify the population with drug side effect cases or the area of the disease outbreak for instance, drug side effects in children and outbreak in specific area of Sweden respectively. This usefulness reflect the operational validity and meaningfulness of the information since the knowledge extracted is used in the agency

activities, and also goal relevance since interview D affirmed that KDD assist in attaining the agency goals (see Table 5.2).

6.2 Usefulness of KDD in the Design phase

As mentioned in the literature section (see 2.4.1), the design phase of decision making aims at finding and analyzing possible courses of action for a problem or opportunity (Turban et al. 2007 & Holsapple 2008). However, in order to achieve the above aims, useful knowledge is required to enable decision makers to better understand the problem and hence develop appropriate course of actions. Literatures indicate that KDD has been used to assist the decision makers to understand hidden relationship between data thus enable to develop the courses of actions, also it assist in plan for the resources allocation (Bach 2003, Fayyad et al 1996) see also Table 6.1. In addition of generating the alternative courses of actions, decision makers are required to develop the criteria which can access the acceptability of generated actions or outcomes. Our empirical findings revealed that, government agencies have realized the use of KDD and has used it to assist in development of the courses of actions and guide to analyze the potential of such actions. The four interviews (see Table 5.3) revealed and attested that KDD assist them to understand and identify hidden relationship between data thus guide them to develop course of actions. Findings also showed in three interviews (A, B, D) see also Table 5.3, KDD assist them in direction of resources during the process of developing the course of action.

The interview A illustrated that KDD helps to identify signal (drug side effects) and highlight problem. This enables the direction of resources which leads the clinical experts get focused during the analysis and development of the course of actions. The perceived useful of KDD in this case is related to meaningfulness, goal relevance and operational validity as it helps the decision makers to have precise knowledge which is used to carry out decisions in this phase.
Also in interview B, KDD is perceived to be useful in forecasting and prediction of possible side effects of drugs based on available historical data of specific cases. The expected drug effect provides information that enable to establish and test the possible course of action or solutions. As due to the KDD knowledge, the direction of resources is facilitated leading the pharmacologist to focus on identified problem during the process of design the solutions. The perceived usefulness of KDD in this context is more focused in the ability that the discovered knowledge provides to the decision makers to enable them to understand the problem identified, develop and analyses the possible course of action for such problem. Additionally, the KDD is perceived useful as it provides appropriate knowledge for timely decision to prevent harms to humans that can be caused by drugs side effect.

The interview C identified that KDD is useful in predicting customer purchasing behaviors for a certain set of product. In this case, KDD is perceived to be useful as it helps to generate the customer purchasing behaviors which guide the decision makers to focus on what to produce and what customer to target, this enable the organization to adjust the production of products in order to meet customer demands. This perceived usefulness of KDD in prediction of customer behaviors can be reflected in terms of goal relevance, innovativeness and meaningfulness. This is because it enable the decision makers to acquire useful and meaningful knowledge which help them in generating, testing and validate the generated course of actions. Generally, it results to efficient and effective decisions into this phase.

In the last interview D, KDD is useful as it provides useful information regarding the drugs side effects which enabling epidemiologists to evaluate and development of the course of actions. Moreover, the discovered KDD knowledge enables the direction of resources to formulate appropriate course of actions. As a result of KDD, the decision makers assured better and accurate knowledge guiding them to develop appropriate courses of action .This is in consistent with meaningfulness, goal relevance, operational validity and innovativeness.

6.3 Usefulness of KDD in the Choice phase

Literature had described the choice phase as where the decisions are being made to solve the given problem; it includes the choice of the best course of action based on organization criteria, concept, goal or even a special model developed for this purpose (Turban et al 2007). The literatures could not reveal the usefulness of the KDD in assisting the decision makers in choosing the best course of action among the available alternatives (see Table 6.1). However, our empirical findings reveal that KDD assist the decision makers in helps in evaluating the possible course of actions and choosing the best course of actions (see Table 5.4).The usefulness of KDD in this context had been confirmed by two interviews (A and B) see also Table 5.4 . In interview A , the KDD reveals the information about the long term drug side effect which gives them an insight on the actual status of the drug side effect hence leading to take decision as whether to redraw the drug or not. The KDD knowledge in this case, provides a meaningful and valid knowledge which affect the choice of best decision.

Also, interview B, the knowledge extracted by the KDD process assist the decision maker in evaluating the possible courses of action which permit the use of the risk benefits criteria hence choose the best course of action such as redraw the drug from the market or not based on the adverse reaction of the drug. Also, the perceived usefulness of KDD knowledge in this case is valid and useful from interview B point of view.

6.4 Usefulness of KDD in the Implementation phase

According to Turban et al. (2007), the implementation phase is where the solution is put into action. Literatures could not indicate the use of KDD in assisting the decision makers to implement the final solution to the problem (see Table 6.1). Also, the empirical findings collected could not reveal the usefulness for KDD knowledge in supporting this phase.

Table 6.1: *Usefulness of KDD identified in literature and agreed on by interviews*

	General Usefulness from Literature					
	Intelligence		**Design**		**Choice**	**Implementation**
	identify the problems and/or opportunities	discover relationship between different variables and factors	understand hidden relationship between variables to develop the courses of actions	plan for the resources allocation	Nothing has been mentioned	Nothing has been mentioned
Interview A	Yes	Yes	Yes	Yes		
Interview B	Yes	Yes	Yes	Yes		
Interview C	Yes	Yes	Yes			
Interview D	Yes	Yes	Yes	Yes		
Agreed on	4	4	4	3	0	0

7. Conclusion and Future Research

In this chapter, we present the conclusion from our research and answer our research question. We also provide the recommendation to the future research of this project.

7.1 Conclusion

In this study, the purpose is to explore the perceived usefulness of the knowledge discovered through the knowledge discovery in database (KDD) in the decision making process of government agencies. The following question was posed

- What is the perceived usefulness of KDD in the Government decision making process?

To answer this question, we have developed criteria to assess the usefulness that the KDD knowledge provides to the decision makers to enable them to make decisions in rapidly changing environment which government agencies experience. These criteria are: meaningfulness, goal relevance, operational validity and innovativeness. The perceived usefulness of KDD in the decision making process in the government have been derived from four decision making phase namely intelligent, design, choice and implementation. These four phase constitutes the decision making process.

The empirical findings confirm that, the perceived usefulness of KDD in intelligent phase is to identify pattern or relationship between data which enable the decision makers to gain a clear and precise understanding thus leading to the problem and/or opportunities identification. The KDD is also perceived useful as it helps the decision maker to overcome his cognitive limits as it permit to scan and extract meaningful knowledge and relationship from a large amount of data. Moreover, KDD is also useful in decreasing the information overload by presenting to the decision makers an appropriate amount of knowledge extracted from the huge amount of data thus facilitates the timely and efficient decisions. Findings also revealed that KDD is perceived useful as it enable the identification of the problem owner during the data collection and analysis (see Table 5.2).

In comparison, the literature pointed out the two general use of KDD in intelligent phase such as identify pattern or relationship between data, for example as in identify crime situations and identify suspicions financial transactions (see 3.6) , secondly KDD helps decision makers to identify the problem and/or opportunities (see Table 6.1). These two uses have been also agreed by all interviews as indicated in Table 5.2. However, literatures could not provide the detailed usefulness of KDD in this phase. In contrast, the empirical findings explore more usefulness of KDD in intelligent phase (see Table 5.2). Findings attempts to explore in details the use of KDD in assisting the decision makers in the activities for identifying problem or opportunities which is main activities in intelligence phase. Furthermore, findings from all interview affirmed to the use of KDD in time reduction while its use in identify problem owner, decrease information overload, overcome cognitive limits has been distributed among the interviews as indicated in Table 5.2. Therefore the empirical finding contributes and provides deeper understanding on the usefulness of KDD in intelligence phase.

In the design phase, empirical findings showed that KDD is perceived useful as it enables the decision makers to acquire useful and meaningful knowledge which help to understand the

problem thus guide them to develop, test and validate the generated courses of actions. In addition, empirical findings shows that KDD is useful in the direction of resources which a course of action or an aid for developing them (see Table 5.3). The empirical findings are consistent with the literatures which also shows that KDD has been using to assist the decision makers to understand hidden relationship between data thus enable to develop the courses of actions, also it assist in plan for the resources allocation (see Table 6.1). Furthermore, findings have attempted to depict the usefulness of KDD in choice phase of decision making process, KDD provides meaningful knowledge that enables decision makers to evaluate complex decision situations thus contributing to the ability of examine different alternative courses of action and choose the best solution (see Table 5.4). This finding serves as important contribution to the understanding of the use of KDD in choice phase. The literatures could not indicate the usefulness of KDD in choice phase.

This research finding also identified that the KDD has not been used in the implementation phase, which is also supported by the same result from the literature. The reasons for why KDD has not been exploited in this phase can be the subject of further research.

The following table summarizes the usefulness of KDD derived from the empirical investigation.

Table 7.1: *Usefulness of KDD in the decision making process*

		Decision Making Process			
Usefulness from the research		**Intelligence**	**Design**	**Choice**	**Implementation**
	Literature	• Identify pattern/trends/relationship between data • Identify opportunity or problem	• Assist in developing the appropriate courses of action • Direct the resources	-	-
	Empirical Findings	• Identify pattern/trends/relationship between data • Identify opportunity or problem • Overcome cognitive limits • Decrease information overload • Identify problem owner • Time reduction	• Assist in developing the appropriate courses of action • Direct the resources	• Assist decision makers in choosing the best course of action • evaluating the possible courses of actions	-

As a conclusion, the cross-case analysis permitted to identify the similarities between interviews and compare it with the usefulness derived from the literatures, therefore we generalize the following usefulness of KDD only in the context of this research and the sittings investigated.

1. *Identify pattern/trends/relationship between data*
2. *Identify opportunity or problem*
3. *Time reduction*
4. *Assist in developing the appropriate courses of action*

7.2 Future research

The explosion of data created a need to develop new technologies to address this important phenomenon. The data generated every day contains important information which needs to be extracted in an efficient and cost effective way, and then transform this information into knowledge by evaluating it. This knowledge could be used in different ways for different subjects; business, health, non-profit services and many other areas. KDD is currently used to create knowledge out of raw data to support different aspect of the organization. It is important to know that the main engine and heart of the KDD process is the data mining algorithms which identify patterns hence lead to knowledge. In this research we tried to explore the perceived usefulness of KDD in the decision making process of the governmental agencies. During the research process we encountered many possible future researches that we couldn't address because of our limitation. We briefly describe the possible future research as follow:

- The data mining is the central process of KDD which aims to scan the data repository for related and significant patterns. Despite of that KDD is an intelligent tool to extract knowledge but it needs the human interaction to follow and support the process for effective output. One of the important areas of this interaction is setting the Sensitivity and specificity of the KDD process in extracting knowledge. The sensitivity control the amount of patterns detected; low sensitivity leads to large number of patterns discovered which could contain insignificant output however high sensitivity will lead to small number of discovered patterns hence increase the chance for missing some important patterns. This requires a special balance sensitivity which is based on the application area on the KDD. This could be more clarified and explored by investigating the way that the different organization (private and public) set the Sensitivity and specificity of data mining.

- As per this research, it is clear that the KDD had not been used in the implementation phase of the decision making process. This raise a that why it is not used and how can we benefit from the KDD capabilities to support this phase.

APPENDIX A

Interview Guide:

Interviewee name:	……..	**Interviewer 1:**	……..
Agency:	……..	**Interviewer 2:**	……..
Department:	……..	**Date :**	……..
Position:	……..	**Time:**	……..

Procedure:

Step 1: Introduce ourselves to the interviewee in aspect of education background.

Step 2: Explain the purpose of the interview. What is the purpose of the interview and why we are conducting.

Step 3: Explain the rights of the interviewee in context of his/her confidentiality, anonymity of the interview and request for inform consent.

Step 4: Ask the interviewee if he/she has any questions, clarifications or concerns before starting to record the interview.

Step 5: Ask for the permission to use interview tools such as tape recorder, note etc.

Step 6: Start Recording and ask questions according to the themes below.

Step 7: Questions

General question:

1. What is your experience and background in the use of Knowledge discovery techniques?

Technical theme: KDD to produce knowledge

a) Data warehouse

2. What are the sources of data in Data warehouse (Name internal and external sources)?
3. What type of data warehouse is implemented in the organization? (Data marts or Centralized data warehouse or Real Time Data warehouse)

4. How old the data is in the data warehouse?

b) Data Analysis Techniques (Knowledge Discovery techniques)

1. What are the analytical techniques employed to access the data warehouse (OLAP, KDD, Data Mining, ANN, data visualization tools)
2. Does the technique identify and /or discover knowledge?
3. How such analytical tools support and improve decision making?
4. Is the Data mining being used to perform predictive analysis and forecasting?

c) Discovered Knowledge

1. How accurate and relevant the discovered knowledge is?
2. How do you manage the knowledge created or discovered?
3. How do you classify the knowledge discovered in aspect of relevance, incompleteness and irrelevance?
4. How the knowledge is presented to the users?
5. How do you transform the decision maker requirements into rules?
6. What are the major departments that use the discovered knowledge?

Decision makers theme:

1. What types of decisions are made in the agency?
2. What is the general decision making process agency in your department?
3. What is the degree of structurdness of the decisions encountered?
4. Do you use the Knowledge discovery techniques such as KDD and/ or data mining?
5. How do you use knowledge from KDD to identify a problem or opportunities in your department or organization? (extends to usefulness during interview)
6. How knowledge discovered assist you in developing and exploring the courses of actions for your decision problem(s)? (extends to the usefulness)
7. How knowledge discovered assist you to evaluate and select the best course of action for your decision problem (s)? (extends to usefulness during interview)
8. How is the knowledge being used in forecasting and/or predicting the unforeseen situations?
9. What is the perceived usefulness of knowledge discovered through knowledge discovery tools in decision making process?
10. Do you need external information to assist you in decision making?
11. What is the perceived usefulness of knowledge available in decision making process?

Finally, stop recording and ask the interviewer if he/she wants to share any ideas about the interview and if he/she has any comments or questions that can be answered by the interviewers. (Debriefing)

THANK YOU FOR YOUR PARTICIPATION.

APPENDIX B

Introduction Letter

I, Kelvin Kiritta and Imad Bani-Hani, we are conducting a research study for the purposes of obtaining a Masters Degree in Information Systems at the University of the Lund. Our research is focused to investigate the perceived usefulness of the knowledge discovered through the knowledge discovery in database (KDD or Data Mining) to assist the decision making process of eGovernment agencies (government agencies).

The research will involve interviews which aimed to collect information of how the knowledge is discovered and use of such knowledge to assist the individuals making decisions.

With your permission the interview/s may be recorded in order to ensure accuracy. Participation is voluntary, and no person will be advantaged or disadvantaged in any way for choosing to participate or not participate in the study. All of your responses will be kept confidential, and no information that could identify you would be included in the research report unless you permit us by checking the "I agree …" statement at the bottom of this paper. The interview material (tapes and transcripts) will not be seen or heard by any person at the school or elsewhere, and will only be processed by myself. Any disclosure of personal identifiable information will be agreed upon in a formal consent.

The interviews/discussions will only be processed by ourselves and will be used for education purpose only. The final output of this research will be published at the Lund University Thesis Database which can be public accessed through web address http://biblioteket.ehl.lu.se/olle/

You are free to withdraw your participation at any point without any further explanations or any personal consequences. We need your written consent for participating in this study.

Your participation in this study would be greatly appreciated.

Kind Regards

Kelvin Kiritta
Imad Bani-Hani

Consent Form

I,_____ have received information from Kelvin Kiritta, Imad Bani-Hani on the study "Knowledge Discovery in Government decision making process ." I am aware that my participation in the interviews are voluntary, and that I can freely withdraw my participation at any time.

☐ I agree to disclose my personal information as well as actual position and organization/agency activities and name.

Signature Date

APPENDIX C

C1.Transcriptions of the Interview with Niklas Norén – Uppsala Monitoring Center (Interviewee A)

R stands for Researcher; **N** stands for interviewee 1

R: What is your experience and background in the use of Knowledge discovery techniques?
N: My current position is Acting Manager, R&D at the Uppsala Monitoring Centre and am involved with research and development of new methods for knowledge discovery in adverse drug reaction surveillance.
R: ok thank you. my first question is What are the sources of data in Data warehouse (Name internal and external sources)?
N: The source of data in our databases includes individual case safety reports and they're contributed by national centers of pharmacovigilance across the world. So in Sweden it may be the medical product agency and USA it would be the food and drug administration and other countries it may be independent centers responsible of the collection of these reports in the respective countries. The individual case safety reports are reports of suspected adverse drug reactions incidents in real world clinical practices where actual patients having received treatment for some illness and then experience adverse events that the doctors believe was or might be related to the medication.

R: So it's just not internal data from Sweden, it's internal and external sources?
N: Well it' depends on how you define external sources, its coming from national centers then we store the data in-house.

R: What type of data warehouse is implemented in the organization? For example: Data marts or Centralized data warehouse or Real Time Data warehouse?
N: Well I don't think that's really my area. My area is ah is the research of method to analyze data, I am not directly involved in the architecture of the data warehouse, I am not sure how I could classify according to those, I would have to pass for that question.

R: How old the data is in the data warehouse?
N: There various of course, the oldest report on database comes from I think I can't say for sure, I think it is 1968 maybe 1967 or even 1969 but it is late sixties. So those are oldest report in the database, then we have of course new report coming in all the time, the most recent report would be from 2009.

R: What are the analytical techniques employed to access the data warehouse? For example: OLAP, KDD (Data Mining), Data Mining, ANN, data visualization tools maybe.
N: Well, we have a range of knowledge discovery method develop specifically for the purpose of analyzing this type of data, most of them have been developed in-house, you may be call them KDD or Data Mining Method (DMM). It doesn't really say very much, I mean, there the range of pattern discovery methods basically looking for a pattern in the database, striking the association between different fields on report or groups of drugs and they all tend to report together the syndrome of various medical event, so is range of different patterns, so I would say we focus on pattern discovery since is not really prediction or forecasting, it is more trying to discover a report pattern which are interesting in some sense and worthy of further following up.

R: so it mainly about new knowledge discovered from database?
N: Basically for this kind of data we need domain experts in this case. In this case clinical expert, go through the findings but the purpose of the knowledge discovery is divided in two, firstly is to screen the database for the interesting report pattern to help direct the resources of the clinical reviewers so that we

can use our domain expert as effectively as possible, so that they can focus on the right issues and secondly is to identify patterns which may be missed in manual review of the database and pickup things that you may not reacted to due to the human cognitive capabilities.

R: Does the technique identify and /or discover knowledge?
N: Ah yes, I mean any knowledge discovery application is critical, we have to follow the context message first step, there are examples of drug safety issues first highlighted with KDD or data mining method which gone beyond clinical review which need to be communicated to the community and have later been supported by some other publications as literature or changing to the official safety information, So yes ,there are range of examples where identify patterns do corresponding to new knowledge.
R: So how relevant is data discovery, on a scale of 1-10?
N: Sorry, in what sense, relevant to whom?

R: useful for the decision makers?
N: I think is difficult to say I mean it is certainly very important part, I mean in our case there is no way that our domain experts could go through the whole database searching for information for the decision makers, so it is critical, we need to screen the data otherwise we could miss important information, so what we do is quite important. I am not sure I am ready to put them on a scale but I think 8 or 9, but I think it is very important for the decision makers.

R: How such analytical tools support and improve decision making?
I: Well I think it is important to support decision making, I think the most, and most important part is it that helps us to direct the resources. It helps us, to focus on those decision or those issues that or most likely to be decision or going to be and then show that there is something we should follow up. They are important in supporting the decision making through limiting the number issues that the domain experts have to look at.
And also some KDD (data mining) method will highlight other aspect of the data which can of course impact decision making. Quantitative pattern can indicate that there is this, and might explain the general tendency of this drug for the same substances in the data, and it says they are all report for young children. Yes I mean in that sense could also impact directly the decision making I suppose…

R: So it depends actually on the meaning of the data? If you discover some new knowledge that has really an impact on some subject or people, then it really affects the decision making.
N: Oh it could, in long term it could of course if you mean long term, yes, absolutely, if we find or discover a new suspected side effect, not directly but in the long terms, it could affect that drug stays on the market or not.

R: How accurate and relevant the discovered knowledge is?
R: I suppose that you answered this question before.

N: Yeah I think, as I said, before we spent a lot of time to analyze the data and which produced a lot of findings but with KDD (data mining) it is easier now. Also we have very problematic data sets; I mean it's also related to the data quality issues so often many of the content patterns can cross from data quality issues as well.

R: How do you manage the knowledge created or discovered?
N: Well, If we go beyond say something is highlighted with data mining methods and then clinical reviews is going through them say that well there is probably something or this is something old enough that we want or something we have to communicate then what we will do is to publish in a restricted document that that goes to national centers so we will communicate our findings to the national centers that provided us with the data.

R: How the knowledge is presented to the users?
N: we present the findings by statistical measures, one of the statistic measure that we used which measure the degree of association between two events, we also strive to present the underlying data so we would not just say there seems to be some association but we would say there is association indicating that the observed expected ratio is 2 so there is 2 times many reports as we we're expecting and the observed number of reports is 20 and expected number of reports is 10 It depends on the specific peace of knowledge discovered.

R: do you use any kind of visualization tools to ease the way of understanding this data?
N: We do as well, I mean we have visualization as well; we look at trends over time which are displayed graphically, also of course if you look at the general distribution of sex, the age distribution for example that would be presented graphically as well, so I guess a mix of numeric data and graphically information.

R: How do you transform the decision maker requirements into rules?
N: The domain expert are not guided by rules, they are quite free to do their vector of analysis in a way that they see fits so there is no sort of strict rules for how they act I mean we want to use their capabilities to see what is best in each case I mean it's not regularized in that sense.

R: So what I understood from you is that there is no relationship between what could be needed from decision making aspects and what is being produced?
N: What I think is that there is quite good relationship that's why we produce this information for them, to be able to do good analysis but there are no sort of strict standardized operating procedures for how they should exactly or what steps they should take for analysis, there are recommendations or guidelines for how to do the decision making, but there is no strict rules they have to go by, I think u can't have that in real world data analysis, you would use the benefits of having someone with the domain expertise, they need to be able to go outside of the box and do the analysis as appropriate for each given situation, I think you lose if you trying to regularize too much, you going to miss things because you just can't predicate anything, that's point of knowledge discovery if you want to detect the unexpected and you need to leave some flexibility, I think that's risks over regularizing it or restricting them in some sense.

R: What are the major departments that use the discovered knowledge?
N: So that would be the national centers, the ones providing the data and in-house it would be the group that is responsible for doing the drug safety analysis, so I mean you can refer to them as drug safety group.

R: What is the exactly usefulness of the knowledge discovery in data base (Data Mining)?
N: I think as I said before there are two important aspects first of all it's a necessitate which we can't possibly go through all the incoming data, it's a necessity to sort of focus the attention on most critical issues and then secondly to detect pattern which may not be directly apparent to domain expert who goes through the data so it last them to look at things which would otherwise may not. So two things really I mean, focusing on the right issues and discover things that they may not otherwise picked up on. These are the major benefits to us

C2.Transcriptions of the Interview with Lennart Waldenlind- Swedish Medical Product Agency (Interviewee B)

R stands for Researcher; **L** stands for interviewee 2

R: As my colleague introduced our purpose of this interview, we have 11 questions about the decision maker, how the decision makers used the knowledge from Knowledge Discovery in Database (KDD) to assist in decision making process, firstly can you give us your background and experience on the use of Knowledge Discovery in Database?
L: My responsibility is to manage the signal detection work within the Drug Safety Department at the MPA. To manage signal detection includes detecting and evaluating signals extracted from data warehouse through Knowledge Discovery techniques.
R: What types of decisions are made in the agency?
L: And then so called approval of new drugs, and then following up already approved drugs. And so, then you can say we have three important things, first is Quality of products and that is regarding the manufacturing and what is the quality really of tablets, suspension etc, the second is efficacy's and then the third is the safety of the product. So we have three major things you can say, and they are quite different, I mean for the approval of the quality is for the chemistry and physical principles and that I think you can more easily investigate by different methods. Regarding the efficacy then you have to indicate trials for patient with this diseases and then you have some standards you're investigating to see if the drugs isn't better that plausible usually that's sugar pills that's tablets without anything and you compare and you have them blind. Patient and physician do not know and then afterwards you make statistics on the results. And then you can see if drugs was ethically statistically significant and of course this significant also need to have clinical important it's not only enough statistical significant also need to have changes or the difference that is clinical useful
And the third is safety and this is most difficulty safety begins study on animals to see something with them in forms of kinetics and effect of drugs on animals, and then you start with patients, then small group of patients where you study the drugs intensively. If everything is alright then you expand investigation to more and more people and usually you have many thousands (1000k) involved which you evaluate before approval. For the safety is not valid because there so many things that they popped up you're not aware of, then you follows the drugs all throughout the market by following the reporting of adverse drugs reaction from WHO.

R: Are this decision related to Research &Development?
L: It is, and of course if you take the quality it is easier because you have some standards you can measure for example if the substance has contents it's proven you can measure. For Efficacy you run the studies and the statistics, if the study is relevant you can make conclusion from that. From the safety is more difficulty because is not so clear. You can see common side effects when study are ready .For uncommon you cannot see, and you don't have all patients populations so this is more difficulty, and then you make synthesis there is a lot of different data from animals, from clinical studies and from use of the products so that is global evaluation for the safety you can say for safety.

R: considered the policy that is used, is it developed here in-house or?

L: You can say is the policy that comes time by time in Europe. We have control risk benefits. We have risk and benefit. Benefits should be greater than risks. And that's very different from different drugs let's say if you take cancer drugs then you can accept a lot of side effects because disease is so serious. But if you use drugs that should lowers the cholesterol level then just when you treat healthy person that they will have less disease on the future then you cannot accept side effect so much at all ,then it's very different from the types of drugs you have ,and the concept is the risk benefits.

R: What if there is a new disease, is there special treatment/procedures?

L: It could be, if you do have disease where have no treatment earlier then you can accept little more risks, than if you have already treatment, then it can be harder on the new one. But let say if you have HIV for example and you had no treatment for that. Then there was a lot pressure from United States that the approval should go quickly and we did that in those times because we had no treatment .Today we have treatment, so today it would be probably harder when we evaluate because patients have some sort of treatments. So for examples for swine its difficulty to say of course is very dependent on serious of the disease, if swine would be seriously let's say killing 10 of diseased people. Then you need to take something very quickly and accept higher risks.

R: Ok. Does the decision depend on environment severity of disease?
L: yes, and is always risks benefits. Let's say the flu would be very serious killing 10 % ,then if you have treatment the benefits is huge because you save lives by treatment ,but let say if the flue only kills 1 person or 1000 then of course that would also be important,, but as not important as 10 % ,then you can be more carefully when you develop, you must look more to the risks, the risk will be more important than relative than if the flue would be killing 10 percent for examples..And that was example before your treatment you could take may be some your risks. But today you're not willing to take the same type of risks. And you're not willing to take any risk that is higher than the current treatment. So in principles we want the risks to be lower than the current treatment. Then Since the risks is not 100 obvious clear there always hesitate when we are approving and that we need to following up them because there could be serious things coming up and we must absorb them to change the efficiency. We want to be so sure we can when we approve the drugs

R: What is the general decision making process agency in your department?
L. We are involved with, we have different types of drugs, national approval drugs and we decide our self here. But many drugs now are centrally approved. And that's means they are approved at all Europe at the same times so then we work on European environment. Then the decision making process is European. We have Committee for Medicinal Products for Human Use (CHMP) a board where all experts from all the countries meet and they decide, the we are part of the team.
R: is there a decision making framework?
L: Yes, of course in fact even if they are national approved, drugs are usually existed in other European countries. The process will also be European. So if we have problem here, we analyses the problem here and discuss internally and usually is taken to European boards for discussion, since if we think it won't be in market it shall also not be on the market in other countries as well. But we have the formal rights to decide our self. Let say, if we don't want to agree with UK we can decide our self. For central approval there is common decision.

R: Is there a specifically details criterion to follow up on taking decisions?
L: You cannot say specifically criteria. Is just a risk benefits always and that varies so much between different drugs so .You cannot have specified criteria because it so different for different drugs. If you treat serious diseases for example you may accept side effect .If you treat not severe disease you can't accept so much and this also depends on what other treatment around the markets, other treatment that are better that you can't accept anything that is worse. That is whole situation, is risk benefit. You look at disease type and other current treatment

R: Do you have some criteria or benchmarking for approval of drugs?
L: Approval of drugs is the preclinical findings, clinical findings and clinical trials that we have around in the new drugs because that we do not have market experience that here we involve with all the drugs. So when are in market we follow them and the criteria is always risk benefit. So it is so difficult to specify criteria. We can also say that criteria may varies on time

R: Is there a factors interfering the approval of drugs?

L: Yes, also the media maybe very important. Homosexual group in US press very hardly on earlier approval and that may also make the politician in US to work more quickly for approval for the drug .so political pressures But politician can not only say on political grounds but general principles can be important.

R: so, can say you call this turbulent environment??
L: And that's the picture at the moment and that may change with time..it may change for example let's say you have drug with risks,, let say you have drug for HIV ,they have risk today. But let say new drug was coming without any the risks we have seen, the risk benefits of the new drugs will be lower since we have new now that are better. Also, there are problem like in developing countries that sometimes cannot pay for drugs. So this brings discussion for how to handle this situation.

R: Speaking on the decision that you normally takes, what is the degree of structuredness of such decisions?
L: We have such, I mean when we look for new signals, when methods look for new signals for example and they such structured methods, but when we have such findings we cannot be sure that at least we have decided that it could depend on other factors as well. So we have structured methods but we need to discuss the result in groups to evaluate because they are so many things that can influence the interpretation of the findings. So they are both structured principles but then final decision is not unique, is sort of global decisions. And if u take this benefits again you cannot weigh exactly let say one patient in 1000 gets depression may be serious disease and then average weight loss is 2kg how do u know it , I meant is not the same thing, sometimes is a matter of taste. Is it worth this 2kgs? Is it worth 1 patient in 1000 .that's question is not so easy. Because it depends on your own interpretation .You need to discuss with different people having different background. Its decision that take consideration many factors

R: Do you use the Knowledge discovery techniques such as KDD (data mining) ?
L: yes, we use the data mining when we look at new side effects, we have this database with the side's effects and then we have the data mining looking for new signals (patterns)

R: How do you use knowledge from KDD to identify a problem or opportunities in your department or organization?
L: first we identify the signals which usually is a statistical signal, and then we take that for a group decision every week and then we discuss according to the table what could be the reason for this, sometime it is just the background of the disease itself so it is not the drug, it is something else and if it is not so serious we can wait for more reports to be more convinced but if it is very serious we need to act more rapidly , so this signal it taken for more discussion ; there is always the risk benefit that lead the decisions .

R: How can data mining help in treating previous diseases?
L: if there are serious side effects, then you should not use this drug so much maybe you should use in very severe patients, so it may lead of less use of this product.

R: Did you ever face a situation that the output of the data mining identified an opportunity in you agency or it identifies just problems?
L: this is a matter of how you look at the problems, I mean they are all opportunities because if you see a problem with a drug and make the right decisions then based on that you can improve the health of the people, it is always positive for people but it is negative for the manufacturer, for us it is an opportunity because If the information is correct we can have better health protection.

R: How knowledge discovered assist you in developing and exploring the courses of actions for your decision problem(s)?

L: yes, let's say that you have some sort of side effects that can be explained by the action of the drug for example, then off course that can improve your understanding of the drug, but sometime we don't know, I mean you have a side effect you can't explain it but it is there, it may take some time before you start understand why do you see this.

R: How do you develop the solutions for a problem?
L: it is the risk benefit again, the solutions are more/less, if you take the lightest one; if you have a problem you just write it in the prescribing information, Let's say you discover you have a head ache in the first day you take the drug , then you write it in the information inside the package insert, the patient can see that he can get a head ache in the first days, then you can say that you increased the risks a little since you can also get a head ache, if that is not so severe then it does not matter so much, but then you may also for example If it is more severe you cannot treat this type of patients, you can treat only severely ill patients then you have a reduction of risks, and the last thing is that we redraw the drug from the market, so there are always different levels.
We need to put all the problems to the correct level of action. One level of action is that we are not sure because it is not black or white it is grey , so we cannot know always if it is a side effect because we don't have enough knowledge at this moment we need to wait in order to have more knowledge and if the problem is not so severe we can do that (wait), we can wait half a year more to see if we have more reports on the problem which is extracted by the data mining, to look at the data mining after a half a year again for example. The more severe the problem is the more rapid you need to act even if you don't know but if it is not a severe problem then you can wait and have more facts before you act.

L: the problem is also or what we call sensitivity and specificity, if the data mining is very sensitive you will get a lot of signals and many of them are not real and that is a problem and if it has less sensitivity you don't get so many signals and maybe lose some of them so it is a balance.

R: Whenever you get signals, how do you evaluate these signals?
L: we have a criteria, the first one is if they cause death that is the most important and the second is if they are serious, for instance if they lead to hospitalization or make the patient handicapped or cause a cancer and the third one is none serious which do not cause serious side effects.

R: How knowledge discovered assist you to evaluate and select the best course of action for your decision problem (s)?
L: if data mining gives us the more correct course for the problem, it may have and it may not.

R: did you face this kind of situations?
L: yes, let's say you have liver problems for example and they are not serious then it is maybe enough to add this to the prescribing information explanation but of-course if they cause very serious liver problem then you may have to withdraw the drug from the market.

R: How is the knowledge being used in forecasting and/or predicting the unforeseen situations?
L: we have the risk management plan for drugs, we try to foresee things drawn from the earlier data mining problems, let's say you have seen small liver problems that are not so sever then in some patients these are maybe very important and then you could expect some sever cases you will follow if you get them or not also some problem you cannot foresee because they come as surprises.

R: So when you do the prediction, do you use data mining?
L: it is one of the methods. If you take the liver example, if you find liver problem in data mining but you can also find liver problems in animal studies or when you perform clinical trials and that is not data mining so we look to different sources.

R: What is the perceived usefulness of knowledge discovered through knowledge discovery tools in decision making process?
L: it is very useful, its help us with 30%-50% in the decision making because the other ones are global such as the risk benefit, type of disease and so many factors you need to consider that are very important as well, data mining is one factor for instance if you have 10 factors for decision data mining is one of them.

R: could data mining help you in focusing the attention to a specific area?
L: yes if data mining discover liver problem, you could ask the company to run a study for that. This is how data mining could lead to some actions in resources direction and decisions. But it is not the golden standard, I mean it does not solve the problem but it is one of many factors.

R: you mentioned that data mining assist you in the decision making by 30%, do you think that this number is a good contribution?
L: yes it is a good number, everything we can have is important

R: how do you compare the data mining with other techniques used in decision making?
L: we need everything, the decision is global, and we need all facts on the table we can't say it is less or more important, it is very important.

R: what is the contribution of data mining on the decision making?
L: it can be very large and it can be minor depending on the content and signals identified, so you cannot say that it is generally very important I mean it is important but it varies very much on the case. Sometime the problem could be better understood by running a clinical study in a few patients for example, so it also depends on the problem.

R: so data mining is used whenever you want to investigate a huge population?
L: yes huge population with unknown effects.

R: Do you need external information to assist you in decision making?
L: yes we need, we have the cooperation with different European countries and that is very important (CHMP) that is very important, it is discussed among experts, sometimes we need to take external experts to hear their view on things because it is maybe complicated.

R: Is the communication is being performed through reports, data into database…?
L: yes, reports, data and meeting because you need to discuss the problem.

R: Do you share the data in your databases?
L: yes we share it but they are not 100 % opened for everyone.

R: Do you have any cooperation with Uppsala monitoring center in aspect of data mining?
L: yes, I mean if they discover signals we look at them and since they are located here in Uppsala so it is easy to cooperate but we could have better cooperation and the relation could be developed.

R: Do you have internal data warehouse and data mining specialists?
L: yes, also we have the European database that is called eudravigilance.

R: Are the sources of data in the data warehouse just local or global also?
L: it is global, so everyone needs to report these effects, so it is from all Europe to start with and for all drugs that are approved in Europe and sold in other countries then all serious report are coming in our data warehouse then we apply data mining on the data to discover signals.

R: what was the motive of employing the data warehouse (DW) in the agency?
L: we have this common database for all the European countries, the DW is developed by the EMEA in London so we have cooperation with them, they have also an expert group working with this DW and I am a member of that expert group as well, it's been developed since 1995, but before that the country was reacting more separately, there was a problem before that in aspect of integration of common data, but now the data is coming continuously from everywhere so the problem disappears. So the data warehouse solved the problem of having a scattered data all over.

C3.Transcriptions of the Interview with Y- Public Owned Company X (Interviewee C)

R stands for Researcher; **A** stands for interviewee 3

R: What is your experience and background in the use of Knowledge discovery techniques?
X: as at the moment am not employee of the Public Owned Company X, but I used to be three years ago, so I had been working in IT department for four years from 2002 to 2006, then I had different roles but when I left the Public Owned Company X, I was responsible for business performance system. Before I joined the Public Owned Company X, I used to work with Oracle in data warehouse and Business intelligent (BI) solutions. So, I left the Public Owned Company X and continued to work as consultant with Bizintel responsible in data warehouse and Business intelligent (BI) solutions. So am now working in Apoteket as consultant also responsible for data warehouse and BI solutions.

R: Ok, thank you very much. Lets starts with data warehouse section, what are the sources of data in the data warehouse? (Name Internal and external sources)
X: yes, we have of course both internal and external; we have a lot of data, all the sales figures, products and financial data.

R: What type of data warehouse have you implemented in your organization? It could be Data marts or Centralized data warehouse or Real Time Data warehouse
X: We have three systems. We have something called ABS which is financial data mostly, and then we also have something called XPLAIN that's mostly sales figures. We have also XPLAIN PLUS which is recently developed, no one has managed it properly, and this is used more for assortment and stock analysis, mostly logistic. If you look at their dimensions they have a lot in common, they have same sources from sales figures, the Public Owned Company X stores or e-business sales. So all the sales records from different channels are fed into our enterprise data warehouse. Also we have products dimensions which are loaded from product databases and also we have ERP systems which contain some accounts. We have efficient ETL process, more because now we have same data here and here, the data from all channels are put into the same database called T-base before loaded into data warehouse.

R: ok, thank you...The third question is how old is the database?
X: Some data are being loaded every day, some weekly and some monthly. The main party is loaded monthly but everyday sales figure for examples you load them everyday

R: do you consider the data in DW as historical data?
X: Yes, We started feeding this warehouse with data in beginning of year 2000, so we have historical data.

R: ok, Thank you, the next section is the Knowledge Discovery Tools, the first question will be what are the analytical techniques employed to access data in data warehouse. There is OLAP, KDD/ Data Mining, data visualization tools
X: yes of course, we have OLAP, reporting services (Ms ROLAP) used, we have portal, dashboards, also stuff in add on excel and data mining. Mostly used is OLAP. Data mining is only used in market department; we have small application in market department that we employ data mining. We have some part of the system that we extract information from data, there is an external company they perform analysis for us which is mainly data mining, then the marketing department use it in their activities but we don't have this tool here. So we extract the data set from our data warehouse and send it to them to perform the mining. So we outsource the part of the data mining to another company.

R: do this data mining techniques extract useful knowledge that you use in your activities?
X: yes, they are analyzing the marketing campaign and take a lot of campaign (sales campaign) decisions based on that.

R: Do the techniques that you previously mentioned (OLAP, ROLAP and data mining) how do you think they do help the decision maker in making decisions?
X: yes, it give them a lot of information that they use to identify trends and perform forecasting, which is most case on campaign planning, they use previous experience from old campaign results, they also identify the factors that affect the campaign like the weather , purchasing behavior and the distance since Sweden is big country, which is being performed by this external company that is performing the data mining and discover patterns, then they collect all the information and discuss that together with a marketing company that uses this information together with the market department people so then they take decision based on the knowledge available to launch new campaigns in the future.

R: can you mention some problems that these techniques solved or what are the possible problems that could appear without these techniques?
X : generally speaking, if you can't analyze your data you will make wrong decisions or make decisions too slow or you don't get the knowledge in a format that could be understood, we need to have information like now this daily information (sales information) that was asked by us from the stores we need them every day, we need them very fast because we need to control our sales process, for instance they use the data to compare the activities of different stores in order to identify the problems. Also it gives the stores their actual status compared to others which create a competition hence boos the productivities, this technology is useful for each store manager to keep an eye on the store activities.

R: so it enables them to take a responsible and timely decision which is important for the management?
X: yes, they need to be very reactive to what the market asks, and also they would like to be proactive. In other words they try to react before the market.

R: Do you think that this information that is extracted from OLAP and KDD increases the understanding of the market problem or the internal activities problem?
X: yes... so that is one part. Then we have the explain part for the sales statistics, we were obliged to produce statistics about the sales from different context for the government (different organization), this data was stored in a data warehouse.

R: How accurate and relevant the discovered knowledge is?
X: we have a very good data quality, which is due to the fact that for instance the data for the sales is coming from another database before they goes to the data warehouse so there is a lot of cleansing also the data entered by the people working in the pharmacy is very precise due to the fact that the data is related to peoples life so precision is demanded. So because of all these factors we have a very high data quality in our data warehouse hence we have little mistakes in our reports. The users are very happy with the software services and they always ask for more.

R: how do you manage the knowledge extracted from the data warehouse?
X: yes the users have the ability to store their output and share it with others.

R: how do you present the knowledge to the users?
X: we use reports, graphs, cubs, dash-board, and score-cards, statistics data

R: is there a relation between the decision maker need and the provided services?
X: the decision maker provide the business rule which will be stored inside the application in a certain way.

R: what are the major departments that use the knowledge provided from these Knowledge Discovery tools?

X: Do you mean the ones that use the information mostly?

R: yes, the ones that much real need this data
X: operational, tactical and strategic information so the knowledge is been used in different department at different level, But in general the financial, economic and statistical departments.

R: what types of decision being made in the agency?
X: Well, for instance we have budget, policy making, forecasting and then you have of course I mean different kind of decision, it depends on how you control your daily business. For examples if we see some stores are unprofitable then we have to make the decision to close such stores now. If you have kind of BI or DSS you can have chances to look at it in different ways. We used to have more balance score cards (BSC). Only one part of business still works a lot with BSC

R: How this knowledge discovered from data mining helps decision makers to identify problem or opportunities?
X: yes, a lot, they help in analysis of market campaign market and plan for future campaigns. They can go further down and analyses the data more and they take action after what they see. So that's important part of their daily work.

C4.Transcriptions of the Interview with Dr. Anette Hulth- Swedish Institute for Infectious Disease Control (SMI) (Interviewee D)

R stands for Researcher; **A** stands for interviewee 4

R: What is your background and experience on Knowledge Discovery techniques?
A Ok. So the application that I will describe to you, that its system for computer supported outbreak detection. If you have any thing you don't understand just ask me straight away. So, I have been working at SMI as a researcher for bit more than 2 years; I have background in computer and systems science and I have been involved in this project of computer supported outbreak detection since it started. That's the project I started at institute. We have a systems which give an automatic alert alarms when the level of any of notifiable disease are in Sweden has reached the level that might indicate that there is outbreak is going on. So there are number of 63 notifiable diseases in Sweden at this moment this seeks little from year to year. And everybody, the doctor that sees the patient, and if this doctor suspect that the patient has so infectious disease any of this notifiable disease, he or she is obliged to report both to SMI and also to local county medical officers. So we get this clinical reports from the doctors in the country, and every reports contains some information on the patient also we get info because the doctors want to have confirmation if it is real this disease or not, in most case she or he can't be sure. So she will send test to laboratory and also this result from laboratory test must be reported to SMI. So in our database we have got a plenty of records of the notifiable diseases, this report their cases. And then we are applying data

mining methods to try to find variation to find something is not normal. Data mining models comprise of statistical methods .They are number of methods that have been developed by other peoples for, some have been developed for other purposes and some developed more for infectious diseases in mind, there is a method called SatScan which is a method for geographical, spatial temporal cluster analysis. So you will cluster the case and will try to see if you have more case in within the spot or outside of the spot. So you have to make circles. These are statistical algorithm that takes geographic into account then you have particular point geographical point from which you start to make this cluster so the point will be the centre of each this cluster.

R: What are the sources of data in your databases?
A: Ok , so these are the case reports, the reports that we get from medical doctors across the countries for the notifiable disease and reports from laboratories that verify if the sample that taken from the patients is indeed final of particular diseases .

R: Do you require some data from other agencies??
A: We do buy data from Swedish statistics from time to time when we need for different projects or we will use the data that are readily available, we also have number of cooperation with other agencies for example National boards of health and welfare and SMI work closely. But when it comes it comes to this particular KDD application then we don't have any cooperation really. Then the data that is collected from this database is also accessed by local medical offices as wells .So that is database you will go if you want to know how many people suffered from Chlamydia last year.

R: What kinds of decision are made in the agency?
A: We don't have any policy decision to make at SMI that is not our task. Our task is to surveillance, to keep track of the status for the infections disease in the country and also suggest what could be done by the decision makers, so SMI is an expert agent

R: ok, Do you use this information from KDD to assist the decision makers in your agency?
A: yes, but this information from this application is just one small part of work being done here, so it's more aid to epidemiologists since they do different diseases and because there is a large amount of data they need to look at and sort of making them focus more on what is probably relevant by getting these alerts so that they know ok we have a lot of chicken pox last week and can look into that maybe at particular account or particular part of the country.

R: How do you use knowledge from KDD to identify a problem or opportunities in your department or organization?
A: yes definitely, it helps in outbreak detection and which is the mission of SMI, it is one of the core things that we do. As I said previously, there is a lot of manual work that needs to be done and the KDD (data mining) is been a great support, and the manual work is mainly used to validate and evaluate the knowledge discovered by the data mining. We use experts (epidemiologists) to evaluate the outbreak discovered in order to formulate appropriate actions, and this evaluation is being done manually. And this actions would be a kind of communicating these outbreak discovered with other experts or decision makers.

R: Does these techniques help you in reducing the time in looking for outbreaks?
A: yes, it is a tool that we need because we have many diseases and since the KDD (data mining) helps us in the directing the resources in a timely manner.

R: How knowledge discovered assist you in developing and exploring the courses of actions for your decision problem(s)?

A: yes, it helps us to find the outbreak, and this technology is good in contacting the doctors. For example the doctors can say that they don't have problems with salmonella, so we say that our tools says that there is an important signals regarding this problem and we rely need to investigate this problem. And sometimes we present this information to the doctors when it gets more serious. In other words it gives us a starting point that the normal scan could not identify.

R: What is the perceived usefulness of knowledge discovered through knowledge discovery tools in decision making process?
A: it is again the outbreak, I mean that it help us to decide if there is an outbreak or not, it also help us in managing our resources by pointing out what to investigate and finally, it is the time reduction since it is hard to go through all the available data to identify the outbreaks.

R: What are the limitations of the KDD (data mining)?
A: in our particular area, it would be very good if these techniques could follow up the signals generated, and if they could help in validating these signals generated. Also we need to reduce the false alarms generated by the system, which could be done by creating a loop-back to the system which could create an internal validation before presenting the alarm. These issues are common in other institutes for disease surveillance.

R: based on this conversation and your activities, I suppose that you have old data (historic data) stored in your database?
A: I believe we have data from the sixties. It also depends on the disease; if it is chronicle disease or it is a temporary disease.

References:

Alshawi, S., Saez-Pujol, I.,Irani, Z. (2003). Data warehousing in decision support for pharmaceutical R&D supply chain. *International Journal of Information Management*, 23 (3), 259-268.

Al-Zoabi, Z; Al-Noukari, M. (2008). Human and Electronic-Based Knowledge Enablement in EGovernment. *3rd International Conference on Information and Communication Technologies: From Theory to Applications*, 1-4.

Ang, J., Teo, S H. (2000). Management issues in data warehousing: insights from the. *Decision Support Systems*. 29 (1), 11-20.

Bach, M.P. (2003). Data mining applications in public organizations. *Proceedings of the 25th International Conference on Information Technology Interfaces* (ITI), 211-216

Bhowmick, Sourav S., Madria, Sanjay K., Ng, Wee K. (2004). *Web Data Management*. USA: Springer.

Bieber, M. (1998). Data Warehousing in Government. Available at: https://www.msu.edu/~biebermo/Data%20Warehousing%20in%20Government.pdf. [Accessed 27 March 2009].

Bloor, M (1997). Techniques of Validation in Qualitative Research: a Critical Commentary. In: Miller, G. and Dingwall, R. (Ed's.). *Context and Method in Qualitative Research*. London: Sage.

Bots, P., Lootsma, F. (2000). Decision support in the public sector. *Journal of Multi-Criteria Decision Analysis*, 9, 1-6.

Chaudhuri, C. & Dayal, U. (1997): An overview of data warehousing and OLAP Technology, *SIGMOD Record*, Vol.26, No. 1, pp.65-74.

Chen, G. (2008). A Design of Data Rebuilding for Decision Support in EGovernment System. *The 9th International Conference for Young Computer Scientists*, 965-970.

Codagnone, C., Wimmer, M (2007). *Roadmapping eGovernment Research: Visions and Measures towards Innovative Governments in 2020*. [E-book]. Italy: MY Print di Guerinoni Marco & C Via San Lucio. Available at: http://*www.egovrtd2020.org/EGOVRTD2020/FinalBook.pdf* [accessed 15 April 2009]

Connolly, T., Begg, C. (2004). *Database Systems, A Practical Approach to Design, Implementation and Management,* 4th ed. Addison Wesley.

Creswell, J. W. (2007). *Qualitative inquiry and research design: choosing among five traditions.* 2nd ed. Thousand Oaks Calif: Sage.

Daft, R (2004). *Organization Theory and Design.* 8th ed. USA: Thomson South- Western.

.

Dror, Y. (1997) Strengthening government capacity for policy development. *International Journal of Technical Cooperation ,3*(1): 1–15.

Dunn, W (1994). *Public Policy Analysis: An introduction*. 2nd ed. US: Prentice Hall.

Erdmann, M. (1997). The Data Warehouse as a Means to Support Knowledge Management. *Proceedings of the 21st Annual German Conference on AI '97.*

Fayyad, U., Piatetsky-Shapiro, G., Smyth, P. (1996a). Knowledge Discovery and Data Mining: Towards a Unifying Framework. *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining*, 82-88.

Fayyad, U., Piatetsky-Shapiro, G., Smyth, P. (1996b). From Data Mining to Knowledge Discovery in Databases. *AI Magazine* , 17(3), 37-54

Goh, D., Chua, A; Luyt, B., Lee, C. (2008). Knowledge access, creation and transfer in eGovernment portals. *Online information review: the international journal of digital information research and use,* 32 (3), 348-369.

Han, Q., GAO, X. (2009). Research of Decision Support System Based on Data Warehouse Techniques. *Second International Workshop on Knowledge Discovery and Data Mining,* 215-218.

Harper, M. (2004). The Data Warehousing and the organizational of Governmental Databases. In: Pavlichev, Alexi and Garson, David G. (eds). *Digital government: principles and best practices.* USA: IGI Publishing. 236 – 247

Hipel , K.W., Radford, K.J., Fang L. (1993). Multiple participant-multiple criteria decision making. *IEEE Transactions on Systems, Man and Cybernetics,* 23 (4), 1184-1189.

Holsapple, C. W. (2008). Decisions and Knowledge. In: Holsapple, C. W; Burstein, F. (eds). *Handbook on Decision Support Systems.* 2nd ed. Berlin: Springer.

Hoss, D. (2001). Top Ten Trends in Data Warehousing. Available at: http:// www.information-management.com/infodirect/20011026/4191-1.html [Accessed 20 March 2009].

Hovy, E. (2008). DATA AND KNOWLEDGE INTEGRATION FOR. In: H. Chen, L. Brandt, V. Gregg, R. Traunmueller, S.S. Dawes, E. Hovy, A. Macintosh (eds). *Digital Government.* US: Springer. 219-231

Huber, G., McDaniel, R. (1986). The decision-making paradigm of organizational design. *Management Science*, 32(5), 572–589.

Inmon, W. H. (1996). The data warehouse and data mining. *Communications of the ACM,* 39 (11), 49-50.

Inmon, W.H. (1993). *Building the Data Warehouse*. 4th ed. NY: Wiley.

Israel, M. & Hay, I. (2006): *Research ethics for social scientists: between ethical conduct and regulatory compliance.* London: Sage.

Khorshid, M. (2004) "Model-Centered Government Decision Support Systems for Socioeconomic Development In The Arab World", *Proceedings of The International Conference On Input-Output and General Equilibrium: Data, Modeling and Policy analysis"* Brussels, Belgium, 2-4 September.

Kvale, S. (1996): *Interviews: An introduction to qualitative research interviewing*. Thousand Oaks Calif: Sage.

Layne, K., Lee, J. (2001). Developing fully functional eGovernment: A four stage model. *Government Information Quarterly,* 18, 122-136.

Matheson, J., Matheson, D. (1998). *The Smart Organization: Creating Value Through Strategic R&D.* USA: Harvard Business School Press.

McClure, D.L.(2000) Electronic Government Opportunities and Challenges Facing the FirstGov Web Gateway Available at: http://www.gao.gov/new.items/d0187t.pdf [Accessed on 27 April 2009]

Menon, A., Varadarajan, P . (1992). A Model of Marketing Knowledge Use Within Firms. *Journal of Marketing,* 56 (4), 53-72.

Mintzberg, H. (1977). Book Review of the New Science of Management Decision. Rev ed. (by H.A Simon). *Administrative Science Quarterly,* 22 (2), 342-351.

Misra, D. (2007). Ten Guiding Principles for Knowledge Management in EGovernment in Developing Countries. *First International Conference on Knowledge Management,* 1-13.

Mitra, S; Acharya,T. (2003). *Data Mining: Multimedia, Soft Computing, and Bioinformatics.* New Jersey: Wiley.

Nabukenya, J., Van Bommel, P., Proper, H.A. (2009). A Theory-Driven Design Approach to Collaborative Policy Making Processes. 42nd *Hawaii International Conference on System Sciences* , 1-10.

Norris. Fletcher. Holden, S, (2001). Is Your Local Government Plugged In? *Public Management*, 83 (5), 4-11.

Nutt, P. (2006). Comparing Public and Private Sector Decision-Making Practices. *Journal of Public Administration Research and Theory*, 16 (2), 289-318.

Ogut, A., Kocabacak, A., Demirsel, M. (2008). The Impact of Data Mining on the Managerial Decision-Making Process: A Strategic Approach. *The Journal of American Academy of Business* , 14 (1), 137-143.

Pandya K., Cong, X, (2003). Issues of Knowledge Management in the Public Sector. *Electronic Journal of Knowledge Management*, 1 (2), 25-33.

Prabhu, C (2006). *E-Governance: Concepts and Case Studies.* India: Prentice-Hall.

Radford, J., Hipel, K.W., Fang, L. (1993). Strategic and tactical analyses in complex decision situations. *Proceedings of the International Conference on Systems, Man and Cybernetics*, 1, 143-146.

Rifaie, M., Kianmehr, K., Alhajj, R., Ridley, Mick J. (2008). Data Warehouse Architecture and Design. *IEEE International Conference on Information Reuse and Integration*, 58-63.

Sabatier, P (1999): *Theories of the Policy Process.* Boulder, CO: Westview Press.

Savvas, I., Bassiliades, N. (2009). A process-oriented ontology-based knowledge management system for facilitating operational procedures in public administration. *Expert Systems with Applications*, 36 (3), 4467-4478.

Seale, C. (1999): *The quality of qualitative research.* London:Sage.

Sharma, P., Kumar, P (2004). *E-governance: The New Age Governance*. New Delhi: APH Publishing.

Singh, H.A (1998). *Data Warehousing Concepts, Technologies, Implementations, and Management.* Prentice-Hall: Englewood Cliffs, NJ.

Turban, E., Aronson, J., Liang, T-P., Sharda, R (2007). *Decision Support and Business Intelligence Systems*. Upper Saddle River, New Jersey: Prentice Hall.

United Nations Department of Economic and Social Affairs (UNDESA). (2008). UN e-government survey 2008: From e-government to connected governance. [Online] .Available at http://www.ansa-africa.net/uploads/documents/publications/UN_e-government_survey_2008.pdf [accessed 10 April 2009]

Vasu, M; Stewart, D; Garson, G. (1998). *Organizational Behavior and Public Management.* USA: Marcel Dekker.

Verschuere, M. (2009). The Role of Public Agencies in the Policy Making Process: Rhetoric versus Reality. *Public Policy and Administration* , 24 (1), 23-46.

Viktor, H; Arndt, H; Oberholzer, M. (2000). Enhancing Government Decision Making through Knowledge Discovery from Data. *ECIS 2000 Proceedings*, 0 (71), 1-8.

Watson, H., Annino, D., Wixom, B., Avery, K. & Rutherford, M. (2001). Current practices in data warehousing. *Information Systems Management*, 18(1), 47-55.

Wimmer, M. A. (2007). The Role of Research in Successful EGovernment Implementation. In: *E-Government Guide Germany*, ed. by Zechner, A.

Wimmer, M.A. (2002), Integrated service modelling for online one-stop government. *Electronic Markets*, Vol. 12 No.3, pp.149-56.

Winterman V, Smith C, Abel A. (1998). Impact of information on decision making in government departments. *Library Management*, 19 (2), 110-132.

Yager, R. (2008). Risk Modeling for Policy Making. In: Ruan, D; Hardeman, F; Meer, K (eds) *Intelligent Decision and Policy Making Support Systems.* Berlin: Springer. 318.

Yin, R. K. (2003): *Case study research: design and methods.* 3rd ed. London: Sage.