A Soft Systems Methodology Perspective on Data Warehousing Education Improvement

R. Goede, and E. Taylor

Abstract—This paper demonstrates how the soft systems methodology can be used to improve the delivery of a module in data warehousing for fourth year information technology students. Graduates in information technology needs to have academic skills but also needs to have good practical skills to meet the skills requirements of the information technology industry. In developing and improving current data warehousing education modules one has to find a balance in meeting the expectations of various role players such as the students themselves, industry and academia. The soft systems methodology, developed by Peter Checkland, provides a methodology for facilitating problem understanding from different world views. In this paper it is demonstrated how the soft systems methodology can be used to plan the improvement of data warehousing education for fourth year information technology students.

Keywords—Data warehousing, education, soft systems methodology, stakeholders, systems thinking.

I. INTRODUCTION

STUDENTS in information technology (IT) study data warehousing (DW) as part of their fourth and final year. Most students experience some kind of paradigm shift when doing this module. They start thinking about IT from a wider organisational perspective. The DW module must be designed in a way to facilitate the development of the students' understanding. It also needs to cater for the expectations of various role players. A holistic understanding is therefore required. The systems thinking movement developed from the need to have a more holistic understanding of problem situations. This paper aims to show how the problem situation of module planning in data warehousing can be improved by using systems thinking ideas and specifically the soft systems methodology (SSM).

The paper provides a discussion on data warehousing and the specific characteristics of DW education in section II. Section III provides background on systems thinking and the SSM. As the aim of this paper is to demonstrate how the SSM can be used in DW module planning, section IV provides a discussion on the aspects of SSM applied to this problem situation. Conclusions on the advantages of SSM in this situation are given in section V.

Roelien Geode is with the Vanderbijlpark Campus of the North-West University, South-Africa (e-mail: roelien.goede@nwu.ac.za).

Estelle Taylor is with the Potchefstroom Campus of the North-West University, South-Africa (e-mail: estelle.taylor@nwu.ac.za).

II. DATA WAREHOUSING AND DATA WAREHOUSING EDUCATION

The aim of this section is to present some background knowledge on DW, including the differences between data warehouses and transactional systems, to demonstrate the shift in thinking required by the students to master the field. The section starts with a brief discussion on the historical development of DW methodology. The section ends with a brief description of the current DW module at the North-West University in South Africa where this research was done.

Data warehousing developed in the 1990s from the need to integrate data from different information sources in large corporations to support decision making. Inmon wrote what later became the pivotal monograph in the field in 1996. He defines a data warehouse as: "A subject oriented integrated, non-volatile, and time variant collection of data in support of management decisions." [1]. He advocates a data-driven methodology where data in the organisation are integrated into a central data store and accessed by end-users through star joins. This process starts with data and ends with requirements. He is sometimes criticized for neglecting the business perspective. This is only partly true since he tries to empower the business user in taking full responsibility for using the data in the data warehouse [1]. In reaction to this data-driven approach Ralph Kimball developed a requirements driven approach to data warehousing, where the development is driven by a business sponsor aiming to seek information to address key business problems [2].

Although the order of activities might differ in the two methodologies the activities are similar. Fig. 1 depicts the main activities in the Kimball methodology:
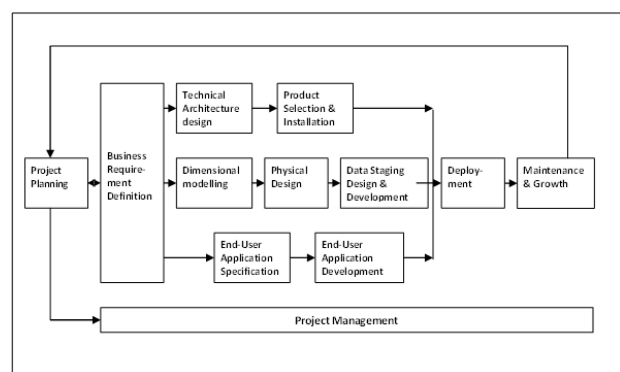

Fig. 1 Kimball lifecycle [4]

From the development perspective there are three main differences between a data warehouse and other information systems. The first is that in information systems used for transactional systems the developer has control over the source of data. Typically data entry systems are developed to support the information needs of the system. In a data warehouse environment data is received from more than one source system typically working in different architecture environments. A large amount of work is done to extract the data from the source systems, transform it in the DW format and loading it into the data stores of the data warehouse. This process is called data staging or ETL (extract, transform and load).

The second major difference is the fact that in a DW environment the focus is on getting data from the system, rather than entering data into the system. In typical transactional information systems great care is taken to minimise the data input time and to minimise the number of place a certain data element is stored. This is called minimising redundancy and students are taught the principles of normalisation of data from very early in their study. The problem with normalised data is that it takes a fair amount of processing work to create a report from the data, especially one that combines a large number of different data entities. In data warehouse design the focus is on getting information out of the system. Quick data input is less important as source comes from the normalised source systems. A fair amount of work is done to de-normalise the source system data in order to improve response time of end-user queries.

A third difference between transactional systems and data warehouses is that in transactional systems users access the data through fixed reports while in DW users access the data also by *ad hoc* queries. These are once off queries business users have to support business ideas. Business users want flexible query environments where they can run queries to find data to support their ideas without working through the IT department in the organisation.

In order to facilitate fast data access data warehouse data models differ from transactional systems. Dimensional modelling is used in DW and entity relationship modelling is used in transactional systems.

Our students are therefore confronted with a paradigm shift in information system development. Firstly, gone are the days of nice data entry screen design, now they should integrate data from different sources. Secondly, gone are the days of normalisation and report writing, now they should understand dimensional models and provide business managers with access tools that are focussed on flexibility.

The methodology of Kimball is used in the data warehousing module at the NWU to the fourth year IT students. The students are exposed to several data warehousing methodologies including that of Inmon, but they perform their practical work according to the work of Ralph Kimball. The business orientation of this methodology is one of the key motivators for this approach. Students entering this module should have mastered many technical skills in their first three years of training in modules such as programming,

data base management and systems analysis. Simply integrating data from different sources should not be a problem for them. The data warehousing module presents them with a different perspective, that of business intelligence (BI). The focus is shifted to how business decisions can be supported from different data sources in and outside the organisation. The students should not only know how to integrate data from different sources but more importantly, they should know how the data will be used to facilitate better business decisions. Whereas the aim of their other IT modules is on mastering the techniques, the aim of DW is to understand the use of BI technology holistically in the organisation.

In the DW module at the NWU, a theoretical course is supported by a practical project where students are provided with data and requirements from an industry partner of the university. The students are divided into teams and each team is expected to create the data warehouse from the data, and to provide access tools to meet the business requirements, according to the methodology of Kimball. A representative from the industry partner is on the evaluation team. Often the students find their first employment opportunity at the industry who supplies the data for the project.

III. SYSTEMS THINKING

Systems thinking developed as a reaction to the reductionist approach of management science in the period around World War II when management problems were identified and solved using mathematical models. A system is a set of interrelated components or subsystems that work together to achieve a goal [3]. It has emergent properties which are not identifiable in the subsystems and it has built-in control mechanisms to ensure effective achievement of the goal [4]. The environment of the system is the constraints in which it has to function [3]. Management problems are part of problematic situations, influenced by many complex social factors and therefor difficult to approach by mathematical models alone. Checkland argues that a soft systems thinker views a problem situation as a "mess" and uses systems to make sense of the situation [4]. This is in contrast to early hard systems thinkers who view the problem situation as a group of systems working together. For the soft systems thinker the system is a method of understanding the situation from a specific worldview.

Checkland developed the soft systems methodology (SSM) as set of guidelines to understand a problem situation from different world views and to guide purposeful action to improve the situation [4]. A concise explanation can be found in reference [6]. In this paper the SSM is used to plan the improvement of the DW module for IT students at the NWU in South Africa.

IV. THE SOFT SYSTEMS METHODOLOGY APPLIED TO DATA WAREHOUSING MODULE PLANNING AT NWU

Most people find it easier to internalise the aspects of the SSM when it is presented by means of an example. In this paper the "example" is the purpose of the paper: improvement

of the DW module for fourth year IT students at the NWU. In this paper the SSM is presented in sections, each demonstrated by its application in the stated problem environment.

A. Overall structure of the SSM

A simplified flow of the SSM is depicted in Fig. 2 [6]. A real world situation exist where there are problems which somebody wants to address. Models are developed representing purposeful activity systems of different worldviews in the situation. Each module (depicted by a square shape in Fig. 2) represents a different worldview.

The modelling process will be discussed in the next paragraph. The modules are not descriptions of the current problematic situation, but rather activity diagrams that represent the desired actions from various worldviews ("weltanshauung" [5]). The models become discussion aids when compared to the real world situation, and often are the source for discussion and understanding. A process of remodelling is followed to design a model that accommodates the ideas of the different world views. This should yield a model of purposeful activity that everybody can live with. When the purposeful action is taken, the situation is hopefully improved, but a further cycle of analysis is sparked.

There is also a parallel process present of cultural analysis consisting of social analysis – focusing on roles, norms, and values – and political analysis – focusing on power and the commodities thereof [6].

B. Analysis One: Building the Models

Model building usually starts with the identification of the transformation that is required in the problem situation. Often rich pictures are drawn to indicate the different stakeholders in the situation. Models are built to demonstrate the different worldviews present. In this paper the module is built from the perspective of the worldview of the lecturers, which can initially be stated as:

Some industry involvement in the module is good and students should do practical and theoretical work.

The transformation required in this environment is:

An unsatisfactory data warehousing module that needs to be transformed into an improved data warehousing module.

To understand this transformation better, some background information is given about the history of the module. When the module was started 10 years ago, it was mainly presented as a theoretical module where the industry focussed textbook (reference [2]) was taught in a face-to-face model of instruction.
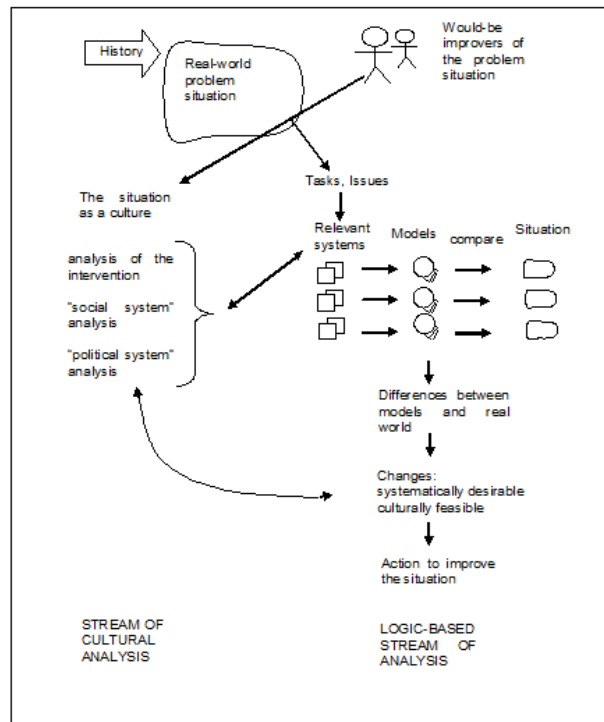

Fig. 2 The overall process of the SSM [6]

In was soon realised that the module needs a practical section and data was generated for the students to create a data warehouse. Soon the students who completed the module found themselves in the DW industry in South Africa. As more and more students completed the module, some of companies employing these alumni became more interested in the NWU and specifically this module in DW. They wanted to be more involved. The lecturer welcomed this as it was difficult to provide new data for practical work each year and one company was willing to provide desensitised data and requirements. In the past year another campus of the NWU also started to offer the module. This created the opportunity for the two lecturers involved to have meaningful discussion on the content and methods of instruction used. After discussion it was decided to that there are room for improvements, some of the identified difficulties include:

• Representatives from the company only visit the students twice a year. The company is a two hour drive away from the one campus and one hour away from the other campus.

• It takes the students a long time to understand the given data.

• Usually it ends up as a data-driven process instead of a requirements-driven process.

• The lecturers feel inadequate to assist the students.

• The lecturers feel that they have lost control over the module.

• The industry partner feels that they do not get enough benefit for the hours they invest.

• The students feel the workload is too high and that they do not get enough guidance.

Since all these problems are linked to perceptions of different role-players, it was decided to use the SSM to redesign the module. A rich picture is a hand drawn picture that depicts some difficulties in the current area of application. Fig. 3 depicts simplified rich picture for this problem situation.
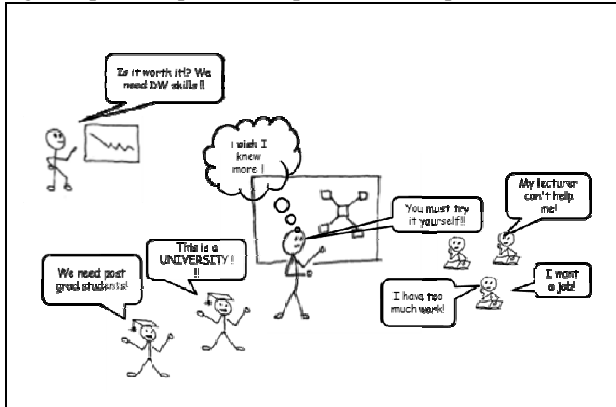

Fig. 3 A rich picture of the problem situation

A process of PQR analysis then followed to aid the formulation of the root definition, where P indicates what should be done, Q how it should be done and R the higher goal to be achieved. For this situation:

P: Redesign the data warehousing module

Q: Have discussions with stakeholders and review teaching strategies

R: Provide students that meets expectations of stakeholders

After PQR analysis, CATWOE analysis is done to better understand the situation and the worldview represented by the specific model. The letter C depicts the customers of the action; A depicts the actors that can achieve the transformation (T). Each model is developed to represent a specific worldview (W), which is explicitly defined. The owner (O) is the party that has the power to stop the transformation. Finally E is the environmental constraints in which the transformation should take place. For this situation:

C: Students, industry and faculty staff

A: Lecturers and industry

T: Redesign DW module to address problems

W: The DW should include industry involvement and students' self-directed learning skills should be improved.

O: Lecturers and faculty management and to some extend the industry.

E: Constraints of study hours per module; distance to current industry partner; available infrastructure; workload of lecturer.

All models include measures for monitoring and controlling the system in terms of performance criteria. Checkland advises criteria for at least efficacy (E_1), efficiency (E_2), and effectiveness (E_3), of the transformation [4]. For this situation:

E_1: It is realistic to involve all the parties identified as stakeholders. A module is designed with outcomes that satisfy the expectations of identified stakeholders.

E_2: The study hours available of students for this module are used effectively; the industry partner is satisfied with his return on investment; the time required by the lecturer is within university guideline.

E_3: Students are delivered that has suitable skills for industry but are well educated individuals with potential to continue with post- graduate study

Analysis one in terms of PQR, CATWOE and the three E's led to the development of a root definition of the purpose of the system from the perspective of a specified worldview. A good root definition includes many of the aspects of PQR and CATWOE. For this situation:

A process managed by the lecturers from the two campuses where the module is offered, to improve the data warehouse module in such a way that expectations of students, industry and faculty staff are met, while taking into account the limitations of available study hours of students and teaching hours of lecturers.

Next, an activity diagram is developed that demonstrates how the transformation will be done in terms of separate activities. Dependencies and flows of activities are indicated by activity numbers. Fig. 4 depicts a simple activity for this situation.

C. Analysis Two: Social System

As indicated in Fig. 2, a process of social analysis is done in parallel with the module building process of analysis one. During analysis two roles (both formal and informal) are identified in the problem environment. Norms, or expected behaviours associated with the roles are identified. These are linked to values. Values, in this instance, are the standards or criteria used to judge the behaviour of roles [4]. In this situation there are formal and informal roles present in each of the stakeholder groups. For example there are top-performer students which are students aiming for academic awards, there are also informal student roles such as moaners and enthusiasts. A full discussion on roles in this situation is outside the scope of this paper. Norms of the top-performer students include that they will not exaggerate the hours it took them to complete a task and they will spend all the hours associated with the credits on the module on studying for this module. Values for the top-performers include aspects such as honesty and self- directedness.
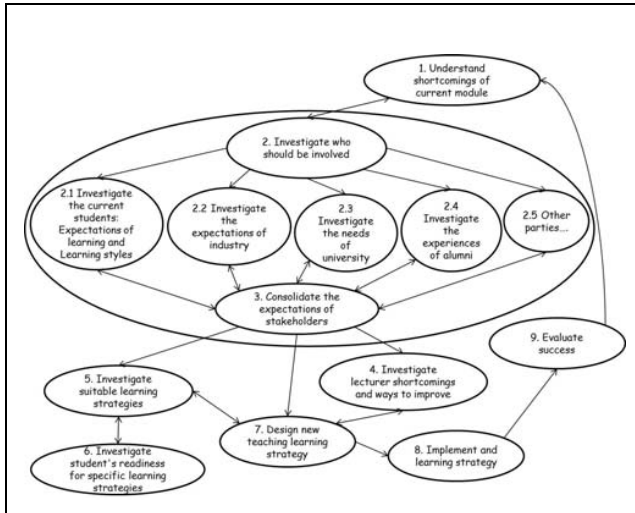
Fig. 4 A simplified activity diagram of the problem situation

The output of analysis two is a description of all the roles in the problem situation and the norms associated with them. This document is often only used by the facilitator in the situation. In this case it is the lecturers trying to redesign the module. They will do analysis two to better understand the stakeholders in the situation.

D. Analysis Three: Political System

Every situation has its own "politics" that ultimately decides what gets done and what not. The focus of analysis three is "to find out the disposition of power in the situation and the processes for containing it" [4]. Power can be present in different commodities. Commodities of power include things like personal charisma, membership of committees, access to senior decision makers, intellectual authority and reputation. In this situation experience academic staff members are holders of some power. The industry partner possesses another kind of power. Even the students have political power – if they tell their other lecturers about all the DW problems a negative image is created in the faculty. As in the case of analysis two, the lecturers facilitating the improvement of the module should keep up an explicit description of power issues in the situation. Such a document will assist in determining the best method to implement changes in the DW module content and presentation.

As indicated in Fig. 2, the results of analysis one, two and three are taken into account and compared with the actual problem situation to suggest improvement. Any proposed changes should always be culturally feasible which implies that parties in the situation should be willing to accept them. The implementation of the activities in the activity diagram of Fig. 4 is outside the scope of this paper.

V. CONCLUSION

This paper aims to demonstrate how using the SSM can help to improve the DW module at the NWU. Using SSM proved to be advantageous in terms of:

- As the module is offered at two campuses of the NWU the SSM process gave an opportunity and structure for the two lecturers to discuss issues intensely.
- The facilitators (lecturers) were forced to look beyond the current situation. The SSM forced them to have a fresh view on the situation.
- The SSM gave structure to the thought process; it provided traceable material that documented the thought process of the lecturers.
- When doing CATWOE analysis the facilitators had to think about their own perspective on the importance of specific role-players.
- The resulting activity diagram provides a roadmap for actions to be taken.
- The SSM is a logical process that can easily be followed.
- Analysis two and three made it possible for the two lecturers to discuss issues they would not normally discuss. The power of certain faculty staff could be discussed in term of power commodity rather than personal preference.

Some problems with the usage of SSM can be identified. It is accepted that these problems might be due to the lack of experience in using the methodology. Problems experienced include:

- It felt that when building the models and the activity diagrams it was easy to get carried away into a state of thinking everything is possible!
- The scope of the problem kept on increasing – there were always more people to consider – what about future students? What about alumni? What about future employees? What about the lecturers of the other modules? A strategy was needed to limit the number of parties to be consulted.

It was clear that this DW is part of a bigger system and it is difficult to define the boundary of the DW module in terms of the bigger system.

The process depicted in Fig. 4 now acts as the roadmap for further research in this situation. The resulting module changes will be implemented and evaluated – again using the SSM.

REFERENCES

[1] W.H. Inmon, Building the data warehouse. 2nd ed. New York, NY: Wiley. 1996
[2] R. Kimball, M. Ross, W. Thornthwaite, J. Mundy & B. Becker. The data warehouse lifecycle toolkit. Indiana: Wiley Publishing. 2008. P. 636.
[3] C.W Churchman. The systems approach. New York, N.Y.: Delta. 1968.
[4] P. Checkland & J. Poulter. Learning for action. Chichester: Wiley. 2006.
[5] P. Checkland. Systems thinking, systems practice. Chichester: Wiley. 1981.
[6] P. Checkland & J. Scholes. Soft systems methodology in action: Includes a 30-year retrospective. Chichester: Wiley. 1999.