Data Quality in Educational Systems for Decision Makers

Geoff Sandy and Bill Davey
Victoria University and RMIT
Geoff.Sandy@vu.edu.au, Bill.Davey@rmit.edu.au

Abstract: Significant research has shown that effectiveness of an educational administrative system can be measured by acceptance testing. These research efforts have suggested that an additional measure should be of system quality. ITEM systems are used by Universities and funding bodies to make important decisions. Case study research reported here shows that there is a likelihood of poor decisions being made because of the continued existence of data quality problems. This means that a decision maker should be able to see what reliability is relevant to each piece of data. Data quality literature shows that data quality depends on some important and known factors and case studies show that only some of these factors are included in current ITEM systems.

Key words: University, data quality, administrative systems.

1   INTRODUCTION

Recently, a former Director of the CIA revealed that a real-life version of this fictional scenario was actually played out when a test tape was inadvertently installed and the screen at a similar center warned of a nuclear attack. As computers play a more and more important role in the real-world – a world in which the computer-generated outputs often present a picture of the real-world for critical activities – it is increasingly vital that that picture be correct! (Orr, 1996)

Although educational management seldom deals with nuclear warfare, decisions in education do depend on information and do have consequences. Each of the universities studied in this research has a revenue from the

national and state governments of the order of $500 million and $200 million respectively (RMIT annual report 2002, VU annual report 2003). These funds are received as a result of statistics collected from and stored in the university information systems. One of the very few studies that examined the integrity of this data was conducted in 1992 by a team of researchers that examined statistics student by student. This study found: "This information was obtained for the five year study period but, when checked against faculty records, was found to be only about 50% reliable." (Burns, Davey, Hill, Leveson 1992). Obviously, in 1992 the University information systems were not reliable and very large amounts of money were being distributed on the basis of these poor figures. Each of the universities in this study has recently upgraded its information systems. This exploratory but timely study of the new systems reveals that whilst fundamental improvements have been made we are still not confident about their data integrity.

2   THEORETICAL FRAMEWORK

The general literature of data quality contains some coherent principles regarding the means of keeping data pure. According to the U.S. Defense Information System Agency (Cykana & Stern, 1996), the root causes of poor data quality can be attributed to four primary areas:
1. Process problems.
2. System problems.
3. Policy and procedure problems.
4. Data design problems.

There are a number of general Data Quality Rules that one can deduce from an FCS view of information systems:
DQ1. Data that is not used cannot be correct for very long;
DQ2. Data Quality in an information system is a function of its use not its collection;
DQ3. Data Quality will, ultimately, be no better than its most stringent use;
DQ4. Data Quality problems tend to become worse with the age of the system;
DQ5. The less likely some data attribute (element) is to change, the more traumatic it will be when it finally does change;
DQ6. Laws of data quality apply equally to data and meta-data (the data about the data). (Orr 1998)

Similarly Becker (1998) categorizes seven common data quality problems seen by end-users: (1) data corruption due to incorrect conversion, (2) historical and current data having different meanings, (3) the same data

having more than one data definition, (4) missing data, (5) hidden data, (6) missing granularity, and (9) violation of integrity rules. Clearly both of these sources indicate the need to correctly structure data, and then to use the data regularly in support of daily operations. The need for structure starts with the intended uses of the data. To see what that means in an educational setting the case studies started with analysis of user complaints about the two old systems. From this analysis the data quality rules could be seen in terms of three main areas common to both the case studies.

First is the accessibility of the information required by the decision makers. Second, is the integrity of the information used by the decision makers and finally, is whether the information supports the business rules of the organizational unit. Each of these is discussed in section 4 with examples drawn from each of the case studies.

3   METHODOLOGY

To test this framework two universities were chosen for an in-depth study. Both had acknowledged problems with their administrative systems but had taken different paths to solving these problems. The researchers are both senior academics at the respective universities with significant administrative responsibilities. A range of administrative staff and academics provided input into the findings through interviews and other verbal input. A study was also made of documentation of the administrative systems. Specifically, data was collected about the accessibility and integrity of the information from the old systems. In addition data was sought as to how well the information supported the business rules of the organizational units of each university. This includes both the academic administration and the academic teaching units.

4   THE CASE STUDIES

Two universities in Victoria, Australia were chosen for the study due to the ability to obtain very detailed data and the recent revamp of their information systems. The universities have taken different directions in addressing the problems of data quality. VU has produced a data warehouse solution we will call the MIS. RMIT has used a People Soft product to support a business process reengineering model called the AMS.

4.1   Victoria University Data Warehouse (MIS)

In 2002 Victoria University (VU) introduced a Data Warehousing system and MIS (hereafter referred to as MIS). Funds were first voted for this project in 1999. The main data sets of the Warehouse are the Victorian Tertiary Admission Centre (VTAC) Popularity Polls, the VU Student Information System (VUSIS) and data VU provides to the Commonwealth Department of Education, Science and Training (DEST). Wherever possible official point of time data captures are used in preference to VUSIS data.
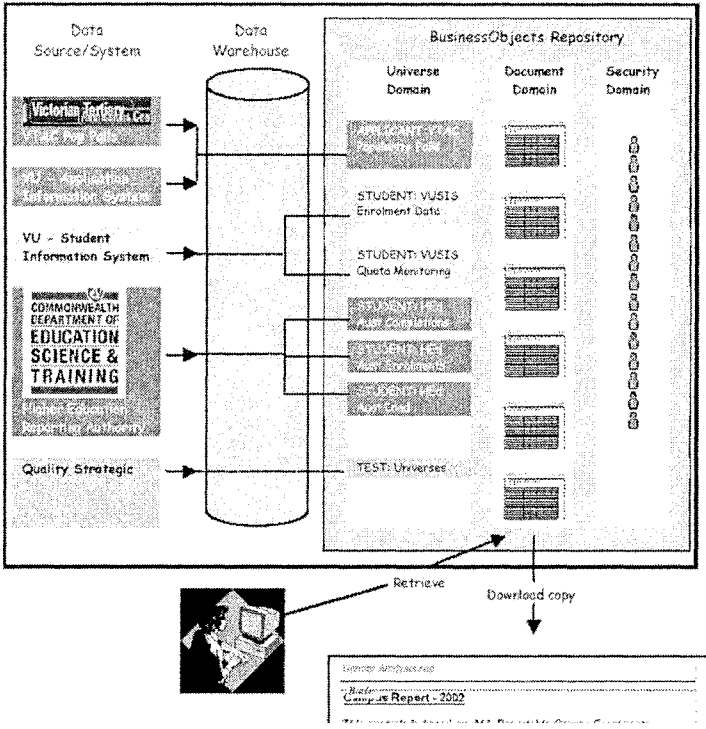

Figure 1: MIS at Victoria University

BusinessObjects is a software program that allows authorised users to retrieve and analyse corporate data. The BusinessObjects Repository is a database that controls all BusinesObject activities. There are three domains:
1. Universe Domain.
2. Document Domain.
3. Security Domain.

Users are concerned with corporate documents housed in the Document Domain. They can access and print standard reports and are able to create and access customised reports. The Course Analysis Reports for the Higher

Education Sector of VU is an example of a class of standard reports. They have been widely distributed and a response to their usefulness has been sought. The results have yet to be published.

The MIS provides relatively easily access and all relevant information is warehoused. The MIS provides information for each course for 3 years (2001-2003). Figure 1 shows that the MIS is a data warehouse, taking data from a number of sources and providing it to users in the form of Business Objects to the user.

4.2 RMIT University Academic Management System (AMS)

In 1998 RMIT Council adopted the Information Technology Alignment Project (ITAP) recommendations, including the acquisition of a functionally rich and inherently flexible integrated academic management system to manage the student learning experience and to facilitate the academic teaching process. In 1999, RMIT conducted a tender for a software replacement to the existing HP3000 legacy systems. The selected software was PeopleSoft Student Administration, with AST Consulting selected as the Implementation Partner.

Business Process Re-engineering (BPR) is a recognised management tool used by organisations to improve their performance across a range of indicators. The purpose of reviewing an organisation's business processes is to gain a comprehensive understanding of the way the organisation currently operates, to identify bottlenecks and difficulties in the current system and to redesign those systems to produce an improved outcome for the organisation and its stakeholders. The following abstract from the AMS process description website, http://www.ams.rmit.edu.au/business/bp_sr.htm, show the type and extent of processes that have been reengineered.

0	Current	Enrolment	Enrol Student -Single Subject-Enquiry
1	Current	Enrolment	Enrol Student -Single Subject-Web Based
2	Current	Enrolment	Enrol Student -Single Subject-Paper Based
3	Current	Enrolment	Enrol Student -Higher Ed Onshore Local (UGRD, PGRD, RSCH)
4	Current	Enrolment	Enrol Student -Higher Ed Offshore International (UGRD, PGRD, RSCH)
6	Current	Enrolment	Enrol Student -Preparatory (Incl VCE)
7	Current	Enrolment Variations	Context Enrolment - Registration/Variation/Exemption
8	Current	Enrolment	Enrol Student -Offshore (PREP, TAFE, UGRD, PGRD, RSCH)
9	Current	Enrolment	Enrol & Re-enrol Student OLA
10	Current	Re-enrolment	Re-enrolment preparation

There are approximately 500 of these high level processes in the list, each of which has a detailed process map, a small section of which has this level of detail.
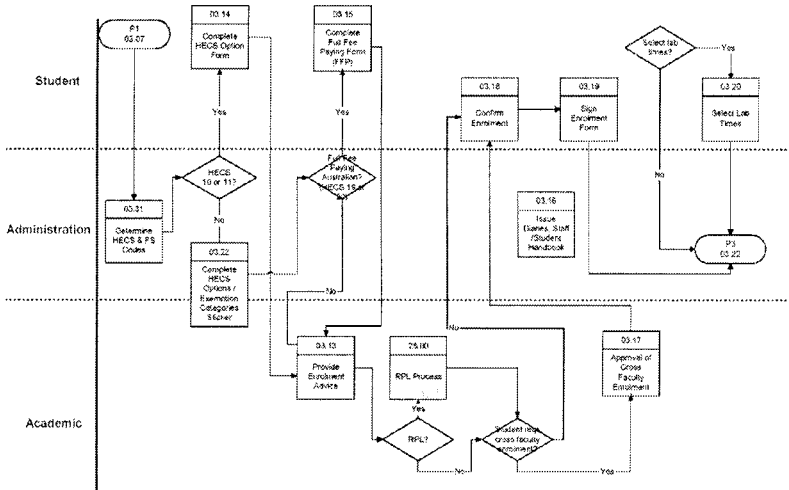

Figure 2: Typical process definition form the AMS

Accurate measurement is important for the University and specifically at the course level because the Australian Government like many others is intent on mandating many more quality control processes that must be reported upon by the University.

The AMS team made an important initial decision. That was to create robust process descriptions for every process prior to implementing that process in the new system. An unattributed adage "a computer allows us to turn a manual mess into an automated mess" is typical of the outcome of many systems that do not ensure robust systems before automation.

5   PROBLEMS WITH DATA QUALITY

The study focused on how the old systems performed in regards to three key aspects of data quality. Reference is also made about the new system. Appropriate information must be accessible to the decision makers. Information cannot be used by people who cannot find it. In both cases senior academics and administrators were found to not use systems because of their inaccessibility.

Similarly information will not be used if it suspected of lacking integrity. This can be due to the structural factors pointed out by Becker, or due to lack

of use resulting in out-of-date and incomplete data. In both cases integrity concerns were prominent in the decision to introduce a new system.

Finally, information must support the fundamental business rules of the organisation. Both case studies showed that the old systems did not fully support the important business rules. The new systems were also prompted by a desire to have some business rules re-engineered and implemented in new systems.

5.1 Accessibility of Information

A Victoria university the legacy system, called VUSIS, was necessarily complex and therefore difficult to navigate. It presented difficulties to anyone but the most experienced users. Also some important information was not stored on VUSIS.

A Head of School (HOS) requires information on how many fee-paying students are enrolled in each undergraduate and postgraduate course. In addition a break-up based on onshore/ offshore is required. The information is to establish the degree of dependence on full-fee income for each course. A threat, like a SARS outbreak, may put a course or School at risk if there is a very high dependence on full fee income. Such information is also required because of a strategic decision to attract more students onshore (more profitable) than operate offshore. In the past such information was difficult to obtain because of difficulties associated with the VUSIS and that relevant information was stored in multiple places.

At RMIT the legacy system was an HP product, distributed through an interface called Reflections. Most people knew the system as Reflections. This system was available to one person in each department in the University. Because the system was mounted on very old hardware it became so slow as to be unusable during peak periods. Most departments reported that they used paper-based back-up systems for 'the real work' although the system was robust and reliable in the features that were offered. Accessibility of information was a problem with the Reflections system since an emulation program needed to be installed on each client machine and the server was very old technology. This meant that only a few people in each faculty were able to be connected. In practice this meant printing of lists for use by staff.

5.2 Information Integrity

Recently, the Senior Management of Victoria University decided to continue with the practice of annual increase of the ENTER score required

of students to enrol in an undergraduate course. This is the score graduating secondary school students receive. An official document analysing the impact on each Faculty, Course and Campus was distributed. It sought to use 2003 ENTER scores to assess the impact of raising the ENTER from 65 to 70 in 2004. The Faculty of Business and Law's executive officer believed that the numbers in the document were inaccurate in regard to the Faculty. Using BusinessObjects the results revealed this to be the case Specifically, the numbers enrolling with an ENTER in 2003 was 51 not 103.

A manual check would confirm the accuracy of the MIS but this was not undertaken as it was considered too resource-intensive and laborious. However, the Faculty of Engineering, Science and Technology did undertake such a manual count. It revealed that the senior management data was very inaccurate and the MIS sound.

As with many institutions, RMIT determines program viability through a number of measures. These include survey data from graduates. In recent times a number of programs were marked for deletion due to unfavourable survey results. On enquiry it was found that the underlying survey results were from less than five graduates. This level of granularity is an issue with data quality. Any critical data values that are used for decision making should allow the manager to 'drill down' to determine how much reliability is supporting any given figure. This simple example was repeated in most of the decision making within the University. Eventually senior management decided that a new system was required to provide timely and accurate information.

5.3   Business Rule Support

One of the quality measures the success of a course is evaluated against is the student retention rate across the years of the course. DEST uses a measure referred to as the Apparent Retention Rate (ARR). It is an important business rule about what makes for a successful course.

The MIS project team at Victoria University has identified the need for the University to upgrade its information systems so it can accurately measure course completions. Currently, VU reports ARR data to DEST based on graduation information. However, it is recognised that the number of students who graduate may be less than the number of students who complete and that this may have an adverse affect on the retention rate. It is also recognised that the ARR is based purely on the student and not the student course combination. This means that a student who transfers to another higher education course will be counted as a continuing student.

At RMIT the issues of qualification to graduate, recognition of prior learning and prerequisite courses are all seen as academically essential. The

old system supported none of the relevant business rules and full-time staff were allocated in every department to perform these tasks. The errors due to this massive amount of manual processing had become of concern to senior administrators and were seen as a justification of the new system. Prior to introduction of the AMS a considerable amount of effort was expended in determining and integrating business rules from around the University. Some business rules were found to be so complex that the University decided to simplify all rules. Documentation and standardization of business rules became an important step in producing an effective administrative system.

6   BENEFITS AND CONTINUING PROBLEMS OF THE NEW SYSTEMS

The MIS at VU represents an important development in minimising data quality problems. It confirms many of the data quality rules. This can be illustrated with reference to two examples previously discussed. The first concerns accessibility of information to make decisions on overseas full-fee students. The second concerns decision making needs on identifying high and low demand courses.

The MIS is an easily accessible system for important course reporting information than has been the past experience of users. The VUSIS system is necessarily complex and difficult to navigate. It presents difficulties to anyone but the most experienced users and does not contain all information.

Despite the benefits conferred by the MIS at VU there remain continuing data quality problems. Again this may be illustrated with reference to two examples that have arisen recently at VU. First, concerns the inability to define and accurately identify student attrition in a course. The second example is the failure to access in a timely and meaningful way, basic staffing information on a University, Faculty and School basis.

At RMIT, the AMS was transported from the United States where it had been used in the tertiary education sector. It is mounted on modern hardware and provides reporting features that have the potential to be very useful. The system is restricted in use, but reports are available to most officers of the university who might need the information. To this extent the data is used daily and any errors in the database do show up quickly.

The AMS sought to improve on this lack of availability by making information available through a web interface. The interface exists but security and training concerns have meant that only a few people in each department have password access to the system, particularly for input. This fundamental balance between security and universal access remains

unsolved with the new system. The process of adapting a program from one system to another across international boundaries has lead to a very large problem. The University has found many problems with the interaction of adaptations. It has been decided to restore the original version of the programs and to change University processes to match those inherent in the original.

7   CONCLUSION

The literature suggests that the issues 'Accessibility of Information', 'Information Integrity' and 'Business Rule Support' underlie the important issues of data quality in university administrative systems. Two universities took different approaches to implementing a new administrative system. One designed a data warehouse to provide data objects to users, the other took a business process engineering approach, in each case to address real data quality concerns. In both cases an important design criterion was to make information more generally accessible and hence improve its quality. In both cases significant improvements were created in accessibility, and this was reflected in perceived data quality. The data quality model of using three area of concern was found to be a useful means for studying information systems in the tertiary education field.

8   REFERENCES

Becker, S. (1998), "A Practical Perspective on Data Quality Issues," *Journal of Database Management*, 9, 35-37.

Burns B, Davey B, Hill G, Leveson L. (1992). Articulation and Credit Transfer, Northern Industry Education and Training Link Melbourne.

Cykana, P., Paul, A., and Stern, M. (1996), "DOD Guidelines on Data Quality Management," *Proceedings of the 1996 Conference on Information Quality*, Cambridge, MA, 154-171

Orr, Ken (1998) "Data Quality and Systems Theory - A Paper by Ken Orr, Ken Orr Institute"; in Communications of the ACM, May 1998.

RMIT (2002), Report of the RMIT Council for the period 1 January to 31 December 2002 http://mams.rmit.edu.au/wmhpedjibaj9.pdf.

Tatnall, A. and Davey, B. (1995). "Executive Information Systems in School Management: a Research Perspective". *World Conference on Computers in Education VI. WCCE'95. Liberating the Learner*. Tinsley, J. D. and van Weert, T. J. London, IFIP / Chapman & Hall: 579-588.

Victoria University (2002). Annual Report, http://www.vu.edu.au/home/Media_and_Marketing/Online_Publications/2003_Annual_Report/indexdl.asp.

Visscher, A. J. and Bloemen, P. P. M. (1999). "Evaluation of the Use of Computer-Assisted Management Information Systems in Dutch Schools." *Journal of Research on Computing in Education* 32(1): 172.