CRC Construction Innovation
BUILDING OUR FUTURE

CRC-CI Project 2001-002-B

Life Cycle Modelling and Design Knowledge Development in 3D Virtual Environments

Industry Report

Program Number:	B
Program Title:	Sustainable Built Assets
Project Number:	2001-002-B
Project Title:	Life Cycle Modelling and Design Knowledge Development in Virtual Environment
Project Leader:	John Gero, USYD
Project Manager:	Rabee Reffat, USYD
Project Researchers:	Wei Peng, USYD
	Julie Jupp, USYD
	Pak San Liew, USYD
	Lan Ding, CSIRO
Project Contact:	Peter Scuderi
Date:	November, 2004

Distribution List

CRCCI
Authors

Disclaimer

The Client makes use of this Report or any information provided by CRC CI in relation to the Consultancy Services at its own risk. CRC CI will not be responsible for the results of any actions taken by the Client or third parties on the basis of the information in this Report or other information provided by CRC CI nor for any errors or omissions that may be contained in this Report. CRC CI expressly disclaims any liability or responsibility to any person in respect of any thing done or omitted to be done by any person in reliance on this Report or any information provided.

Please direct all enquiries to:

Chief Executive Officer
Cooperative Research Centre for Construction Innovation
9th Floor, L Block, QUT
2 George St
Brisbane Qld 4000
AUSTRALIA
T: 61 7 3864 1393
F: 61 7 3864 9151
E: enquiries@construction-innovation.info

Summary

Experience plays an important role in building management. “How often will this asset need repair?” or “How much time is this repair going to take?” are types of questions that project and facility managers face daily in planning activities. Failure or success in developing good schedules, budgets and other project management tasks depend on the project manager's ability to obtain reliable information to be able to answer these types of questions. Young practitioners tend to rely on information that is based on regional averages and provided by publishing companies. This is in contrast to experienced project managers who tend to rely heavily on personal experience. Another aspect of building management is that many practitioners are seeking to improve available scheduling algorithms, estimating spreadsheets and other project management tools. Such “micro-scale” levels of research are important in providing the required tools for the project manager's tasks. However, even with such tools, low quality input information will produce inaccurate schedules and budgets as output. Thus, it is also important to have a broad approach to research at a more “macro-scale.”

Recent trends show that the Architectural, Engineering, Construction (AEC) industry is experiencing explosive growth in its capabilities to generate and collect data. There is a great deal of valuable knowledge that can be obtained from the appropriate use of this data and therefore the need has arisen to analyse this increasing amount of available data. Data Mining can be applied as a powerful tool to extract relevant and useful information from this sea of data.

Knowledge Discovery in Databases (KDD) and Data Mining (DM) are tools that allow identification of valid, useful, and previously unknown patterns so large amounts of project data may be analysed. These technologies combine techniques from machine learning, artificial intelligence, pattern recognition, statistics, databases, and visualization to automatically extract concepts, interrelationships, and patterns of interest from large databases. The project involves the development of a prototype tool to support facility managers, building owners and designers.

This Industry focused report presents the AIMM™ prototype system and documents how and what data mining techniques can be applied, the results of their application and the benefits gained from the system. The AIMM™ system is capable of searching for useful patterns of knowledge and correlations within the existing building maintenance data to support decision making about future maintenance operations.

The application of the AIMM™ prototype system on building models and their maintenance data (supplied by industry partners) utilises various data mining algorithms and the maintenance data is analysed using interactive visual tools.

The application of the AIMM™ prototype system to help in improving maintenance management and building life cycle includes: (i) data preparation and cleaning, (ii) integrating meaningful domain attributes, (iii) performing extensive data mining experiments in which visual analysis (using stacked histograms), classification and clustering techniques, associative rule mining algorithm such as “Apriori” and (iv) filtering and refining data mining results, including the potential implications of these results for improving maintenance management. Maintenance data of a variety of asset types were selected for demonstration with the aim of discovering meaningful patterns to assist facility managers in strategic planning and provide a knowledge base to help shape future requirements and design briefing. Utilising the prototype system developed here, positive and interesting results regarding patterns and structures of data have been obtained.

Table Of Contents

Summary .... **iii**

Table Of Contents .... **v**

1. SCOPE **1**
- 1.1 Report Organisation: 2

2. DELIVERABLES **3**
- 2.1 Deliverable on Statement: How Technology will Improve Maintenance Management and Building Design 3
- 2.2 Deliverables on Demonstration 1: Data Mining and Knowledge Discovery Algorithms 4
- 2.3 Deliverables on Systems Framework 4
- 2.4 Deliverables on Demonstration 2: Prototype System 4
- 2.5 Deliverables on Outcomes: Improved Connection between Maintenance and Design Knowledge 5

3. BACKGROUND **7**
- 3.1 Information Standardisation 8
- 3.2 The role of 3D CAD models in design and construction 8
- 3.3 Building Maintenance 8
- 3.4 Life Cycle Costs and Life Cycle Modelling 9
- 3.5 Virtual Environments 9
- 3.6 Existing Commercial Data Management Systems 10
- 3.7 Existing Knowledge Discovery Systems 10
- 3.8 Available Data Mining Software 11

4. SURVEY OF AVAILABLE INDUSTRY DATA **13**
- 4.1 Queensland Department of Public Works 13
- 4.2 Central Sydney Area Health Service 14
- 4.3 Industry Data Conclusions 14

5. APPROACH TO LIFE CYCLE MODELLING **17**
- 5.1 System Composites 17
- 5.2 3D Object-Oriented CAD Database & Building Representation Module 19
- 5.3 Agent-Based Virtual Environments Module 20
- 5.4 Data Mining Module 20
- 5.5 Stages of the Knowledge Discovery Process 21
- 5.6 Demonstration and Results of Data Mining Techniques 21
- 5.7 System Components and Results Summary 23

6. AIMM PROTOTYPE SYSTEM **25**
- 6.1 Architecture of AIMM 25
- 6.2 AIMM Interface and User Scenario 27

7. DEMONSTRATION OF AIMM™ ON BUILDING NO. 10 AT RPAH **31**
- 7.1 Demonstration of using AIMM on Building no. 10 at RPAH 31
- 7.2 Discussion 32

8. RECOMMENDATIONS AND FUTURE WORK **35**
- 8.1 Data Mining Requirements 35
- 8.2 Requirements for 3D CAD Modelling 38
- 8.3 Potential Industry Partners Survey 38
- 8.4 Integrating Life Cycle Cost Analysis in AIMM 38
- 8.5 Recommendations for the CRC 40

9. CONCLUSION **43**

10. REFERENCES **45**

1. SCOPE

As the construction industry adapts to new computer technologies, in terms of hardware and software, computerized design, construction, and maintenance data are becoming increasingly available. The growth of many business, government, and scientific databases has begun to far outpace an individual's ability to interpret and digest the data. Such volumes of data clearly overwhelm the traditional methods of data analysis such as spreadsheets and ad-hoc queries. The traditional methods can create informative reports from data, but cannot analyse the contents of those reports. A significant need exists for a new generation of techniques and tools with the ability to automatically assist humans in analysing the mountains of data for useful knowledge.

The increasing use of databases to store information about facilities, their use, and their maintenance provides the background and platform for the use of data mining techniques for future projections. The current technology for facility maintenance uses databases to keep track of information and for notification of maintenance schedules. These databases are so far not well linked with an interactive 3D model of the building and are generally presented in tabular form.

Data Mining (DM) and Knowledge Discovery in Databases (KDD) are tools that allow identification of valid, useful, and previously unknown patterns. These technologies combine techniques from machine learning, artificial intelligence, pattern recognition, statistics, databases, and visualization to automatically extract concepts, interrelationships, and patterns of interest from large databases. The DM and KDD techniques are capable of finding patterns in data that can assist in planning. Patterns and correlations identified from data mining existing records of maintenance and other facilities management activities provide feed back and can improve future maintenance operation decision making, inform strategic planning as well as the design of new facilities.

This research is motivated by several observations of the current situation in the building industry. Since the cost of maintaining a facility over its life span is more than the capital cost, a marginal increase in capital cost can be shown to produce an amplified reduction in facility maintenance with a concomitant reduction in overall cost of ownership. The prototype system developed in this research provides a tool to assist in such decision making during the life cycle of a building.

The primary aim of this project is to develop and test a prototype system and report its capabilities as a tool suitable for discovering meaningful patterns, correlations and to report useful information via filtering techniques. The filtering system is based on heuristics derived domain specific knowledge and provides direct assistance to users. The interface system has been developed to provide a platform to facilitate set-up, data mining, filtering and reporting capabilities. The system offers a user-friendly interface to support: data configuration, control of information flows, communication, and effortless interpretation of data mining results.

This report presents the AIMM™ prototype system and summarises how and what data mining techniques can be applied to building maintenance data, its results and the benefits of applying such techniques. The report includes discussion and recommendations of the overall performance and capacity of maintenance data requirements, CAD requirements and IFC conversion, the EDM engine, data mining algorithms and virtual environments.

The scope of this project is outlined as follows:

- Background research on issues of information standardisation, building maintenance, life cycle cost modelling as well as potential future trends of information technology in AEC;

- Analysis of data sources from industrial partners and external sources;
- Demonstration of data mining, its algorithms, software and applications on industrial partner data;
- Documentation and implementation of systems architecture, building life cycle model and integration of design knowledge in a virtual environment;
- Testing AIMM™ prototype system. Experimentation tests three building assets including: air conditioning unit, thermostatic mixing valve, and battery charger components;
- Development of filters and reporting interface system for prototype system using Java
- Demonstration of completed prototype system, its data mining algorithms, filters and reporting interface.

The RPAH (Royal Prince Alfred Hospital) Building No.10 was chosen to demonstrate data mining techniques and test the prototype system since the scope of maintenance data specifies a more complete mapping of building assets and CAD data and also maintains a level of complexity practicable for demonstrating the data mining module's capacity to automate knowledge development.

Technologies incorporated within the overall framework of the prototype system include:

- An object-based CAD system: ArchiCAD7.0;
- Industry Foundation Class mapping for component (element) information transfer: IFC2.0 and IFC2.x.;
- EDM database management;
- A 3D virtual environment, Active Worlds;
- WEKA;
- Agent Technology
- Java and Visual Basic C++ programming languages.

The objective of integrating these technologies is to demonstrate the detection of patterns and discovery of knowledge applicable to building maintenance, planning and design strategies from an object-oriented CAD building model in a 3D virtual environment. Such a tool provides an online interactive tool that automates feedback on any combination of assets and components.

This project is carried out at Key Centre of Design Computing and Cognition (KCDCC) of the University of Sydney.

1.1 Report Organisation:

The topics covered in the following sections of this report are:

- Section 2 covers the deliverables of the project.
- Section 3 covers background research.
- Section 4 provides an analysis of two data sources from the industrial partner and an external source.
- Section 5 introduces data mining, its algorithms, software applications, and a data mining approach to life cycle modelling of buildings.
- Section 6 introduces a detailed systems framework of life cycle modelling of buildings and design knowledge in virtual environments is presented.
- Section 7 introduces the filtering and interface of AIMM™ prototype system.
- Section 8 discusses the projected estimates, recommendations and future project extensions.
- Section 9 presents the conclusions.
- Section 10 provides a list of References

2. DELIVERABLES

The deliverables of this project are as follows:

1. **Statement**: Approach and how it will improve maintenance planning methodology and knowledge to be used in the maintenance and management of existing facilities.
2. **Demonstration 1**: Data mining and knowledge discovery. The demonstration will use industry partner data (where possible) to show the implications of the availability of such a system.
3. **Development:** Systems framework, specify data mining techniques, agent architecture, links and mappings between CAD, maintenance databases and 3D virtual environments.
4. **Demonstration 2**: Prototype modelling tool that can be attached to asset management systems used by industry and government entities responsible for the management of building assets in a 3D virtual environment.
5. **Outcomes**: Improve connection between maintenance and design knowledge. Higher levels of maintenance knowledge should produce improved building designs and improve collaboration between industry partners.

As a consequence of the deliverables listed above the following milestones were outlined:

- ***Research paper***: present approach to maintenance data mining that includes design data modelling submitted to appropriate conference.
- ***Demonstrate***: a working data mining and knowledge discovery algorithms using industry data to QDPW.
- ***Develop***: object-oriented representation to provide 3D interactive environment using data provided by Woods Bagot.
- ***Implement***: object-oriented representation in virtual environments to provide 3D interaction and a link between knowledge development (as a result of data mining), with the building model.
- ***Establish***: basis of agent technology and develop and implement software agent architecture for data mining.
- ***Implement***: populate maintenance database with selected industry data.
- ***Research paper***: present agent-based approach to maintenance data mining submitted to journal.
- ***Research paper***: present architecture of entire prototype system submitted to appropriate conference or journal.
- ***Workshop***: demonstrating prototype system with data mining of maintenance data.
- *Flyer*: derived from reports for CRC promotional purposes.
- ***Develop***: proposal for project submission on extension to time-based modelling.
- ***Final research report***: that documents the research basis of the development and its implementation.
- ***Industry report***: focused report for dissemination.

2.1 Deliverable on Statement: How Technology will Improve Maintenance Management and Building Design

A report, CRC2001_002_B_1, documenting potential future trends of information technology in AEC (Architecture, Engineering and Construction) has been completed. This report supported the need for a tool to assist facility managers in decision making and provide feed back for designers. This was established on the basis of relevant issues such as the need for: information standardisation, building maintenance, life cycle cost modelling, existing life cycle models of building data and their pros and cons of these models are discussed. The report also summarised the requirements of the design information and outlined the basic approach to life cycle modelling.

In addition, a refereed research paper has been published titled "Using data mining on building maintenance during building life cycle" at the 38th Annual Conference of Architectural Science Association, ANZAScA 2004.

2.2 Deliverables on Demonstration 1: Data Mining and Knowledge Discovery Algorithms

A report, CRC2001_002_B_2 presenting the application of data mining algorithms and knowledge discovery on industry partner building maintenance data has been completed. The results reported in this document were demonstrated to industry partners in order to illustrate the technique's success in searching for patterns and correlations within the existing maintenance data and provide a method for supporting decision making on future maintenance operations.

In this demonstration various data mining algorithms provided by WEKA were applied and the maintenance data was analysed using interactive visual tools such as stacked histograms. The maintenance data of a variety of asset types were selected for this experiment with the aim of discovering meaningful patterns and providing a knowledge base to help shape future requirements gathering. A summary of these results are provided in Section 5.7 and a more detailed analysis is contained within the Final Technical Report and its Appendix E.

In addition, a refereed research paper has been published titled "Improving the management of building life cycle: A data mining approach" at the CRC Construction Innovation 2004 Conference on *Clients Driving Innovation*.

2.3 Deliverables on Systems Framework

The overall framework of the prototype system has been developed and described across two reports; refer to CRC2001_002_C_3, and CRC2001_002_C_4. The final architecture of the complete system was documented in CRC2001_002_B_5.

The systems architecture was partially implemented and tested to demonstrate the performance of a variety of data mining algorithms, on a range of asset type. The completed system, results and outcomes are presented in Sections 6 and 7 of this report.

2.4 Deliverables on Demonstration 2: Prototype System

The preliminary AIMM prototype system was demonstrated during the March Quarterly Review meeting in Sydney and the results were documented in report: CRC2001_002_C_4. In addition, the July Industry Partner Workshops demonstrated the completed implementation of the AIMM system demonstrating: (i) object-oriented representation in a 3D virtual environment, (ii) software agents (iii) population of the maintenance database with selected industry data, (iv) filtering heuristics, and (v) user-friendly reporting interface. An industry focused flyer has also been designed and published, refer to the Final Technical Report: Appendix A.The July Industry Partner Workshops (refer to Final Technical Report: Appendix B) were held in both Sydney and Brisbane.

The completed AIMM prototype system, results and outcomes are documented in the Final Technical Report and summarised here in Sections 7. An evaluation of the performances and capabilities of the AIMM prototype system is provided in Section 8.

2.5 Deliverables on Outcomes: Improved Connection between Maintenance and Design Knowledge

This final Industry Report provides results supporting the capabilities, flexibilities and advantages of automated knowledge development as a result of data mining building models in a virtual environment. Evaluation and testing highlights the following:

- Production of timely data on the effects of different maintenance regimes and provision of proactive information for improving the design, maintenance and management of building facilities.
- Provision of testing methods to validate the usefulness and scope of current databases as a platform for guiding future decisions.
- Linking of a 3D model with maintenance data allows both the facility manager and the designer to gain access to information and knowledge that is currently inaccessible.
- Combination of a 3D model with maintenance and other asset data facilitates the ability of building designers and owners to visually model the impact of design, maintenance, refurbishment and extension decisions on the building's life cycle cost.
- Representation of the facility within the virtual environment provides a basis for linking data mining with emerging technologies (such as connecting to WAP phones and other PDAs both in the office and on site) to address a gap in the construction life cycle.

This Industry Report also provides estimates and recommendations for future works. The recommendations for the extension of this research are described in Section 8. In addition, estimates of the impact of up-scaling the system to handle more complete representations of all building assets, components and systems are provided. A discussion of the prospective areas of development for the CRC is also presented.

3. BACKGROUND

With the boom in the information technology at the end of last century, the increase in information availability has become a dilemma due to inefficiency in processing the information for decision-making. This problem becomes critical in the building industry when we consider the high degree of complexity of work flows involved and the accompanying uncertainty for decision making in the lifetime of a building. Efficiently dealing with information from different stages of a building's life cycle to improve profitability, productivity as well as strategic resource planning are important business forces driving life cycle modelling. The maintenance objective for the building is that the cost of any maintenance activity should be less than the expected marginal value of production enabled by the planned activity. To support this objective, it is essential to tackle the maintenance from multiple facets including interpretation of observed data, diagnosis of problems, planning repair and maintenance, and business evaluation of the value-added from different repair and maintenance options. Equally significant is defining the "value" of maintenance from both engineering and business perspectives.

Figure 1 illustrates a conceptual four-way framework that depicts this research in relation to a variety of existing commercial systems and related state-of-the-art systems' development. The four related areas include: (i) Facilities Management, (ii) Construction Industry, (iii) Architectural and Design Domain, and (iv) University and Research.

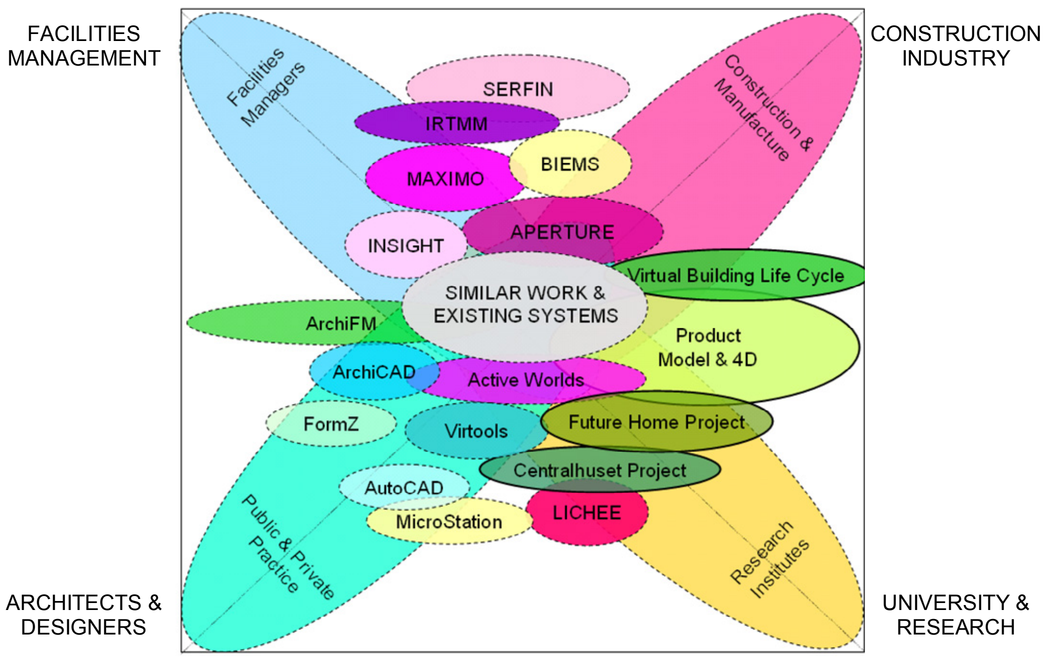

Figure 1. Conceptual framework of related commercial and prototype systems

Based on a general investigation into the four main areas, detailed background studies were carried out which focued primarily on aspects of: (i) information standardisation; (ii) role of 3D models in design and construction; (iii) building maintenance, and (iv) life cycle modelling. Other areas crucial to the development of the project included:

- Trends in Information Technology in AEC Industry;
- Virtual Environments and their significance in modelling;
- Data Management Systems;

- Systems Used for Knowledge Discovery and Data Mining.

The following provides a summary of the above areas and relevant case studies surveyed as part of this research project's development. For a more detailed description of related research, please refer to the Final Technical Report.

3.1 Information Standardisation

To leverage the project data generated in hybrid domains, as well as to improve the efficiency of information sharing, this CRC project has adopted the Industry Foundation Classes (IFC) interoperability standard to exchange project data. IFCs are defined by the AEC/FM industry in which they provide a foundation for shared project models and classes in a common agreed manner. IFC-based objects allow project participants to share a project model and improve efficiency in terms of objects reusability. Interoperability among AEC/FM software applications is achieved via the use of universal AEC/FM objects based on IFC specification. Based on its extensive applications in productivity, time and cost control during the design, construction and maintenance life cycle, IFC is becoming a de facto standard for the building industry. Many of the leading CAD software vendors such as AutoDesk and Graphisoft support importing and exporting IFC compliant data that is platform independent. The latest version IFC 2.x is supported by some leading CAD vendors as an add-on feature.

3.2 The role of 3D CAD models in design and construction

The primary purpose of a 3D CAD model needs to be established at an early stage in a project. 3D CAD modelling can be used in structural, lighting, acoustic, thermal, bio-climatic and spatial analysis. There is a common misconception that CAD systems are just drafting tools for use in the post-design stages of work rather than having a much richer role to play during designing and construction. However, 3D CAD models can help to resolve ambiguities, provide linkages to design data and present computerised visualisation. Discovering design conflicts and inconsistencies early is far less costly than repairing design and construction mistakes in buildings as cost increases exponentially at every stage from conceptual design to construction. 3D CAD modelling allows for such inconsistencies to be discovered before construction and therefore for better design and construction decisions to be made in the very early stages of the design and construction phase.

In addition, there are currently a variety of software packages available designed to facilitate interoperability between the products of various manufacturers and suppliers. They assist in converting and viewing CAD formats; translating between many different CAD/CAM (computer aided manufacturing) formats; and modifying and sharing objects. Moreover, there are Application Service Providers (ASPs) for the translation and adaptation of 3D solid models, enabling rapid sharing of 3D engineering data across design and manufacturing firms, and their clientele, regardless of their installed CAD systems (Reffat, 2002).

3.3 Building Maintenance

A simple statement of the maintenance objective for a building is that building systems are always available to support building function, and where applicable without ever limiting production. Where production is relevant, the building maintenance objective is that the cost of any maintenance activity should be less than the expected marginal value of production enabled by the planned activity. A managerial challenge is to allow the building and plant operating engineers as well as the owner to share information about current building component status, the business situation and for strategic planning in order for them to meet time-varying objectives.

Supporting this objective is difficult. It is difficult to assess the amount of risk posed by an observed non-critical problem to future production. There are multiple goals (e.g., high long-term availability, minimal short-term cost); goals change (e.g., between availability and cost concerns); goals conflict; indicator data are almost never completely reliable or adequate. The problem has multiple aspects, including interpretation of observed data, diagnosis of problems, repair and maintenance planning, and business evaluation of the value-added of different repair and maintenance options. Finally, significant judgment is needed to interpret both available engineering and business data, and clear business policy is needed to define the "value" of maintenance.

The issue of different maintenance options is complex. While *preventative maintenance* may avoid the unscheduled down time and costly repairs associated with reactive maintenance, it may be scheduled more often than is necessary. Thus costs are introduced that are unwarranted and therefore should be minimised without sacrificing plant performance. *Predictive maintenance* has enabled facilities technicians to identify and solve problems before they have a chance to damage equipment and now accounts for over 50% of *best* practice (Bowers). In *Proactive Maintenance*, the general time frame that a component will fail is determined *before* the failure occurs or is about to occur. Typically, a model of the system is used to anticipate failure. This approach minimises the risk of failure by eliminating root causes and currently accounts for approximately 10% of best practice (Bowers). The Data Mining component of the system is aimed at allowing proactive maintenance to occur based on the knowledge and patterns extracted from past maintenance records, thus allowing for generalisation from experience of similar assets as well as pinpointed information specifically relevant to a particular asset.

3.4 Life Cycle Costs and Life Cycle Modelling

Life cycle costs (LCC) are summations of cost estimates from inception to disposal for both equipment and projects as determined by an analytical study and estimate of total costs experienced during their life. LCC not only comprises initial acquisition cost, it also contains other cost like 'ownership cost" – operation costs, maintenance costs, logistics costs, etc, which is usually higher than the original acquisition cost. The major objective of LCC analysis is to choose the most cost effective approach from a series of alternatives so the least long term cost of ownership is achieved (Barringer, 1996). Based on an online report on Life Cycle Cost analysis (Kawauchi and Rausand, 1999), it is believed that a typical range of the ownership costs is 60 percent to 80 percent of the total LCC.

Life cycle cost modelling (LCM) helps facility managers in evaluating alternative equipment and process selection based on total costs rather than the initial purchase price. The multidimensional information that LCM presents is merged from hybrid project domains such as management, engineering, as well as finance. LCM may be applied in a wide range of critical functions, including: (i) evaluation and comparison of alternative design; (ii) assessment of economic viability of projects and products; (iii) identification of cost drivers and cost effective improvements; (iv) evaluation and comparison of alternative strategies; and (v) long-term financial planning.

3.5 Virtual Environments

Virtual Environments (VE) are computer generated synthetic environments in which users are provided with multi-modal, highly natural forms of computer interaction. Virtual Environment research is concerned with creating artificial worlds in which users have the impression of being in that world and with the ability to navigate through the world and manipulate objects in the world.[1]

Five existing life cycle modelling prototypes based on buildings with an object-oriented database in virtual environment platforms were surveyed. These are projects of 'Future Home', 'Virtual Building Life Cycle' (VBLC), LICHEE[2] (Life Cycle House Energy Evaluation), 'Product Model and Fourth Dimension

[1] http://www.ctit.utwente.nl/programme/areas/vr.html

[2] http://www.cmit.csiro.au/innovation/2002-02/lichee.htm

(PM4D)'[3] and 'Centralhuset'. The various benefits derived from the surveyed projects include:

- Use of virtual environments as platforms providing an interactive interface to improve communications between all team members as well as a simulation tool in enhancing predictable strategic planning within the whole life cycle of selected buildings.
- Application of the IFC-compliant object oriented database in standardizing the data exchange and facilitating the manipulation, reusability of project information.
- Linkage of maintenance data to 3D CAD model provides the potential of future development of intelligent life cycle analysis and control capability.

However these existing prototypes do not provide the capability of data mining of the hybrid data gathered from different stage of a building's life cycle, and do not provide performance-gaining life cycle analysis. This CRC project aims to bridge this gap to accomplish an efficient way in achieving an intelligent knowledge-gain.

3.6 Existing Commercial Data Management Systems

Five existing data management systems used in the construction industry were surveyed. These systems include: 'BEIMS[4]', 'INSIGHT[5]', LICHEE[6]', MAXIMO'[7], 'APERTURE'[8], and 'ARCHIFM'[9].

The various benefits derived from the surveyed data management systems include:

- Centralised mechanisms to control, regulate, maintain and track an organisations assets;
- Scheduling of preventative maintenance schedules;
- Online computerised methods of recording job requests, delegating tasks and tracking costs;
- Costing reports, individual transactions and definition of user profiles, access rights, etc.

3.7 Existing Knowledge Discovery Systems

Two existing knowledge discovery systems used in the construction industry were surveyed. These systems include 'SERFIN[10]', and the 'Intelligent Real-Time Maintenance Management' or IRTMM. SERFIN provides facilities management knowledge handling that aims to:

- Identify and capture problems that arise in connection with technical maintenance of buildings
- Make the problem solution process arising during maintenance of buildings more effective
- Make experiences from technical maintenance easily available and accessible in time and space

The IRTMM system provides integrated subsystems for the following aspects of the maintenance and repair planning problem: (i) Situation Assessment, (ii) Planning and (iii) Value Analysis. The various benefits derived from the surveyed data management systems include:

- model-based diagnosis to identify details of possible problems
- heuristic classification to identify idiosyncratic problems

---

[3] http://www.stanford.edu/group/4D/projects/calvin/PM4D.shtml
[4] http://www.beims.com
[5] http://www.sbt.siemens.com/BAU/products/BA_management_insight.asp
[6] http://www.cmit.csiro.au/innovation/2002-02/lichee.htm
[7] http://www.stanford.edu/group/4D/projects/calvin/PM4D.shtml
[8] http://www.aperture.com/solutions/workplace_solutions.html
[9] http://www.graphisoft.com/products/archifm/
[10] http://it.civil.auc.dk/it/delphi/KBS/projects/serfin.html

- case-based reasoning to compare observed data with previous cases

3.8 Available Data Mining Software

There are currently hundreds of mining tool vendors. A large number of reviews of data mining software are also available at (Goebel and Gruenwald, 1999; Elder and Abbott, 1998). Six leading data mining tools have been reviewed and further details can be found in the Final Technical Report. As a result of this survey the focus was narrowed to include only the IBM Intelligent Miner and WEKA since they perform better in scalability and functionality than other packages and are also leaders in commercial and non-commercial data mining software packages. WEKA is freely available, and Intelligent Miner has also been obtained without cost through the IBM Scholars Program.

The IBM DB2 Intelligent Miner Version 8.1 is a set of the following products (IBM Publications)[11]: Intelligent Miner Scoring (IM Scoring), (ii) Intelligent Miner Modelling (IM Modelling), and (iii) Intelligent Miner Visualizing (IM Visualizing). The advantages and disadvantages of IBM Intelligent Miner were identified and listed below as:

Advantages:

- It is a commercial software package, which may be better for the delivered prototype to industrial partners.
- Takes data in SQL format.
- Performs many different data mining methods, including decision trees and neural networks.
- Has text mining capability, which may be useful for free text fields of repair reports (where much of the useful data may be stored).
- Has sophisticated visualisation software.

Disadvantages:

- Does not appear to support meta-learning, i.e., combining the results of multiple data mining techniques on the same data.
- Does not appear to be very easily extensible.
- Source code is not provided.

WEKA[12] is a collection of machine learning algorithms for solving real-world data mining problems. The algorithms can either be applied directly to a dataset or called from your own Java code, since the WEKA package has built in Java which can easily be embedded into a pilot decision support system – one of the expected deliverables of this CRC project. WEKA not only contains tools for data pre-processing, classification, regression, clustering, association rules, and visualization, it is also suitable for developing new machine learning schemes. The advantages and disadvantages of WEKA were identified and listed below as:

Advantages:

- It is an open source software package, making it possible to modify or extend its capabilities.
- It is freeware that can be used in any system as long as its creators are credited.
- Supports meta-learning techniques such as bagging and boosting.
- Has a GUI version that appears to be adequate for our purposes.

[11] http://www.elink.ibmlink.ibm.com/public/applications/publications/cgibin/pbi.cgi
[12] http://www.cs.waikato.ac.nz/~ml/weka/

Disadvantages:

- Does not appear to include software for neural networks, but other ANN packages can probably be added without too much difficulty.
- Input data must be in their own specialised format (ARFF), but this only requires storing the data in comma-separated values format and adding attribute information and a few tags.

4. SURVEY OF AVAILABLE INDUSTRY DATA

A useful and productive application of data mining for improving the maintenance management of buildings is primarily dependent of having an appropriate, rich and diverse maintenance data set. Acquiring such a building maintenance data set has been an extremely difficult task in this project. This Section introduces a thorough analysis of maintenance data sets obtained from Queensland Department of Public Works and Central Sydney Area Health Service. The project's requirements on the quantity and quality of industry maintenance data are:

- Data must be available down to the element level;
- Data must be accompanied by corresponding 2 or 3D CAD data;
- All assets referred to in maintenance data must also be included in CAD data so that a one-to-one correspondence exists for an item as it appears in both types of data;
- Data must cover a period of at least 2-3 years (the longer the better).

This Section concludes by discussing the minimum requirements listed above in relation to the maintenance and CAD data provided by industry and their appropriateness as maintenance data sets to be used for data mining.

4.1 Queensland Department of Public Works

CAD and maintenance data was received for a building constructed and maintained under the supervision of QBuild, a commercial business unit of the Queensland Department of Public Works (QDPW). The data came from the Neville Bonner Building, a state-of-the-art building at 75 William Street, Brisbane. The Neville Bonner Building (NBB) is under the purview of the Portfolio and Housing Unit of QBuild, who, because they deal with complex high-rise buildings and have strict requirements for maintenance data collection and were therefore considered likely to collect data to the component or element level. The data collected for all Portfolio and Housing Unit buildings were deemed as suitable for this project's purposes. The CAD data for the Neville Bonner Building can be grouped into the following seven classes: (i) Air Conditioning, (ii) Ceilings; (iii) PDT (Powell Dods & Thorpe) Drawings; (iv) Raster images folder; (v) Weathered Howe (Civil and structure design works by Weathered Howe); (vi) Base Drawing; (vii) Drawings.

QBUILD databases use the following three layer tree coding:

- **WIC Number:** a 5 digit number that uniquely identifies a building site. For example, a WIC of 22380 identifies Glenmore State High School.
- **Building Code:** a 3 digit number that identifies an individual building, except for the code 000, which identifies elements pertaining to the overall site. For example, a building code number of 22380000 relates to Glenmore State High School as a whole, and a building code of 22380016 identifies building number 16 at Glenmore State High School.
- **Building Component:** a 4 character alphabetic code that identifies a type of system (fixed plant).
    - **Element Number:** a two digit number that uniquely identifies a specific asset. For example, 22380016AIRC01 identifies air conditioner number 1 in building number 16 at Glenmore State High School.

The maintenance database, which consists of work order records, contains a large number of fields, many of which relate to accounting and are not used. In the data received 2048 entries were recorded. Additional details of this data can be found in the Final Technical Report.

4.2 Central Sydney Area Health Service

Data was also received for the Queen Elizabeth II Building of the Royal Prince Alfred Hospital (RPAH), hereafter referred to simply as Building 10. It is a five-storey building, three levels of which are occupied by the Institute of Rheumatology and Orthopaedics.

CSAHS divided CAD data of Level 3-8 of Queen Elizabeth II Building (Building 10) which fall in 7 domains – (i) Architecture and related raw drawings, (ii) BMS (Building Management and Control System) which is in charge of controlling mechanical system as swell as coordinating other building services, (iii) Electrical, (iv) Engineering, (v) Hydrant, (vi) Lifts, (vii) Mechanical Engineering. All CAD data was supplied in AutoCAD 2D.

CSAHS currently uses the BEIMS system (described above in Section 3.1) to store maintenance data. Data for the last two and a half years is available in SQL format and contains data that is highly detailed and structured. In particular, it contains asset numbers that match those contained in the current CAD data. The analysis that follows is based on the approximately 5000 work orders recorded for Building 10 in the period from 1/1/2001 to 9/12/2002. More recent data has also been available to the project.

Data for the period from 1/1/1995 to 31/12/2000 is stored in the Paradox database format. The records are less detailed and contain references to assets that no longer exist because the building was renovated; nonetheless, the enormous dataset provides a rich resource for the data mining system to draw from. There are over 2000 work orders recorded for Building 10 in the period from 1/1/1995 to 31/12/2000, as well as tens of thousands of records for other buildings on the RPAH campus. The detailed work order records kept since 2001 can be found in the Final Technical Report.

4.3 Industry Data Conclusions

In order for the data mining process to succeed, it must have data of sufficient quality so that knowledge can be extracted, as well as sufficient quantity so that there can be a reasonable belief that the extracted knowledge reflects true patterns in the data, rather than mere statistical anomalies. The data we have received from QBuild for the Neville Bonner Building does not adequately meet the minimum requirements that must be satisfied, in terms of both quantity and especially quality, in spite of it being the data set that was deemed to be the best for the purposes.

It is critical that the maintenance data be available at the element level. The QBuild databases use a three layer tree coding, consisting of a WIC Number, Building Code and Building Component with Element Numbers, and is considered ideal to use in data mining. However, of the 2048 data entries for the Neville Bonner Building, only 15 entries specified the asset component.

Although it might be possible to make use of the component code or BOMA code, which are specified separately in another field (see Section 4.1.3), since these codes are present in the majority of entries, in order to distinguish between different assets, they would have to be used in conjunction with other information, such as location. However this information is not available. Without the ability to uniquely identify an asset, it is impossible to create a history of maintenance work done on that asset.

In addition, there must be corresponding CAD data available for the building, and the assets referred to in the maintenance data must also be included in the CAD data in such a way that it is possible to create a one-to-one correspondence between the same items as it appears in both types of data. While substantial CAD data for the building does exist, it is presented in a manner and is labelled in such a way that the assets are not included in the drawings nor identified uniquely. It may be possible to obtain better CAD models for later buildings.

The maintenance data acquired from the Central Sydney Health Area Services (CSAHS) meets all of the specified requirements to allow data mining. A large number of the entries in the CSAHS maintenance database contain a reference to an asset number, which refers to a specific building element or component. A detailed description is also provided for each asset, including an indication of the asset category, along with detailed location information in the form of a building code, room code, level code, and department code. Additional information is also offered when available, such as installation date, purchase value, make, model, and serial number. There is also a large amount of additional information provided in detail and in a highly structured manner and for Preventative Maintenance work orders, a description of the planned task is provided from a list of pre-defined descriptions. The maintenance data for Building 10 spans a period of almost 7 years, from January 1995 to September 2002 and contains over 7000 entries, 5000 of which were entered since January 2001, reflecting the higher level of accountability and detail that has been the practice in recent years.

The CAD data for Building 10 is also well organised and clearly labelled. In particular, the assets are identified uniquely in the drawings using the same asset codes as are used in the maintenance database. This allows an asset to be located with respect to the building, providing for further features that can be used in data mining. This also allows an end-user to access an asset's maintenance history and relevant discovered knowledge by selecting that asset in the 3D virtual environment that will be created from the CAD data.

5. APPROACH TO LIFE CYCLE MODELLING

The Architectural, Engineering, Construction (AEC) industry is seeing explosive growth in its capabilities to both generate and collect data. Advances in scientific data collection, the introduction of bar codes for almost all commercial products, and computerisation have generated a flood of data. Advances in data storage technology, such as faster, higher capacity, and cheaper storage devices (e.g. magnetic disks, CD-ROMS), better database management systems, and data warehousing technology, have allowed the transformation of this enormous amount of data into computerised database systems. As the AEC industry is adapting to new computer technologies in terms of hardware and software, computerised building data is becoming increasingly available. Yet, in many cases, this data may not be used, or even properly stored. Several reasons exist (Soibelman 2002):

1. Project managers do not have sufficient time to analyse the computerised data,
2. Complexity of the data analysis process is beyond the capabilities of the relatively simple building maintenance systems commonly used,
3. No well defined or automated mechanisms to analyse data and interpret results so that site managers can use it.

However, there is a great deal of valuable knowledge that can be obtained from an appropriate use of this data; there is a need to analyse this increasing amount of available data and Data Mining can be applied as a powerful tool to extract relevant and useful information.

5.1 System Composites

The framework of this CRC project is established based on three modules: (i) a Database module, (ii) an agent-based virtual environment module, and (iii) a data-mining module. The general description of this framework is illustrated in Figure 2.

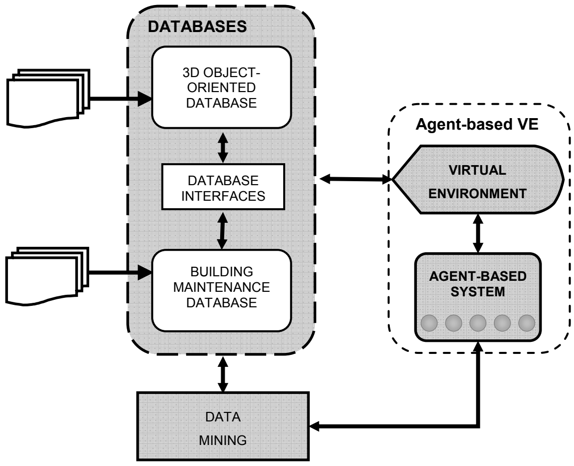

Figure 2. Proposed framework of life cycle modelling in virtual environments

Mining data enables us to understand how systems that were once thought to be completely chaotic actually have predictable patterns (Peitgen et al. 1992). Through KDD, patterns and causal relationships behind apparently random data in AEC projects can be found. By applying KDD to identify novel patterns, project managers will be able to build knowledge models that may be used for the recurrent activities of on-going construction projects, as well as for a future project activities, and avoid unanticipated consequences (Soibelman 2002). Data Mining also presents the potential to address the problem of transforming knowledge implicit in data into explicit knowledge for decision making.

In contrast to traditional methods of statistical data analysis, KDD is an automated process that discovers new trends and patterns without the need for human intervention. KDD takes input variables whose relevance may not be obvious to a designer but which becomes evident as a result of this process.

Thus, the approach is based on a comprehensive view of the building management problem. It views the process of building design, maintenance, and replacement as a process generating an enormous amount of information. While current practice addresses parts of this information generation and management, our approach attempts to account for the life cycle flow of this information.

Figure 3 outlines our model of the flow of information in building design and maintenance. The bold arrows depict the functionality provided by our proposed approach while the dashed arrows describe the scope of present approaches to building information management.

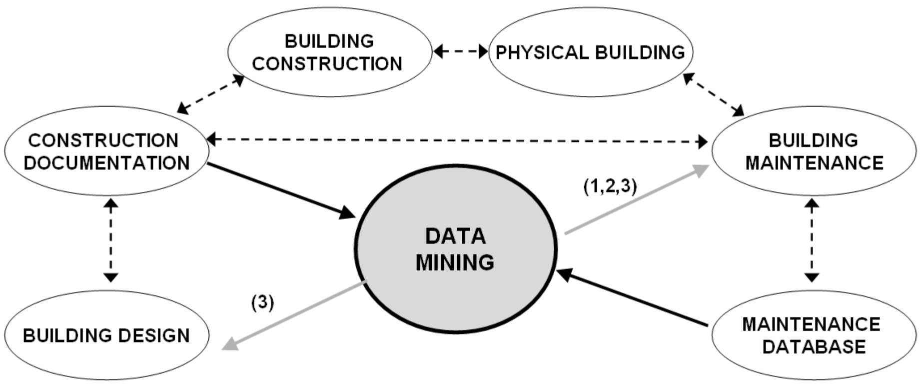

Figure 3. Integrating data mining within the life cycle of building information management.

Data mining techniques can be used effectively on data stored in a Building Maintenance System (BMS) by creating knowledge that can be used in future management and design decision making. Knowledge that implicitly resides in BMS databases includes information about:

1. Components that frequently need maintenance and therefore need careful inspection,
2. Historical consequences of maintenance decisions that may inform future decisions,
3. Components of buildings that significantly determine maintenance cost and therefore may inform future building designs, as well as refurbishment of the building in question

This information can be extracted using Data Mining techniques and used to improve all phases of the building life cycle, both for current and future buildings, as indicated by the numbers (1, 2, 3) and (3) marked on the arrows in Figure 3, which correspond to those listed above.

The Database module consists of 3D CAD files and an object-oriented database. The object-oriented database called Express Data Manager (EDM)[13] is used for managing CAD objects. The reason we adopt EDM comes from the consideration of construction industrial needs for interoperability and the data complexity in AEC domain

The Agent-based Virtual Environment module consists of agent technology embedded in Active Worlds. Apart from visualization and communicational capabilities, this CRC project is more concerned with life

[13] http://www.epmtech.jotne.com/products/index.html

cycle control within virtual environments. Unfortunately, most of artificial intelligence explorations on VE were carried out in other areas instead of in AEC. The integration of life cycle modelling with VE is made potentially viable by including intelligent agents and data mining. Through the agent-based virtual world each object represents an agent with attributes and behaviours.

The Data mining module incorporates WEKA because of its advanced scalability and functionality. WEKA[14] is a collection of machine learning algorithms for solving real-world data mining problems. The algorithms can either be applied directly to a dataset or called from your own Java code. WEKA not only contains tools for data pre-processing, classification, regression, clustering, association rules, and visualization, it is also suitable for developing new machine learning schemes.

5.2 3D Object-Oriented CAD Database & Building Representation Module

EXPRESS Data Manager 4.5 is a powerful object-oriented database management system that provides greater functionality for building application and data modelling. It also serves as a multi-user rule engine to define constraints to improve knowledge management. With the interactive working environment of *EDMvisualExpress*, a user can define and document object models that are capable of integrating domain specific rules for knowledge-based explorations.

Using IFC-compliant design software such as ArchiCAD or AutoCAD, different kinds of CAD data ranging from three-dimensional geometric data to space identity to building material information can be stored in the EDM database for further processing. However, like other methods, the process of transforming the data into a format suitable for knowledge discovery is not automated. The representation of building objects is not limited to the graphical information that illustrates only the structure of objects. There are various other types of non-graphical information that are not less important than the structural representation of these objects, including functional, behavioural and semantic properties.

The research team has coordinated with Woods Bagot (an Industry Partner in this project), to model the selected building (Building no. 10, Royal Prince Alfred Hospital, Sydney) into object-oriented 3D CAD. Woods Bagot modelled Building 10 using ArchiCAD, an object oriented CAD modelling system. One floor only has been modelled as a prototype for the demonstration and testing of the system. Captions of this 3D model are shown in Figures 4 and 5.

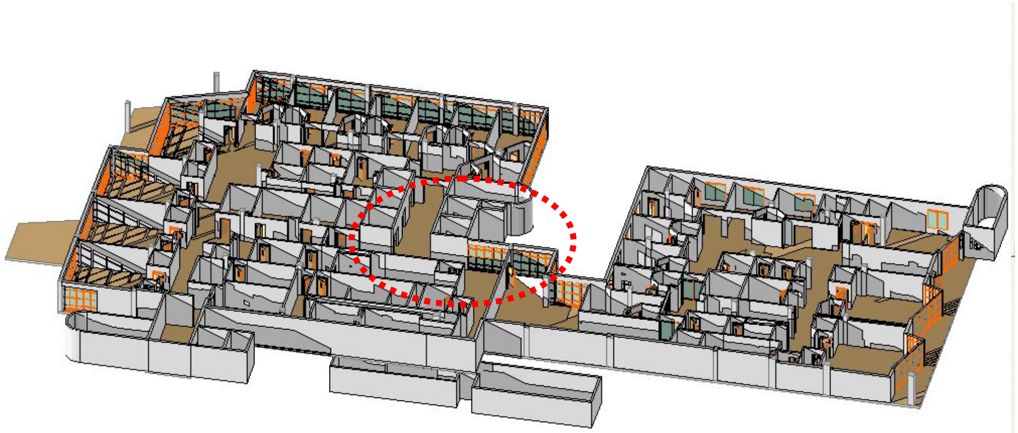

Figure 4. 3D view of typical floor of Building no. 10, Royal Prince Alfred Hospital, Sydney.

[14] http://www.cs.waikato.ac.nz/~ml/weka/

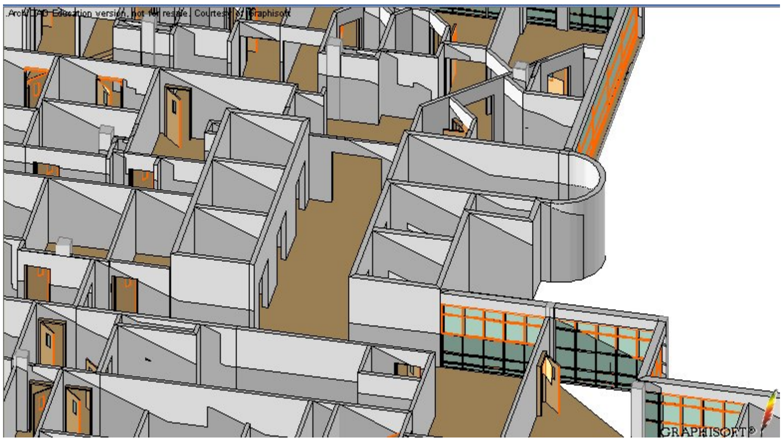

Figure 5. 3D view of vertical transportation area (Lifts and stairs) at Building 10, Royal Prince Alfred Hospital.

For this 3D CAD model to useful for interactivity in the virtual environments (Active Worlds), it needs to be converted into an RWX format in order to view it and interact with its objects. All objects in the 3D model were also exported to IFC (Industry Foundation class) format in order to be accessible to the Express Data Manager to handle both graphical and non-graphical attributes of building objects.

5.3 Agent-Based Virtual Environments Module

The virtual environment of an agent-based system serves as an interactive interface for linking users to the underlying 3D and data-mining modules. The agent model in the virtual environment is represented as objects that have agency and are capable of sensing, reasoning and affecting the environment. In an agent-based virtual world, a behaviour may be triggered by any change about the world, for example, triggers can occur either when a user 'clicks on' a 3D building element or when their avatar enters a room. In doing so, specific actions are performed according to agent's rules. An agent-based virtual environment is thus comprised of objects that have a 3D model and an agent model which encapsulates the following five operations: 'sensation', 'perception', 'conception', 'hypothesizer' and 'action' (Maher and Gero, 2002).

5.4 Data Mining Module

For this project, two types of building descriptions exist that can be exploited for data mining purposes; the complete set of features that can be derived from the object-oriented 3D CAD model of a building, and the aggregated features derivable from the building maintenance system. Based on an evaluation of the building maintenance and CAD data that has been made available to us from CSAHS (described in detail in Section 4.2), we make certain observations regarding the nature of the data that will be used for data mining:

- Data consists of both numerical and nominal values
- Information will be incomplete (fields may be empty)
- Information will be noisy (there may be errors or contradictory data)
- Some fields will consist of unstructured English text
- Some data will be in the form of a 3D object-oriented model of a building
- There will be large data sets available

Based on the above properties, a variety of approaches to data mining were evaluated including: Decision Trees, Clustering and Classification, Association Rules, Artificial Neural Networks, and Hidden Markov Models. Each of these approaches has certain advantages and disadvantages, and can best be used in a complementary fashion. Each of these algorithms have been described in report

CRC2001_001_1 and a summary of their review can be found in the Final Technical Report: Appendix E.

5.5 Stages of the Knowledge Discovery Process

The process of Knowledge Discovery from Databases occurs in several stages, of which Data Mining is one intermediate phase. The stages, of the knowledge discover process is as follows:

1. Select the data to be used for mining purposes
2. Pre-process the data to remove or correct errors, select feature subsets, create additional fields from existing data, etc.
3. Transform raw data into an appropriate format
4. Mine the data for emergent patterns
5. Evaluate the data to decide which patterns are relevant and useful

Data mining the CSAHS maintenance database is an explorative process since new knowledge is discovered and new hypotheses can be formed. The data mining process for extracting hidden knowledge from large databases can be depicted as shown in Figure 6.

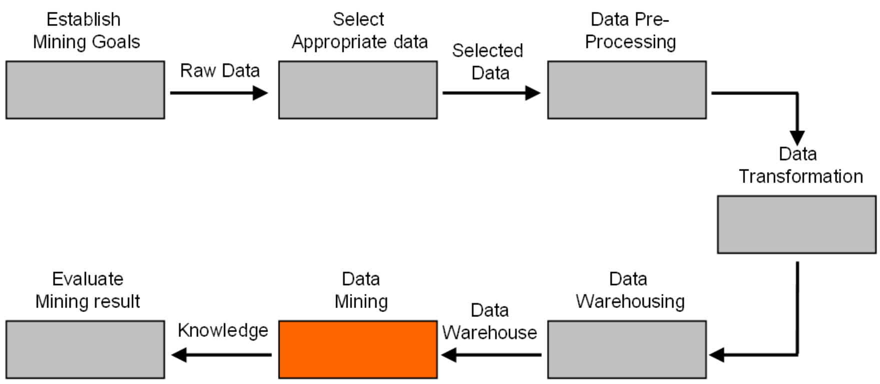

Figure 6. Stages of Data Mining Process (Hui and Jha 2000).

Data mining and visual analysis techniques are applied in the demonstration presented in this report as follows:

- Data analysis using stacked histograms;
- Classification through C4.5;
- Clustering using SimpleKmeans;
- Attribute evaluator for finding groups of correlated attributes using associate rule algorithm “Apriori”.

5.6 Demonstration and Results of Data Mining Techniques

This section discusses the preliminary results of the data mining techniques selected for systems implementation and tested their capacity to search for patterns and correlations of building maintenance data. Based on the industrial building maintenance data, the data mining algorithms provided by WEKA[15], are applied and the results analysed. The selected industrial building maintenance data covered a variety of domains. The aim of this demonstration was to discover meaningful patterns that

[15] http://www.cs.waikato.ac.nz/ml/weka/

can assist facility managers in strategic planning as well as provide a knowledge base to shape future requirements gathering and design briefing.

Through a process of cleaning, integrating meaningful domain attributes, conversion of data formats, a number of experiments were performed in which concentration is given to visual analysis using stacked histograms, classification and clustering techniques, an associative rule mining algorithm.

Visual data analysis and data mining techniques were applied on three selected data sets. The building maintenance data used in this demonstration are for three asset types including Air Handling Units, Thermostat Mixing Valves and Battery Chargers at Building 10, Royal Prince Alfred Hospital. An example of the raw maintenance data is shown in the Final Technical Report: Appendix D. Some interesting results regarding patterns and structures of data have been obtained and the following provides a summary of results from applying data mining techniques on Building 10 maintenance data.

- Results extracted from the available maintenance data of the asset type 1: Air Handling Units

1. Approximately all A/C malfunction belongs to high and medium priority.
2. A/C malfunction concentrates on the problems of: too_hot 32%, too_cold 28%, not working 7.5%.
3. The lowest levels of A/C Malfunction took place in August followed by June and April while other months share similar high A/C malfunction rate.
4. Approximately all the description of *too_cold* or *too_hot* were associated with high or medium priority
5. All 7th floor jobs were of high and medium priority
6. In all 7th floor jobs the cause of repairing was A/C malfunction
7. Maintenance jobs conducted at Floor 7 did not meet the expectations: with 23 out 25 completion_within_expectation = “N” (92%)
8. In August only department 26464 has corrective maintenance work. (only 1 case).
9. The work at 4, 5, 6, 7th floor constitutes most of the reports of A/C malfunctions, with 86% of A/C malfunction reported from these floors.
10. Floors 4,5,6,7 constitute 80% of completion not meeting expectation.
11. No A/C malfunction was reported at level 9.
12. Higher percentage of users’ unhappiness associated with high and medium priority.
13. Higher level of unhappiness related to completion not meeting expectation with a focus at too_hot and too_cold adjustment activities.
14. The department of Cost centre 0 reports 42% of CM work.
15. Department 26462 only reports A/C malfunction. (all 18 cases)
16. For most costcentre = 0 (45 out of 47), the jobtype = CM
17. For floor 5,6,7, workOrder_Status always was completed

- Results extracted from the available maintenance data of the asset type 2: Thermostatic Mixing Valves

1. High priority works constitute of 55/101 monthly work, 22/101 6mthly work, 24/101 12mthly work. (All 6mthly and 12mthly works are of highly priority)
2. 12mthly work happened in the middle of the year – June-Sept, while all 6mthly work is carried out in December.
3. All monthly and 12mthly works were completed. Parts of 6mthly works (50%) were outstanding.
4. All 6mthly and 12mthly works did not meet the expectations of completion date.
5. Monthly work was identified as TMV004, 6mthly = TMV002, 12mthly = TMV003
6. All maintenance of thermostatic mix valves happened at Level 4.
7. All monthly works were supposed to be completed in 0.5 hours and cost $10. All 6mthly and

12mthly works were estimated to be completed in 2 hours and cost $29.

8. All high priority works did not meet the expected completion data.
9. All medium priority works were completed on the expected completion data.
10. There was a trend in emphasizing the maintenance of thermostatic mix valves recently – with increasing of workorder_No, the priority was getting higher.
11. All works between August and December did not meet the expectation of the completion date.
12. All 6mthly works (TMV002) were carried out in December.
13. With higher priority works there was a low level satisfaction of work completion.
14. High priority works take more than 0.5 hour to finish while total hour the duration of <=0.5 must belong to medium priority works.
15. 50% of 6mthly work did not finish (these unfinished 6mthly works' WorkorderNo> 725085)
16. There is an incremental relationship between the work priority, the estimated time to complete the work and the associated budget.
17. There is a pattern in relation to the work priority, level of meeting the expectation and the frequency of the task TMV004.
18. 100% of medium priority works occurs on monthly maintenance and relates to a specific task TMV004.

- Results extracted from the available maintenance data of the asset type 3: Battery Chargers

1. All outstanding works took place at the end of work order list around December 2002.
2. Asset "EPG0101" belongs to cost centre "1000" while the cost centre for the other asset were not available;
3. There is fee charge with asset "EPG0101" while no charge for "EDG1000-01".
4. For all tasks with work order No > 66195, and some tasks with work order No between 48002 and 66195, completions did not meet the expectation of completion date.
5. 57% of workorders were completed within the expected completion date.

5.7 System Components and Results Summary

The detection of previously undiscovered patterns in the maintenance data can be used to determine factors such as the cost effectiveness and expected failure rate of assorted building materials or equipment in varying environments and circumstances. These factors are important throughout the life cycle of a building, and such information could be used in the design, construction, refurbishment, and maintenance of a building site, representing a substantial decrease in cost and increase in reliability.

In order to advance the testing and the feasibility of the proposed approach, a systems prototype was further developed and tested using the building maintenance databases and available CAD drawings. The development of a fully operational prototype system is presented in the following section and the results of further testing are provided in Section 7.

6. AIMM PROTOTYPE SYSTEM

6.1 Architecture of AIMM

As a result of the preliminary data mining demonstration and testing the integration of all system's components and modules were integrated and refined to develop the Agents for Improving Maintenance Management (AIMM) prototype system. The AIMM prototype system includes: an object oriented 3D CAD model of a building modelled in the ArchiCAD package, and a maintenance database in a standard SQL (standard query language) format. The architecture of AIMM was developed to include three agents: Interface, Maintenance and Situated agents, as illustrated in Figure 7. The roles of these agents include:

(a) The appropriate mapping between the building assets of the building model in the virtual environment is maintained by the *Maintenance Agent* that connects data contained within the Maintenance Database with data contained within the EXPRESS Data Manager (EDM) Database via the virtual environment, Active Worlds.

(b) Linking Data Mining techniques to building models in a virtual environment (Active Worlds) is achieved via a *Maintenance Agent* that accesses the maintenance database and applies its mining algorithms on it.

(c) Linking knowledge development with the building model in virtual environments is carried out by the *Situated Agent* that assists in improving maintenance management by providing life cycle implications as feedback whenever building assets (mechanical and electrical elements) are selected in the building model in the virtual environment (Active Worlds).

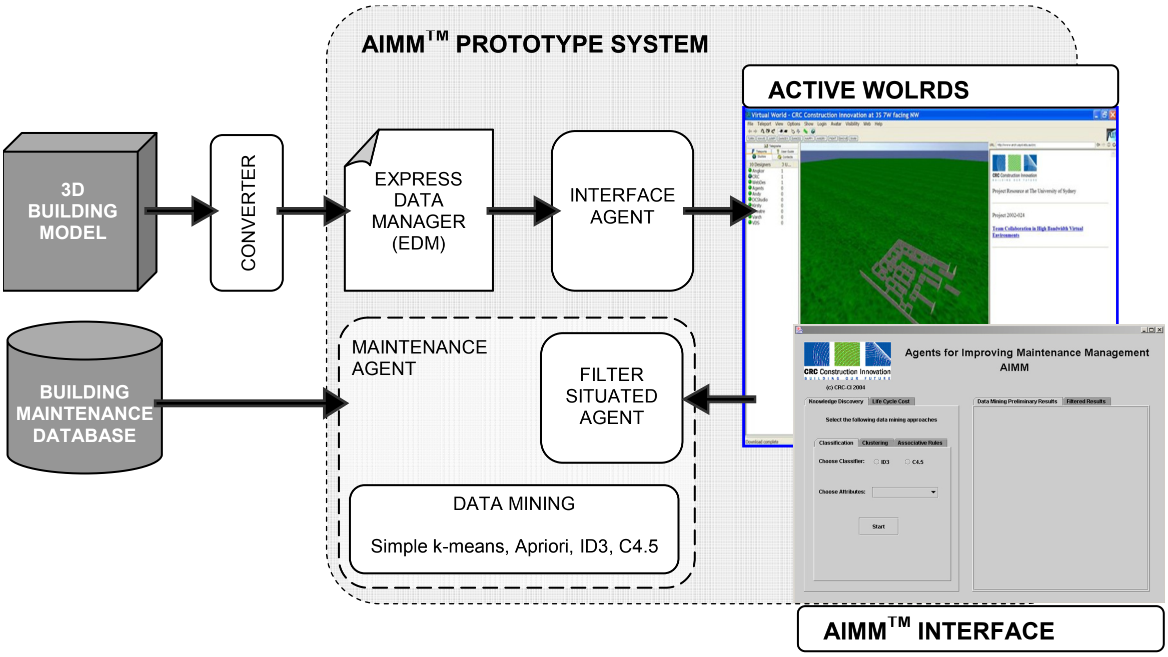

Figure 7. The architecture of AIMM

Feedback of useful knowledge can be discovered by the Maintenance Agent in the application of the four data mining techniques and algorithms (see in Section 5.5) in order to discover various classification of knowledge. The data mining algorithms and the link between its knowledge development and the building model in a 3D virtual environment has been fully implemented.

The following four phases reflect the all key composites (outlined in Section 5) of this project:

- Phase 1: involves the manual pre-processing of data, which removes noisy, erroneous and incomplete data to derive important attributes from original raw data. For example, the raw text description of time of work orders “1/12/2001” which need to be converted to a meaningful attribute such as “month”. Moreover, various “testing” algorithms are run through the maintenance data to find out the suitable data mining approaches. From Phase 1, the quality of the data can be improved.

- Phase 2: adopts the EDM interface agent that was developed in CRC-CI “information flow” project in converting IFCs (Industry Foundation Classes) object model into a Renderware (RWX) format, so as to be presented at the virtual environment (Activeworlds). The virtual environment provides a collaborative multi-user interface and more importantly, a means for the user to walkthrough 3D object model at a real time. The user is able to navigate and select a building asset type to explore useful knowledge.

- Phase 3: instantiates the maintenance interface agent and the maintenance agent. Once the user decides to select a certain building asset (right clicking on the object representing this building asset at Active Worlds), the maintenance interface agent is invoked to load related data from database. The Maintenance Agent performs data mining on the selected asset type. The four mining algorithms that have been implemented in the system prototype of AIMM are the same as those demonstrated in Section 5.6, i.e., clustering using “SimpleKmeans”, associative rules learning in “Apriori”, classification using “C4.5” and “ID3”.

- Phase 4: A situated agent is activated. This is a software agent that performs post-processing of the mined results. The situated agent filters out irrelevant patterns based on the heuristic rules. This architecture has been implemented in the system prototype of AIMM using Active Worlds as the virtual environment platform and developed using Java programming language. Table 12 provides a sample of some of the kinds of heuristic rules defined for the filtering agent. Refer to Final Technical Report for a more detailed list.

Table 1: Heuristics for Air Handling Units

If Clause	Filtering Display
If the result contains string patterns matching “despcriptionofCause = too_cold ……. | floor = 7 || priority = H:N|| priority = M:N|| priority = L:Y”	Within the complaints related to air conditioning “too_cold”, all the “high” and “medium” priority jobs did not complete within expectation. All the “low” priority jobs were finished on time; This may reflect lack of supervision of high and medium air conditioning maintenance works in facility management.
If the result contains string matching “floor=4 …… |month = Dec: O”	For floor 4, air handling unit works that occur in December are not likely to be finished – inspections needed to locate the causes of this failure.
If the result contains string matching “department = 26462 | workorderNo<=50461:1(4.0) | workorderNo > 50461:7(11.0)”	Department 26462 resides only at 1st and 7th floor – this tell you locations of some department
If the result contains string matching “floor=7 25 = = > causeofrepair = A/C_Malfunction 25”	All air handling units maintenance works in floor 7 belong to A/C malfunction – failures concentrate on a particular floor, needs inspections.
If the result contains string matching “deparment = 26462 18 = = > causeofrepair = A/C_Malfunction 18”	All air handling units maintenance works in department 26462 belong to A/C malfunction – failure abnormally concentrate on a particular department, which needs more inspections.
If the result contains string matching “deparment = 21271 16 = = > floor = 6 16”	Department 21271 resides only at 6th floor – this tell you the location of the department

6.2 AIMM Interface and User Scenario

The user is able to navigate a 3D model within a real time virtual environment as shown in the top left of Figure 8. The user is able to instantiate the AIMM prototype system by “right-clicking on the desired building asset or component, thereby presenting the main Maintenance Interface as shown in the bottom right of Figure 8.

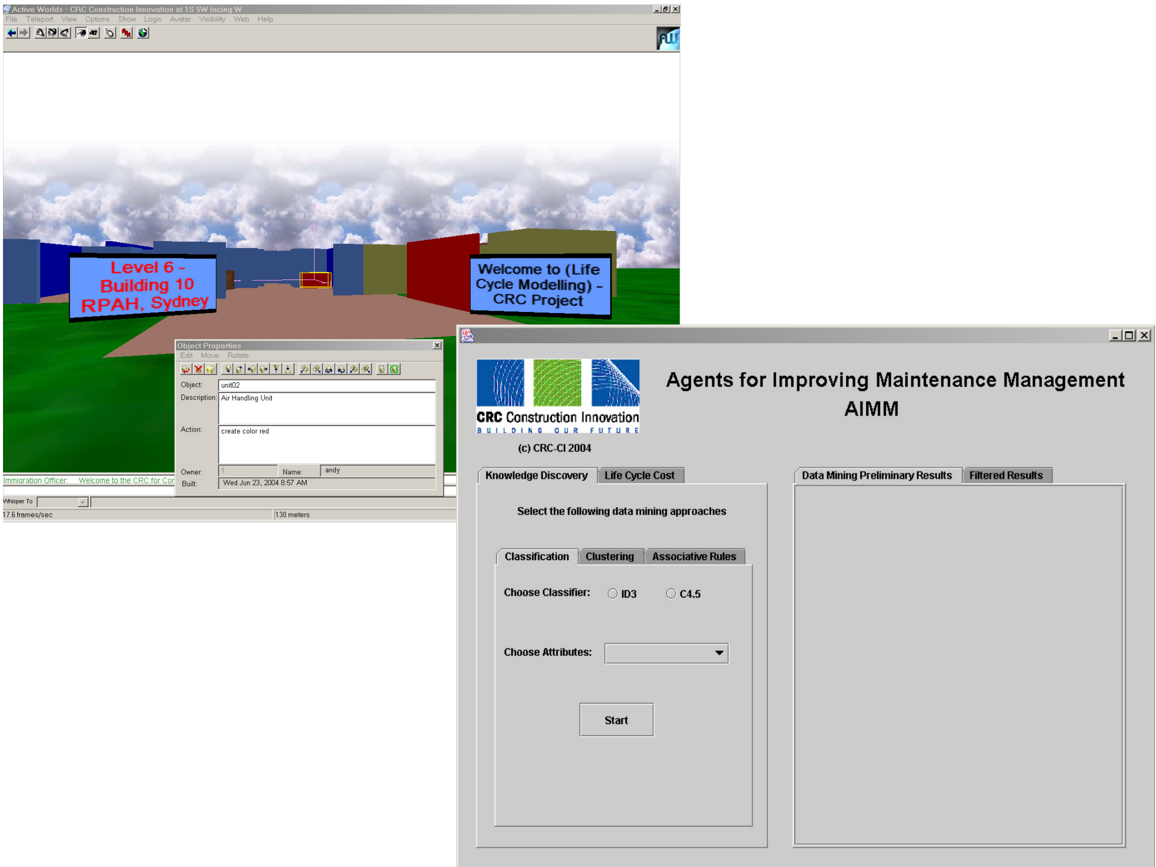

Figure 8. Selecting an asset type in Active World instantiates the maintenance interface agent

The main Maintenance Interface is divided in to equal parts. On the left hand side are two stacked panels: (i) Knowledge Discovery and (ii) Life Cycle Analysis. Here we will only deal with the Knowledge Discovery panel since the Life Cycle panel will be discussed later in Section 8.4. In the Knowledge Discovery panel are located three stacked sub-panels: (I) Classification, (ii) Clustering and (iii) Associative Rules. These panels provide a range of ways for using each different algorithm. This provides the user with greater flexibility and scope since the user may test a variety of data mining approaches for each type of algorithm. On the right hand side are another two stacked panels that are dedicated to reporting results. Results are reported in two ways. The panel named Data Mining Preliminary Results displays the results of the chosen algorithm in their “raw” form. The panel named Filtered Results displays the results in their interpreted form using domain derived heuristics. The overall data mining interface is shown in Figure 9 and illustrates the hierarchy of stacked panels for the different data mining scenarios.

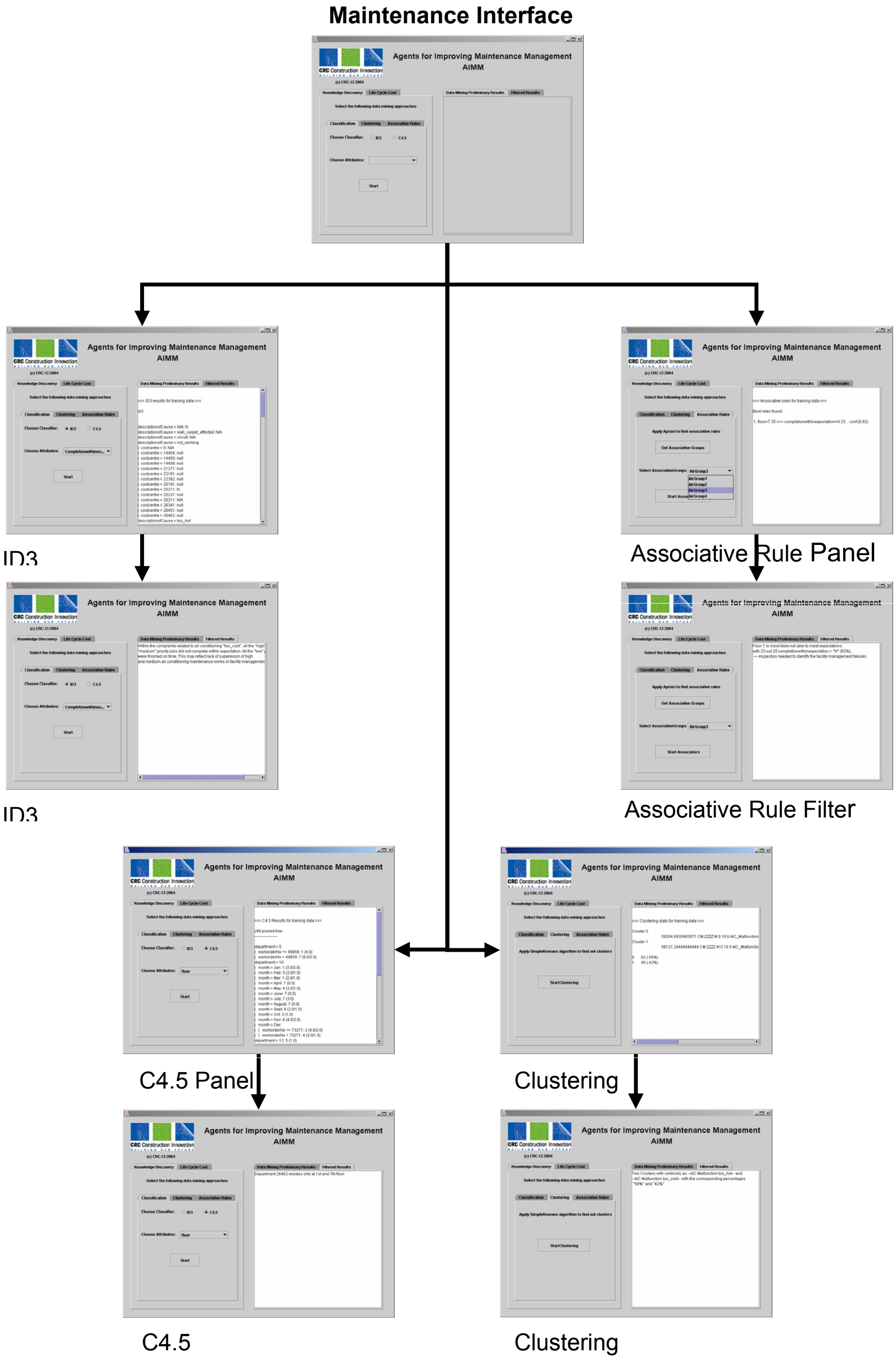

Figure 9 Overall Interface and user workflows

The following scenario is proposed during typical interactions between a user and the proposed system in the previous section. In this scenario, we shall use an Air Handling Unit (AHU) as the chosen building component that a user wishes to apply Data Mining on. The following sequence of actions is then followed:

- The user navigates the building in a real-time and online 3D virtual environment as shown in Figure 10;
- Once the user selects a building asset type such as the Air Handling Unit the object property window pops out describing general information of the selected object as shown in Figure 11;

- AIMM invokes the Interface and the main window pops up to allow selection of algorithms as illustrated in Figure 12;
- The user explores a variety of data mining algorithms and chooses the desired algorithm by clicking on a checkbox or button and running the algorithm as shown in Figure 13;
- AIMM invokes the Maintenance Agent running the algorithm and results are reported first in the Data Mining Preliminary Results panel as illustrated in Figure 14;
- User selects Filtered Results in order to access post-processed knowledge and an example of filtered knowledge is shown in Figure 15.

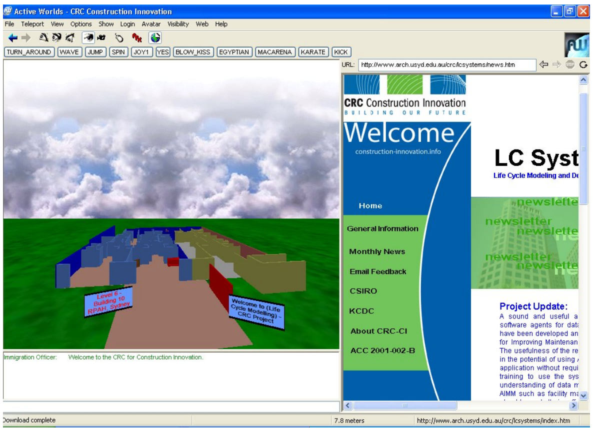

Figure 10. The primary interface of AIMM in an interactive network multi-user environment.

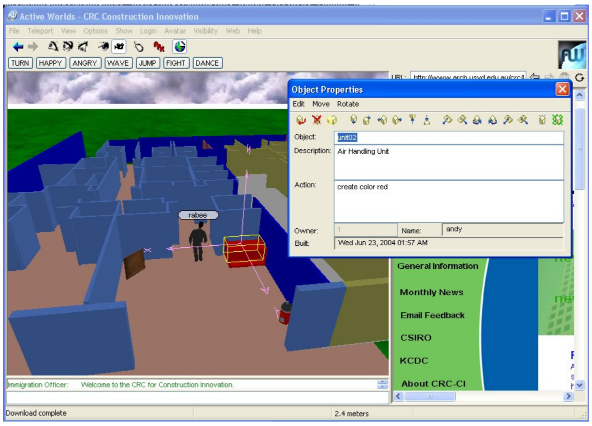

Figure 11. User selects building asset (the Air Handling Unit) and an object property window pops up.

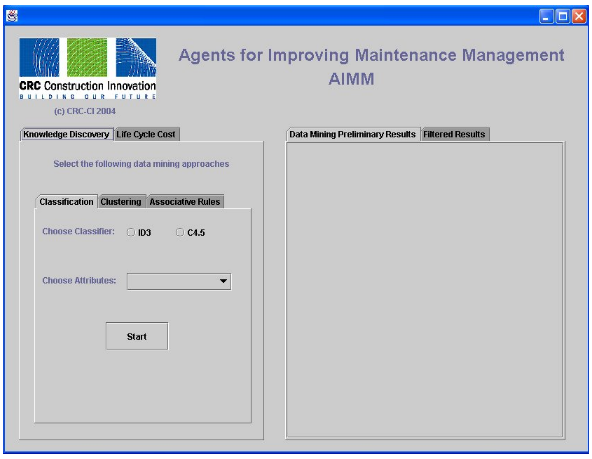

Figure 12. The AIMM prototype system is instantiated once a building asset type has been selected.

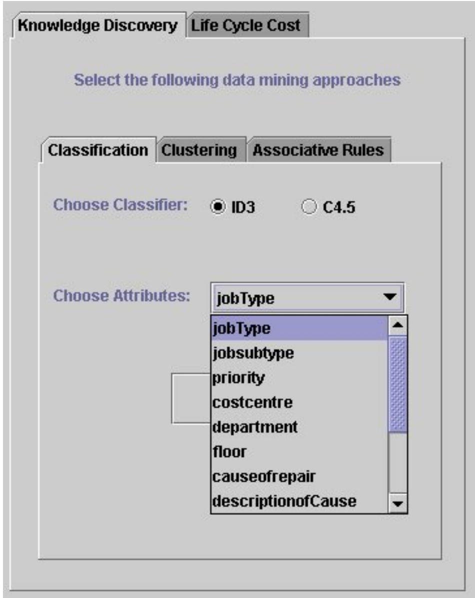

Figure 13. Data mining techniques and different attributes for the user to choose from based on focus and interest.

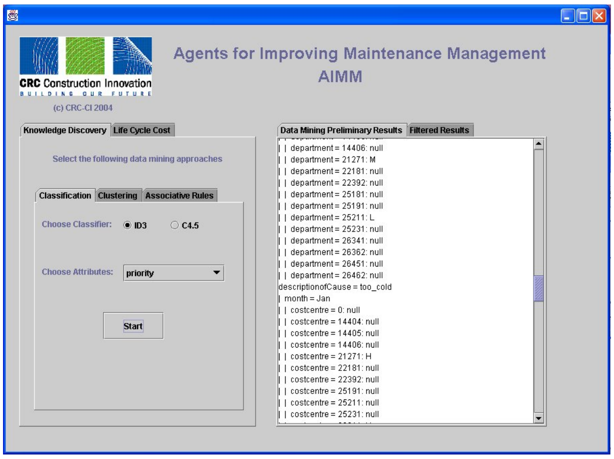

Figure 14. Preliminary results of applying the ID3 with the "Priority" attribute on AHU maintenance data.

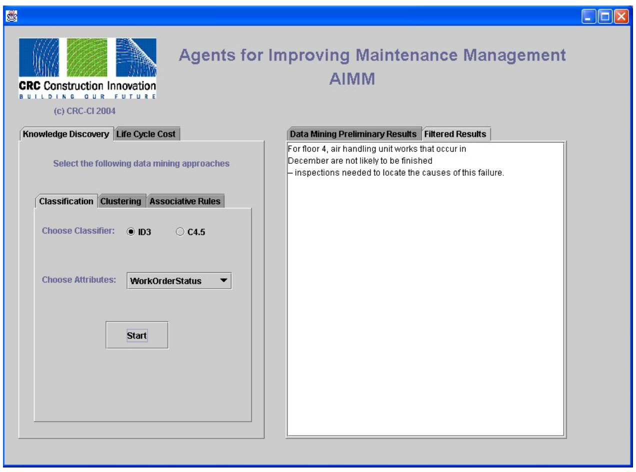

Figure 15. An example of the filtered knowledge presented to the user from the preliminary results of applying the ID3 with the "Work Order Status" attribute on the AHU maintenance data.

7. DEMONSTRATION OF AIMM™ ON BUILDING NO. 10 AT RPAH

7.1 Demonstration of using AIMM on Building no. 10 at RPAH

This Section illustrates a demonstration of a run of the AIMM™ prototype system. This demonstration applies the system on Building No. 10 at RPAH, Sydney. The maintenance data was provided by the Engineering Division of the Central Sydney Area Health Service (CSAHS) and details of the maintenance data can be found in Section 4.2. Maintenance data for the last two and a half years is available in SQL format and contains data that is highly detailed and structured. There are approximately 5,000 work orders recorded for Building 10 in the period from 1 January 2001 to 9 December 2002.

AIMM starts by converting and presenting the 3D Model of Building No.10 in the virtual environment (Active Worlds). The user may navigate the 3D model within a real time virtual environment. We may assume that a user resolves to apply Data Mining on the Air Handling Unit (AHU). Once the user right clicks on the AHU object in the 3D model, the AIMM™ system invokes the Maintenance Agent Interface that activates the Maintenance Agent to apply the four Data Mining techniques and the Situated Agent presents the learned knowledge in a WEKA Learning Results window as shown in Figure 16.

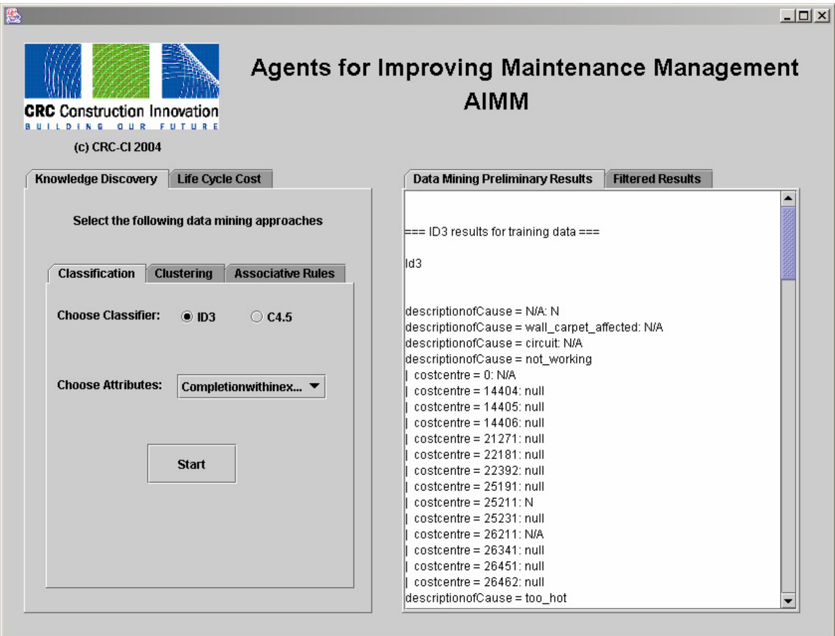

Figure 16. Demonstration output (display Part I) of applying Data Mining on the maintenance data of the Air Handling Unit at Building no. 10, RPAH.

Visual data analysis and data mining techniques were applied on two selected data sets: air handling units and thermostatic mixing valves at Building 10. The evaluation of the results obtained from mining the maintenance data of the above asset type and its impact on improving the maintenance of existing buildings and the design of future facilities is summarised in Tables 2. For more detailed results refer the Final Technical Report and Appendix F.

Table 2. Evaluation of results for air handling units and impact on improving maintenance of existing and design of future building facilities.

Data Mining Technique	Rules Obtained	Potential Impact on Facility Maintenance and Design
Visual Analysis	Approximately all "A/C malfunction"belongs to high and medium priority.	A/C malfunction" is of a major concern in guiding the allocation of maintenance resources.
	"A/C malfunction" is concentrated on the problems of: *too_hot* 32%, *too_cold* 28%, *not_working* 7.5%.	Temperature should be automatically adjusted and a provision of self-reporting faults equipments should be considered.
	The lowest levels of "A/C malfunction" took place in August followed by June and April while other months share similar high rate of "A/C malfunction".	
	The maintenance work on 4, 5, 6, and 7th floors constitutes most of the reports of A/C malfunctions, with 86% of A/C malfunction reported from these floors.	
	Approximately all the descriptions of *too_cold* or *too_hot* were associated with high or medium priority.	The appropriate temperature is of high priority from users' perspective.
Visual Analysis + Decision Tree Algorithm (C4.5)	All 7th floor jobs were of high and medium priority and the cause of repairing was "A/C malfunction".	Investigate the possibility of poor design or maintenance of air conditioning function in 7th floor. A special attention in the design should be given to a specific floor due to its high demand of corrective or preventive maintenance or special design of A/C.
Association Rule Algorithm	For floors 5, 6 and 7, the workOrder_Status was always completed.	Benefiting from successful maintenance practices including both equipments and labour is useful to achieve a high level of an overall maintenance performance.

7.2 Discussion

From the result generated through a single run of these algorithms, it can be observed that the rules generated are relatively interesting. The rules obtained were pulled out manually in a strategic manner and an example is shown in Table 3. The pull process involved cleaning and modification of data files; and selecting various attributes alternatively. For example, when the associative rule algorithm is used on the data set meaningless and less useful rules were produced. Whereas if a small file (containing just 2 to 3 relevant attributes) is fed into the same algorithm, much more interesting rules are discovered. Currently the four algorithms that have been implemented can only be automated according to the selected attribute for classification. For example, if the attribute "outlook" is selected for splitting in the Weather database a different decision tree is generated from selecting the "humidity" attribute for classification.

Table 3. An example of some of the results pulled out manually when applying data mining techniques on air handling units.

Data Mining Technique	Rules Obtained	Potential Impact on Facility Maintenance and Design
Decision Tree Algorithm C4.5, and Association Rule Algorithm	All 7th floor jobs were of high and medium priority and the cause of repairing was "A/C malfunction".	Investigate the possibility of poor design or maintenance of air conditioning function in 7th floor. A special attention in the design should be given to a specific floor due to its high demand of corrective or preventive maintenance or special design of A/C.
	Department 26462 only reports A/C malfunction. (all 18 cases)	A special attention should be directed to certain places in the building wherein maintenance work is required more often.
	96% jobs for cost_centre = 0 is CM (corrective maintenance).	
	For floors 5, 6 and 7, the workOrder_Status was always completed.	Benefiting from successful maintenance practices including both equipments and labour is useful to achieve a high level of an overall maintenance performance.

The results indicate that data mining does not automatically extract all available knowledge that is embodied in a data set. Although it may sound at first appealing to have an autonomous data mining system, in practice, such a system would uncover an overwhelmingly large set of patterns, and most of the patterns discovered in the analysis would be irrelevant to the user. Results indicate that a more realistic scenario would be to communicate with the data mining system, using additional questions to examine the findings and direct the mining process; some of these questions might include (Morbitzer, Strachan and Simpson, 2003):

- What is task relevant data?
- What kind of knowledge do I want to mine?
- What background knowledge could be useful?
- How do I want the discovered patterns to be presented?

Consequently, the first analysis will not necessarily provide the required knowledge since the user might have defined a mining exercise that does not reveal important patterns. Hence, the analysis needs to be refined. The creation of different mining exercises is supported by a very flexible definition of a mining task. The user can therefore quickly change variables to be included in the mining run, in combination with filters that can be defined for all variables.

Herein are some of the scenarios utilised for improving the results obtained from using data mining techniques on maintenance data for Air Handling Units.

Applying ID3

Prior to applying ID3 on air conditioning unit's maintenance data, convert data file into nominal value and replace all the missing values with "N/A" strings.

- The ID3 algorithm takes file CMID3normic.arff and some results are list as below:
    - ID3 classifies on attribute "Completionwithinexpectation"

For “too_cold” descriptionofCause, all the “high” and “medium” priority job did not complete within expectation. All the “low” priority jobs were finished on time;

- ID3 classifies on attribute “WorkOrderStatus”

  For floor 4, works that occur in Dec are not likely to be finished (with WorkOrderStatus = “o”)

Applying C4.5

Classify against the “floor” and “descriptionofCause” attributes

- The C4.5 algorithm takes file AirCondCMupdate4Nov.csv and some results are list as below:
- **Department 26462 resides only at 1st and 7th floor:**
- **Department 21271 only reside at 6th floor:**
- **Department 26462 only reports A/C malfunction.**

Applying Apriori:

Prior to mining the maintenance data using the Apriori algorithm, apply attribute evaluator “CfsSubsetEval” and search method “BestFirst”. The data file is divided into sets of data files.

- Applying Apriori on the data set of (priority, department, floor, causeofrepair, descriptionofCause, workormaterial), the results include:

  All works in floor 7 belong to A/C malfunction
  All works in department 26462 belongs to A/C malfunction
  Department 21271 only resides at floor 6
- Applying Apriori on the data set of (priority, floor, completionwithinexpectation), some of the learned rules include:
  - **Floor 7 is most likely not able to meet expectations: with 23 out 25 completionwithinexpectation = “N” (92%)**
  - **Apply Apriori on relation6.arff (jobType, costcentre, causeofrepair): 96% jobs for costcentre = 0 is CM**

8. RECOMMENDATIONS AND FUTURE WORK

8.1 Data Mining Requirements

As an integrated approach to facilities management, the use of data mining implemented in the AIMM system is a novel approach. Toward this end, the AIMM prototype system requires reliable data and must be able to translate, query, mine and filter this data quickly. For the AIMM system to reach its full potential future research and development must consider data collection, data quality as well as the accuracy and long-term maintenance of the data sets themselves. It is also critical that these aspects also be considered in relation to overall maintenance goals and objectives. Therefore, there is a need to address the (i) necessary specification criteria that enables the identification of user goals, (ii) resulting type and quality of data required, and (iii) method of data collection to ensure useful information is maximised and human error is minimised.

While the outcomes of both the industry data (Section 4.2) and data mining techniques (Section 5 and 7) surveyed and tested in this demonstration, were separated for purposes of discussion, they are closely linked in practice. Indeed, the power of the data mining scenarios implemented within AIMM come largely from the ability to simultaneously access and process both large quantities of data and high quality 3D building components and their corresponding maintenance data. Collection of maintenance data to be stored in the database is typically a large expenditure in implementing any kind of maintenance management system. For this reason in conjunction with general data mining requirements, maintenance data must be as reliable and accurate as possible. Complete and accurate documentation of all maintenance operations is important to assure the integrity of the AIMM system, the reliability of subsequent analyses as well as the ability to maintain the system over time. Remembering that the utility of data mining is a direct function of the data contained within the system, a commitment must be made to maintain the database itself. Only dedicated care of the database can ensure that the data mining scenarios and subsequent analysis will produce results of the highest order

Data mining is a life cycle process in which the knowledge obtained will affect interpretation of data gathering, in terms of the availability and priority of certain attributes. This process is established via the: setting up of mining goals → pre-processing data → data transformation → data mining → evaluation → refinement of data requirements (according to the knowledge obtained). Therefore to secure the more effective use of the AIMM system, future research should begin with addressing a variety of different management scenarios for a range of maintenance goals in order to identify the resulting data requirements.

By investigating a variety of different facilities maintenance scenarios it may be possible to build a better understanding of the requirements of maintenance data collection methods during which crucial building component and attribute data is obtained. Since in practice, initial maintenance goals drive the way in which data is collected, there is a real need for further consultation with Industry partners to identify critical relationships between goals, data requirements and their collection. As a result, data mining the building component's attribute data will yield more effective and helpful results.

In addition, whilst the filters developed here for performing secondary analyses are derived from domain knowledge, this phase does not constitute a complete or comprehensive listing of heuristics and further development is required through consultation with a variety of facilities managers, maintenance, design engineering and construction experts.

The quality of data mining results is a direct consequence of the overall goals of maintenance managers and designers. The following themes are commonly pursued in facilities maintenance management and are considered useful in developing further data mining scenarios, identifying all essential component attributes and structuring the appropriate strategies for data collection:

- Budget and Cost-based analysis – Existing costs, ongoing costs and cost prediction.
- Location-based analysis -- Maintenance anomalies occurring at particular locations within the building or across a complex of buildings.
- Error and Failure-based analysis -- What kind of errors and failures are mostly reported and corrected.
- Temporal and Seasonal analysis – Frequency of failures, occurrence and changes over time, i.e., specific months/ weeks/ days.
- Customer and User satisfaction-based analysis – User satisfaction of facilities maintenance performance and conditions.

During the data acquisition stage, special attention must be paid to data accuracy and quality. Most data collected for a building comes from site managers, maintenance officers and repairmen. In the Industry data provided neither the type, nor scale of a maintenance operation relates to the relative accuracy or precision of the collected data. Generally the data collection survey was composed of the following attributes:

- Work Order No,
- General Job Type,
- Job Priority,
- Floor,
- Room Number,
- Asset Number,
- General Repair Description
- Additional Text

In many cases the "Additional Text" field is a preferred method of entry and often describes valuable information in written form. This kind of entry requires extensive pre-processing before it can be translated into meaningful data that can be used in data mining. The minimum data collection requirements that must be satisfied, in terms of both quantity and quality, and for any type (or scale) of corrective, scheduled or preventative maintenance operation include:

1. Element level maintenance data, involving a three layer tree coding consisting of:
    - WIC Number: a 5 digit number that uniquely identifies a building site
    - Building Code: a 3 digit number that identifies an individual building, except for the code 000, which identifies elements pertaining to the overall site
    - Building Component: a 4 character alphabetic code that identifies the type of system (air conditioning, electrical, etc.)
        - Element Number: a two digit number that uniquely identifies a specific asset
2. The cost of repair or replacement associated to each maintained building asset.
3. The cause of repair or replacement for each maintained building asset
4. Time required to fix or replace each maintained building asset

In particular, if maintenance data is collected about any kind of building component, without all element level data, its utility in data mining is limited.

In addition, data mining does not automatically extract all available knowledge that is embodied in a data set. For example, the industrial maintenance data supplied for Building 10 had to be pre-processed according to the following steps:

1. Removal of “noisy data”: This is maintenance data which contains false or constant values. Attributes such as “Task Number”, “Descriptions”, “Extra Text” which cannot be processed efficiently.
2. Re-interpretation of “noisy data”: useful information is extracted from the deleted attributes “Description”, “Extra Text” can be re-interpreted to form new attributes “causeofrepair”, “descriptionofCause” and “workordermaterial” ;
3. Derivation of new information: from an existing attribute new information can be obtained via interpretation or via the combination number of existing attributes new information can be extracted. For example, attribute “month” is derived from “Start Date” and “Completion Date”, “Completionwithinexpectations” is created from cross-comparing “Completion Date” and “Estimated Completion Date”;
4. Re-formatting of data: The maintenance data received was in Excel file format. The Excel file must be converted into ARFF file format to be processed by the machine learning algorithms used in the AIMM system.

Crucially, the following data was missing from the existing maintenance data provided by Industry:

- Cost related data – no information was provided of how many resources were involved for each maintenance job;
- Human resources – personnel information is required in order to evaluate work performance;
- Failure information – detailed failure information and related analysis is required for each corrective maintenance job;
- User satisfaction – information on user satisfaction can be derived from some pre-processing techniques, however more detailed information is required to obtain more meaningful analysis.

Future development of the AIMM system must also allow users to: (i) easily incorporate (import) data from outside sources, (ii) easily update and alter data, and (iii) ask data-related questions of (or query) the database. Since most database management systems such as the EDM system incorporated within AIMM, provide these capabilities, they should also be available through the AIMM systems interface.

Since data must be imported and reformatted from outside sources, this process should be made available through the AIMM system. Currently, data imported into the AIMM system must come in ARFF file format which is similar to CSV format. Only an ARFF file format can be processed by the machine learning algorithms used in the AIMM system. Since maintenance databases often come in the form of SQL, Excel and other similar file formats the AIMM system should provide a process for importing and re-formatting a number of these formats into ARFF files automatically.

A common task in facilities management is updating or editing the database. Since no user can foresee all future data needs and applications, the AIMM system should incorporate ways to easily modify, refine, or correct the database. The AIMM system should also allow for error checking as new maintenance data records are created or existing ones are updated. Not all errors can be eliminated in this way, however, so care must be taken when collecting, automating, and changing the database. Attribute data are seldom static and therefore, maintain the value of the data depends on updating capabilities.

The data mining algorithms used in the AIMM system have specific data type and format requirements. All algorithms receive data in ARFF file format and the following outlines the general requirements of each algorithm implemented in AIMM:

- **Classification Algorithm: ID3** – Prior to applying the ID3 algorithm on maintenance data, all necessary data must be converted into nominal values and all the missing values must be replaced with “N/A” strings.

- **Classification Algorithm: C4.5** – Although the C4.5 algorithm allows missing values, prior to applying C4.5 algorithm on maintenance data, missing values should be replaced with "?" symbols. The C4.5 algorithm can be applied to both nominal and numeric values.
- **Clustering Algorithm: SimpleKmeans** – Like the C4.5 algorithm, SimpleKmeans allows missing values and prior to applying SimpleKmeans algorithm on maintenance data, missing values should be replaced with "?" symbols. The SimpleKmeans algorithm can also be applied to both nominal and numeric values.
- **Associative Rule Algorithm: Apriori** – Prior to mining the maintenance data using the Apriori algorithm, the data file must be divided into natural associated attribute groups that can be further divided into sets of new data files to allow associations to be found. Like C4.5 and SimpleKmeans, the Apriori algorithm takes nominal value and allows missing attribute value (to replaced with "?" symbols). In addition, an attribute evaluator “CfsSubsetEval” must be applied and the “BestFirst” search method must also be used.

8.2 Requirements for 3D CAD Modelling

The 3D CAD model of a building and its asset should be modelled as object oriented for each component and be transferable to IFCs. The principle benefit of IFC’s is their object description. The IFC protocol preserves the full 3D geometric description in 3D and distinguishes its location and relationships, as well as all the properties and parameter values of each element. This provides access to accurate geometry of building systems, components, structural elements and properties relevant to facilities maintenance requirements.

Since the process of data mining in a 3D virtual environment requires an adequate building model to begin with it is therefore necessary for architects and designers to understand how to better structure building models for data mining a variety of scenarios since information provided within CAD and IFC element descriptions are inadequate for many facilities maintenance descriptions and requirements. It is necessary to extend design information in object oriented CAD systems to support descriptions defined in facilities maintenance operations.

8.3 Potential Industry Partners Survey

The potential benefits gained from the AIMM prototype system have prompted a preliminary survey of large commercial businesses and industries in order to gauge interest levels of such a tool. A variety of potential industry partners were contacted including:

- Sydney University, Facilities Management Office (FMO);
- Woolworths Pty Ltd, Facilities Management;
- AMP Capital: Commercial & Industrial Management Pty Ltd.;
- Resolve FM Pty Ltd, (Contact: John Smith).

In the initial round of discussions, feedback from the above businesses and industries was very positive. The meeting outline included a presentation of the project, research completed to date, current research maintenance data requirements, and an illustration of the benefits to industry partners. In all preliminary meetings the opportunity for active project involvement was presented. A summary of the responses taken from the minutes recorded at all four meetings can be found in the Final Technical Report.

8.4 Integrating Life Cycle Cost Analysis in AIMM

Life cycle costs (LCC) are summations of cost estimates from inception to disposal for both equipment and projects as determined by an analytical study and estimate of total costs experienced during their life. LCC comprises initial acquisition cost and also may contain other costs such as operation costs, maintenance costs, logistics costs. Such costs are usually higher than the original acquisition cost. The major objective of LCC analysis is to choose the most cost effective approach from a series of alternatives so the least long term cost of ownership is achieved (Barringer, 1996). Life cycle cost analysis provides strategic planning on refurbishment and enhances information for decision making. LCC helps facility managers in evaluating alternative equipment and process selection based on total costs rather than the initial purchase price. The multidimensional information that LCC presents is merged from hybrid project domains such as management, engineering, as well as finance.

There are various existing life cycle models available for buildings as a whole and for their component systems that demonstrate the multifarious approaches to LCC. Although there is no one model that have been accepted as a standard, there are some areas of commonality. Life cycle cost models form predictions based on several parameters, some of which include a degree of uncertainty, such as the reliability of a part (Siewiorek and Swarz, 1982). Among the inputs whose values could potentially be predicted for each component by the Data Mining system in building LCC (Dhillon, 1989) are:

- mean time between failure
- mean time to repair
- average materials cost per repair
- labour cost per corrective maintenance action
- average materials cost per preventative maintenance action
- labour cost per preventative maintenance action
- spares requirements

The values of input variables, along with their probability distributions, can be predicted for each component, thus allowing for more accurate estimation of average life cycle cost. By predicting failure rates and repair costs, it is possible to compute the optimal schedule of preventative maintenance for each asset. Existing life cycle modelling systems fail to provide a seamless integration of hybrid information that provides users access to previously unreachable knowledge.

The proposed life LCC formula will be integrated in the system prototype of AIMM™ via a new LCC Agent. The architecture of the AIMM™ system has been developed so as to include the LCC Agent as shown in Figure 17.

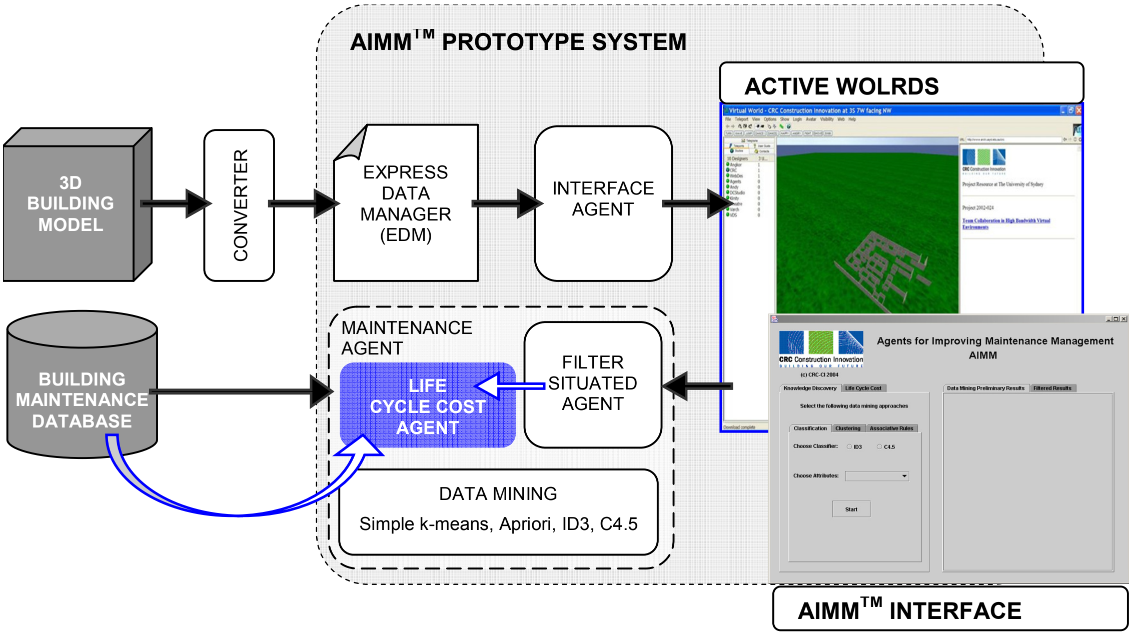

Figure 17. Integrating a new LCC Agent in the architecture of AIMM.

The LCC Agent utilises the results available produced by the Situated Agent and accesses the maintenance database to extract the input variables required in the LCC formula. Figure 18 illustrates the proposed LCC Panel and how it can be integrated with the AIMM prototype system's existing interface.

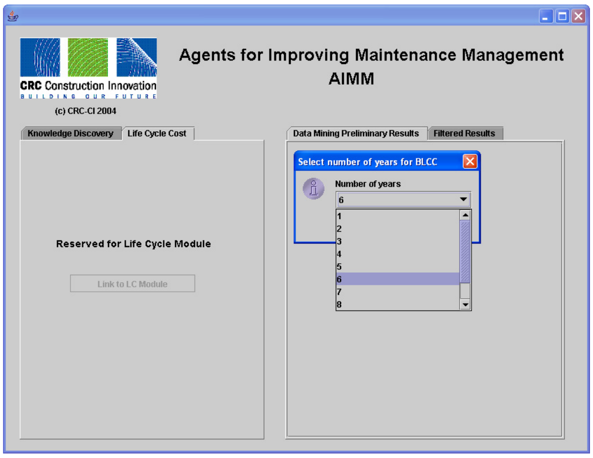

Figure 18. Integrating a new LCC Panel into the AIMM interface.

Some of the possible input variables with the source from where a value may be acquired are shown in Table 4.

Table 4. Examples input variables with the source from where a value may be acquired.

Input Variable	Source
Acquisition Cost	Asset Register
Operating Cost (per year) Includes energy consumption	Manufacturer
Scheduled Maintenance Frequency (per year)	Maintenance Database
Scheduled Maintenance Parts Cost	Structural, Mechanical, Electrical Engineers
Scheduled Maintenance Labour Cost (per hour)	Structural, Mechanical, Electrical Engineers
Mean Scheduled Maintenance Time (in hours)	Maintenance Database
System's Life Time (expected)	Manufacturer
Mean Time Between Failures	Maintenance Database
Corrective Maintenance Parts Cost	Structural, Mechanical, Electrical Engineers
Corrective Maintenance Labour Cost (per hour)	Structural, Mechanical, Electrical Engineers
Loss of Service Cost (per hour)	Structural, Mechanical, Electrical Engineers
Mean Repair Time (in hours)	Maintenance Database
Mean System Response Time (in hours)	Maintenance Database
Real Discount Rate	Government or financial institution

Thus, Life cycle cost (LCC) analysis may be introduced to provide strategic planning on refurbishment and enhances information for decision making. The life cycle cost formulas proposed here for building assets and components are appropriate for adoption within the AIMM™ systems architecture and an LCC Agent may be implemented in future research.

8.5 Recommendations for the CRC

The construction industry has adapted the information technology in its processes in terms of computer aided design and drafting, construction documentation and maintenance. Hence, the data generated within the construction industry has become increasingly overwhelming. The growth of many business, government, and scientific databases has begun to far outpace human's ability to interpret and digest this data. This issue becomes critical when the high degree of complexity of work flow is taken into account in the decision making process during the lifetime of a building. Furthermore, past experience often plays an important role in building management. Therefore, applying data analytic techniques to efficiently deal with information at different stages of a building life cycle has great potential.

A large number of buildings maintained in Australia rely on efficient facilities management. Maintaining building facilities is a major task for many in the AEC industry, ranging from the facilities manager to the occupant and client to the designer. Table 5 illustrates this range of professions and stakeholders that can potentially benefit from the system demonstrated by this research project and the life cycle phase that would be influenced in terms of improving standards and best practices.

Table 5. Potential users and beneficiaries and related life cycle phases.

	Stakeholders and Users	Life Cycle Phase
Users: Immediate and potential	Facilities and Asset Managers	Management in-use
	Asset Planners	Strategic Planning
	Facility and Building Operators	Occupancy in-use
	Architects and Designers	Design Life Cycle
Beneficiaries: Direct and indirect	Occupants, Tenants, Owners	Post Occupancy
	Investors	Asset Investment
	Building Developers	Building Development
	Environment: Water, Material, Energy, etc.	Ongoing
Flow-on effects to AEC industry	Consultants: Engineers, Researchers, etc.	In-use Operations
	Building and Quantity Surveyors	Procurement
	Builders	Project Delivery
	Manufacturers, Environmental Control Bodies	Procurement
	Project Managers	Construction

Facilities Managers will benefit from the development of this project in terms of the feed back generated from identifying patterns and correlations in maintenance records that are presented as meaningful asset knowledge. Since integrating informed design decisions for improving building life cycle is a complex task at the early stages of designs, designers will benefit from this project in reducing the risk of design mistakes. Both Facilities Managers and Designers will benefit from the automated feed back in reducing time and cost. Further, the project can be applied in other potential areas - including checking a building model against the information required for project management, and cost estimates as detailed in Section 8.4.

This project has developed a prototype to support automated feed back. Therefore future research should look towards applications of the prototype for Industry partners. For continued AIMM™ systems development, a needs-based assessment is perhaps the most important step. Discussions about the systems requirements, an understanding of who are the users and what they demand of the system and an evaluation of the educational and training needs of users are all part of such a needs-based assessment. This process will ensure that proper data sets are collected, that all users understand the technology and its role within the organization, and that the specific goals of maintenance management are identified to generate the proper analyses and output. After a discussion of industry scenarios, and establishment of goals and requirements, work towards education, training, and data compilation (i.e., acquiring and digitising proper data sets) can begin. Based on this approach, valued industry and research benefits can be achieved from further project development including:

- Contribution to long-term scientific and technological research and innovation to Australia's sustainable economic and social development.

- Collaboration between researchers, industry and government, and to improve efficiency in the use of intellectual and research resources.
- Creation of a fully commercialised tool to deliver innovative and sustainable constructed assets to further the financial, environmental and social benefit to the construction industry and the community.

9. CONCLUSION

The development of data mining agents of facilities and building maintenance data in a 3D virtual environment provides useful information for improving the design, maintenance and management of building facilities and guiding future decisions. Virtual environments of building models offer the opportunity for the user to navigate through the model, to manipulate and to interact with its objects. The integration of facilities databases with interactive 3D virtual environments containing building models and data mining techniques provides a visual modelling tool for the simulation and projection of the financial and physical impact of maintenance, refurbishment and major replacement and extension of a building and its components over its life cycle.

A sound and useful architecture and scenario of using software agents for data mining on building maintenance has been developed and implemented in an AIMM™ system prototype. The usefulness of the architecture and scenario lies in the potential of using AIMM™ to maximise the benefits of its application without requiring its users to have an extensive training to use the system or to have a comprehensive understanding of data mining and its techniques. Users of AIMM™ such as facility managers, designers and building developers should be reflecting on the discovered knowledge acquired from the application of data mining on maintenance records and utilise AIMM results to improve overall maintenance management as well as the operation of buildings and their life cycle.

By demonstrating how and what data mining techniques can be applied on maintenance data of buildings this research has discovered and addressed patterns, correlations and useful rules within existing building maintenance data. The results from the initial data mining studies were improved in the final experiments of the prototype system by applying the appropriate data mining scenarios and filters needed to provide the required knowledge. Appropriate filters chosen according to the task were proposed and with future development can be extended further to include a variety of scenarios. The techniques within the tool demonstrate how a building or facility manager can identify meaningful and useful patterns, correlations and trends of knowledge from large amount of building data.

Further, this final report provides results supporting the capabilities, flexibilities and advantages of automated knowledge development that is a result of data mining building models in a virtual environment. Evaluation and testing has highlighted how the connection can be improved between maintenance and design knowledge development in the following ways:

- The combination of a 3D model with maintenance and other asset data facilitates the ability of building designers and owners to visually model the impact of design, maintenance, refurbishment and extension decisions on the building's life cycle cost.
- The linking of a 3D model with maintenance data allows both the facility manager and the designer to gain access to information and knowledge that is currently inaccessible.
- The development of agents for data mining of facilities maintenance and other data provides a method of testing and validating the usefulness and scope of current databases as a platform for guiding future decisions.
- The representation of the facility within the virtual environment provides a basis for linking data mining with emerging technologies (such as connecting to WAP phones and other PDAs both in the office and on site) to address a gap in the construction life cycle.
- The integration of data mining agents into the maintenance process produces timely data for the facility manager on the effects of different maintenance regimes.

- The development of agents (agents are software systems capable of taking autonomous decisions) for data mining of facilities maintenance data in a 3D environment provides proactive information for improving the design, maintenance and management of building facilities.

Applying the AIMM prototype system to the records of facilities provides a tool that is capable of improving the management and maintenance of existing facilities as well as provides valuable knowledge for the design of new facilities. This leads to more efficient and effective facilities maintenance and management through better planning based on models developed from available maintenance data, and therefore results in a more economical life cycle of buildings.

Although the project focuses on mining the maintenance data in which huge benefits go to the facility management, it is not prevented from attempting to fill the gaps between designing and building maintenance in biasing the future design solutions within the scope of the whole coordinated building life cycle. The major contribution of data mining for this project is to provide a knowledge base which is served as a centre bridge. Furthermore, designers and maintenance managers will be better equipped to achieve higher performance by utilising the techniques incorporated in AIMM in their workplace.

For the AIMM™ prototype system to be useful in the AEC industry, there is a set of minimum requirements of building maintenance data that must be satisfied, in terms of both quantity and especially quality (Refer to Section 8.1)

The fundamental directions of expected commercialization potentials of AIMM prototype system involve: (i) enhancing and expediting the traditional maintenance management activities by live and progressive feedback using data mining techniques to improve management of building maintenance; (ii) opening new possibilities in building design and management by exploring 3D multi-user, online, real-time environments; and (iii) providing interface to generally available maintenance databases and 3D CAD software (seamless integration of hybrid information).

Future work of the AIMM prototype system includes comprehensive consultation with Industry experts and the integration of life cycle modelling at the early design stages. The potential benefits of this approach include: (i) recommending quality systems (building assets) which meet users expectations within cost estimates; (ii) recommending systems that are cost-effective to enhance and maintain; (iii) recommending the exploration of alternative concepts and methods to satisfy the need; and (iv) evaluating costs and benefits of alternative approaches to satisfy the basic functional requirements.

.

10. REFERENCES

Barringer, H. Paul and Weber, David P. (1996) Life Cycle Cost Tutorial, *Fifth International Conference on Process Plant Reliability*, Gulf Publishing Company, Houston, TX

Bowers, S. V., Piety, K.R. Proactive motor monitoring through temperature, shaft current and magnetic flux measurements. *White Paper*, Computational Systems, Inc., Knoxville, TN.

Brachman, et al. (1997) Visual data mining: Recognizing telephone calling fraud. *Data Mining and Knowledge Discovery,* 1, 33–46.

Christiansson, P. (1998) Using knowledge nodes for knowledge discovery and data mining. *Artificial Intelligence in Structural Engineering. Information Technology for Design, Collaboration, Maintenence, and Monitoring,* I. Smith, ed., Springer-Verlag, Berlin, 48-59.

Dhillon, B. S. (1989) *Life Cycle Costing: Techniques, Models, and Applications*, Gordon & Breach Science Publishers, New York.

Ding L., Liew P., Maher L., Gero J., Drogemuller, R. (2003) *Integrating CAD and 3D Virtual Worlds Using Agents and EDM*, CMIT, Commonwealth Scientific and Industrial Research

Organisation, Key Centre of Design Computing and Cognition, University of Sydney.

Elder, J., Abbott, D. (1998) *A Comparison of Leading Data Mining Tools,* tutorial presented at KDD-98, New York, August 28, 1998.

Froese T. M., Lloyd Waugh, L. M., and Pouria, A. (2001) Project management in the year 2020, *Proceedings of 2001 Conference of the Canadian Society for Civil Engineers,* Victoria, BC., May 30-Jun 2, 2001. Paper C15.

Goebel, M. and Gruenwald, L. (1999) A survey of data mining software tools, *SIGKDD Explorations*, 1(1).

Kawauchi Y., Rausand, M. (1999) *Life Cycle Cost (LCC) Analysis in Oil and Chemical Process Industries,* adopted from http://www.ipk.ntnu.no/rams/notater/lcc.doc on 28th, Mar., 2003.

Maher, M.L. and Gero, J.S. (2002) Agent models of virtual worlds, *Proceedings of ACADIA 2002*, Pamona, CA.

Morbitzer, C., Strachan, P., Simpson, C.: 2003, Application of data mining techniques for building simulation performance prediction analysis*, in the Proceedings of Building Simulation 2003* (to appear).

Peitgen, H. O., Jurgens, H., and Saupe, D. (1992) *Chaos and Fractals : New Frontiers of Science*, Springer-Verlag, New York.

Siewiorek, D. P., Swarz, Robert S. (1982) *The Theory and Practice of Reliable System Design*, Digital Press, Bedford, MA.

Smith, G. and Maher, M.L. (2003) *User Guide for KCDCC AW Agent Package*, Key Centre of Design Computing and Cognition, University of Sydney

Soibelman, L., Kim, H. (2002) Data preparation process for construction knowledge generation through knowledge discovery in databases. *Journal of Computing in Civil Engineering*, 16(1), 39-48.

Witten, I. and Frank, E.: 2000, *Data Mining: practical machine learning tools and techniques with java implementations* , M organ Kaufman, San Diego