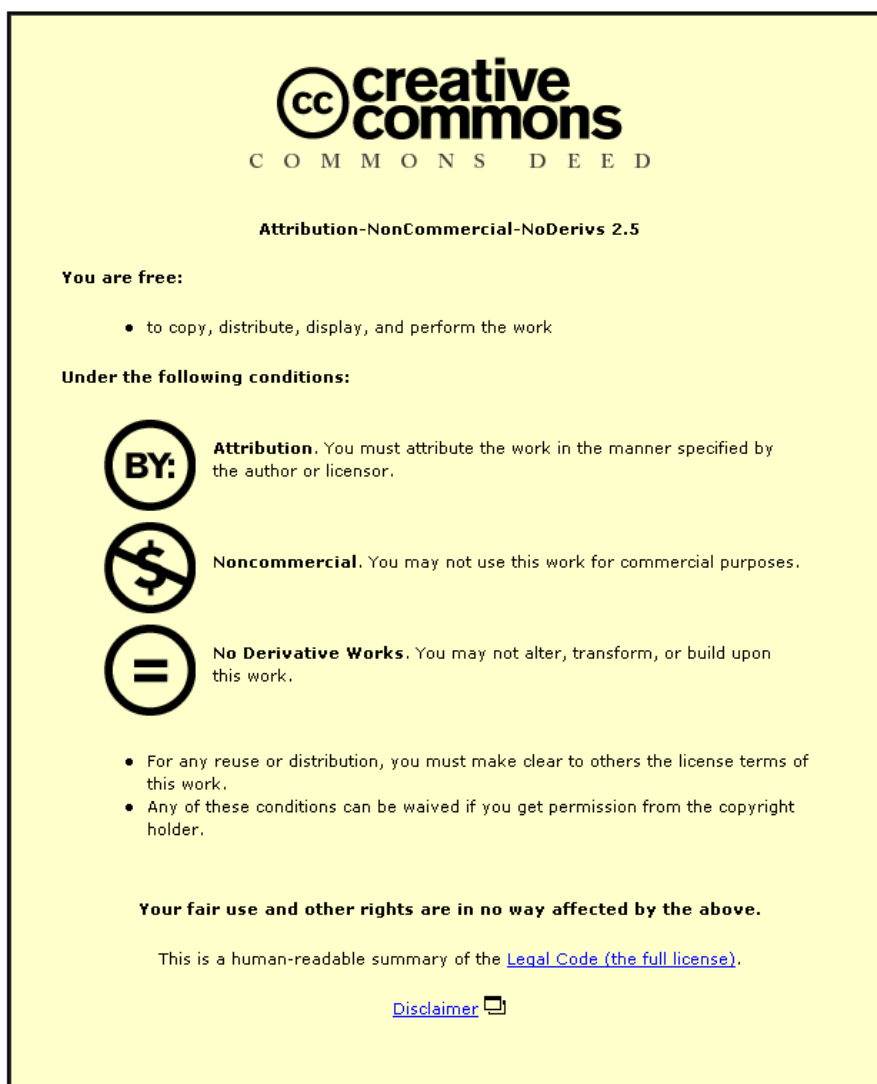


This item was submitted to Loughborough's Institutional Repository (<https://dspace.lboro.ac.uk/>) by the author and is made available under the following Creative Commons Licence conditions.



For the full text of this licence, please go to:
<http://creativecommons.org/licenses/by-nc-nd/2.5/>

An Enterprise Modeling and Integration Framework based on Knowledge Discovery and Data Mining

Elena I. Neaga* and Jennifer A. Harding **

The paper deals with the conceptual design and development of an enterprise modeling and integration framework using knowledge discovery and data mining. First, the paper briefly presents the background and current state-of-the-art of knowledge discovery in databases and data mining systems and projects. Next, enterprise knowledge engineering is dealt with. The paper suggests a novel approach of utilizing existing enterprise reference architectures, integration and modeling frameworks by the introduction of new enterprise views such as mining and knowledge views. An extension and a generic exploration of the information view that already exists within some enterprise models, are also proposed. The Zachman Framework for Enterprise Architecture is also outlined against the existing architectures and the proposed enterprise framework. The main contribution of this paper is the identification and definition of a common knowledge enterprise model which represents an original combination between the previous projects on enterprise architectures and the Object Management Group (OMG) models and standards. The identified common knowledge enterprise model has therefore been designed using the OMG's Model-Driven Architecture (MDA) and Common Warehouse MetaModel (CWM), and it also follows the RM-ODP (ISO/OSI). It has been partially implemented in JavaTM, Enterprise JavaBeans (EJB) and Corba/IDL. Finally, the advantages and limitations of the proposed enterprise model are outlined.

1. Introduction

One of the foremost challenges facing manufacturing industry nowadays is the large-scale integration of their enterprise systems, along with their associated models, data, information, knowledge and web descriptions. To achieve a high level of integration of managerial and technical elements, companies are resorting to *standard reference architectures* and *common enterprise models* usually developed within international projects. On the other hand the enterprise systems generate large amounts of data which are a valuable asset and potentially an important source of new information and knowledge for improving the business of the enterprise, gaining competitive advantage in fierce markets and coping with changes and managerial complexity. Large organizations such as manufacturing companies could respond to changes and challenges in their business and production activities by using intensive and intelligent database processing to identify new trends and to predict and improve their business performance. Therefore *knowledge discovery (KD) in databases* is a promising solution and this paper is focused on using emerging knowledge technologies such as data mining for enterprise integration and modeling. However, a comprehensive and comparative review of literature related to knowledge discovery, data mining and enterprise integration and modeling approaches reveals that there is not any reported interdisciplinary research. This paper suggests that extending the existing enterprise modeling and integration architectures and environments to incorporate *KD* and *data mining (DM)* systems could significantly contribute to improving the decision making process and business performance. Therefore this paper mainly addresses the design and development of a framework for

enterprise engineering which consider knowledge discovery in databases and data mining processing to be essential. One of the most important practical issues of the research reported in this paper is to include and accommodate a generic KD&DM system within the existing standardized and referenced enterprise architectures and models. In order to realize this issue in a systematic, effective and standard manner, the paper proposes introducing new views within the existing reference architectures for enterprise modeling and integration. Finally, a common knowledge enterprise model applied to extended enterprise is designed and partially developed.

The most well-known enterprise architectures in the 1990s include *CIM-OSA*, *ARIS*, *PERA* and *GERAM* (ESPRIT 1993, Vernadat 1996, Rolstadas and Anderson 2000). However, research efforts have been especially devoted to looking for new architectures in order to accommodate the changing conditions, recent technical advances and requirements of new manufacturing paradigms such as agile, holonic, bionic, fractal and virtual. The research presented in this paper also considers the *Reference Model of Open Distributed Processing* (RM-ODP) which is an ISO coordinating framework to support the design of distributed systems in heterogeneous environments based on the standardization of open distributed processing (ISO/OSI 1995).

The paper briefly presents in section 2, the background of knowledge discovery, data mining, related systems, projects, standards and modeling techniques. This paper is directed at defining, especially in section 3, an enterprise modeling and integration framework which intensively considers the previous enterprise engineering approaches and is based on knowledge discovery and data mining defined as emerging knowledge technologies. Generally, a framework is defined as a more general concept than an architecture including incomplete and general design and implementation roadmap or guidelines for a range of enterprise information systems. Hence, different architectures can be developed within a framework (Molina 1995, Vernadat 1996, Molina and Bell 2002, Neaga 2003). Within the framework described in this paper, a *common knowledge enterprise model* is identified and it is analyzed and modeled in section 4. This model is applied to the extended enterprise as shown in sub-section 4.3. The remainder of the paper includes in section 5, the development aspects, and in section 6 the advantages and the limits of the Common Knowledge Enterprise Model are presented.

2. Background of Knowledge Discovery and Data Mining

Knowledge discovery(KD) and *data mining(DM)* are interdisciplinary areas directed at intelligently exploring large databases in order to find and use patterns, new information and knowledge. KD and DM incorporate complex algorithms from statistics and artificial intelligence, including imaginative and intuitive processing. Like other evolutionary systems, especially those based on neural networks, DM applications are tending to use both rational and emotional intelligence defined as affective intelligence (Adami 1998, Neaga 2003).

The main DM techniques are On-Line Analytical Processing(OLAP) and methods based on classification, association rules, clustering, decision trees, sequential patterns, fuzzy logic and combinations of algorithms such as neural networks(NN) and case-base

reasoning(CBR) (Fayyad et al. 1996, Ebecken 1998, Bramer 1999, Adamo 2001, Han and Kamber 2001, Klosgen and Zytchow 2002). It is also possible to consider DM more as a set of organized activities than as methods on their own because the main algorithms are employed from close areas such as statistics and/or artificial intelligence.

Manufacturing enterprises rely on vast amounts of data and information that is located in large databases. This information is a valuable resource, but its value can be increased if additional knowledge can be gained from it. The exploration of database information, to identify and extract deep and hidden knowledge, is made possible by DM techniques (Ebecken 1998). The existing databases of manufacturing enterprises, or indeed of most large organizations, are huge, but largely untapped sources of information, since they contain valuable records of operational and market history. DM techniques can be used to improve strategic and operational planning activities, as databases can be explored to gain feedback on the past performance and business behaviour of the enterprise as shown in Neaga and Harding (2002) and Neaga (2003). Data warehouse is an optional stage in order to perform data mining and it efficiently and effectively collects and stores data from multiple, distributed and heterogeneous operational databases, organizes the data using the data mart concept, and keeps historical data for future analysis (Adriaans and Zantinge 1996, Harding and Yu 1999)

2.1 Overview of Data Mining Systems, Projects and Standards

Nowadays the trends in DM are towards standardization of projects, using common methods and tools, and defining repeatable activities. *CRISP-DMTM* (Cross Industry Standard Process for Data Mining), *SolEuNet* (Data Mining and Decision Support for Business Competitiveness: A European Virtual Enterprise), *Kensington Enterprise Data Mining* (Imperial College, Department of Computing, London) and other projects have established methodologies and developed dedicated languages and software tools for KD and DM processing (SoleUNet 1999, Chapman et al. 1999, 2000, Helberg 2002). *Predictive Model Markup Language*(PMML) is a *Data Mining Group*(DMG) open standard format based on XML specification for exchanging mining models between applications running on different platforms. *Common Warehouse Metamodel* (CWM) is an OMG's UML/XML-based specification for mining models defined as metadata (OMG 2000).

However, these projects, dedicated languages and standards are not correlated with the enterprise modeling and integration projects and related standards for distributed manufacturing. Moreover, most DM products focus on the mining technology rather than on the ease-of-use, integration, scalability and portability. Neaga (2002, 2003) and Harding (2002) have recommended for manufacturing engineering applications, the use of PolyAnalystTM system as well as the data mining source libraries such as Weka and ArMiner.

3. Enterprise Knowledge Engineering Approach

The enterprise engineering approaches define generic modalities of modeling, analysis and design for several enterprise systems including information, manufacturing and

business systems. These approaches usually include methodologies related to building and executing the models of an integrated and/or extended enterprise and related IT support systems. The reference architectures are useful for abstracting different views of the enterprise, specifying modelling approaches to emphasize enterprise properties and defining the life-cycle of the enterprise engineering activities. However, the reference architectures do not include how integration, collaboration and communication within an enterprise and/or extended enterprise are realized especially in the context of sharing data, information, knowledge and web models. For example, a model may show that an information flow between two functions is desirable for integrating a business process but it does not specify how to implement the integrated information flow (Giachetti, 2004).

There are several enterprise modeling approaches and reference architectures such as **CIM-OSA**, **ARIS**, **PERA**, **GRAI**, **TOVE** and **GERAM** which constitute a common environment and methodology for complex system integration, comparative analysis, modeling design and re-design of a manufacturing enterprise using advanced software tools (ESPRIT 1993, Bernus et al. 1996, Fox 1996, Vernadat 1996, Edwards et al. 1998, Aguiar and Edwards 1999, Harding et al. 1999, Toh 1999a, b, Rolstadas and Anderson 2000, Wang et al. 2002, Shen et al. 2004). Consideration of these approaches, especially CIM-OSA, is justified in order to define a new enterprise model, because these reference architectures may already be used by some companies that may become part of an extended enterprise. Although current advanced and intelligent data exploration was not originally included within these early projects, they do create the basis to build a common enterprise model based on knowledge discovery and data mining.

It can be argued that a reference architecture such as **Computer Integrated Manufacturing – Open System Architecture (CIM-OSA)** which is comprised of several main views, for example **function**, **information**, **resource** and **organization views** have been found to be sufficient for the building and execution of several particular enterprise models (Toh 1999a, b, Wang et al. 2002). The **Architecture of Integrated Information Systems (ARIS)** contains **organization**, **function**, **data** and **control** views (Scheer and Kruse 1994). Generally, the existing reference architectures do not include specific views in order to support advanced knowledge processing even though the fact that knowledge has different representations is considered. Aguiar and Edwards (1999) described knowledge capture for enterprise model building and knowledge manipulation for enterprise model execution. However, the associated **Systems Engineering Workbench for Open Systems Architecture, SEW-OSA** (ESPRIT 1993, Edwards et al. 1998, Aguiar and Edwards 1999) has not included advanced knowledge processing elements and/or sub-systems.

Giachetti (2004) presents a detailed framework to review the information integration of the enterprise. He stated that the various information integration issues and supported systems and their relationships to each other have not been sufficiently investigated and defined. His article presents an enterprise information integration framework that aims to bridge parallel approaches towards integration so that the information integration requirements can be better globally understood and generic representation used. This framework has included ontology definitions and different knowledge exchange and communication languages. The data warehouse is considered only for data integration

aspects because it provides a global data view for the purposes of analysis. However within the information integration framework presented by Giachetti (2004) the data view is not exploited further to gain information and knowledge as shown in the approach presented here, and this is especially related to the addition of new views to ARIS. The new information as well as the discovered knowledge constitutes a support for achieving a high level of enterprise integration, and extended enterprise communication and collaboration based on both reference architecture and knowledge modeling.

Knowledge discovery and its advanced processing and maintenance should bring a considerable advantage for improving business performance and the flexibility of future manufacturing system design and re-engineering, and new product development and introduction. These issues have been approached and been demonstrated within several research projects related to information and knowledge modeling (Molina 1995, Harding 1996, Zhao et al. 1999, Dorador and Young 2000, Molina and Bell 2002). Wang et al. (2002) have introduced a few considerations about data sources integration and data warehouse, and Williams (1996) has defined the requirements for data shared between enterprise entities for PERA reference architecture.

Insert figure 1 about here

CIM-OSA includes function, information, resource and organization views which are described in details in (ESPRIT 1993, Bernus et al. 1996, Vernadat 1996, Toh 1999a, b, Wang et al. 2002). The number of views is not limited, and it can be expanded as necessary, but it is recommended that the number be kept to the minimum possible. The existing views enable designers and users to better understand and communicate the structure, purposes, capabilities, resources and relationships within the enterprise and a network of enterprises. However, they do not particularly support them to identify and extract knowledge that exists within the system mainly because they do not capture knowledge in a systematic and organized manner. Hence, different views of the enterprise, and of the extended enterprise are needed to enable efficient knowledge extraction through data mining. Therefore it is necessary to define the **knowledge view** as the description of processed information with an associated meaning, which leads to an action that adds value to the initial data.

The extended CIM-OSA cube depicted in figure 1 defines the high level representation of an enterprise engineering modeling and integration framework based on knowledge discovery and data mining. The knowledge view facilitates enterprise modeling, integration, collaboration and coordination from the knowledge perspective.

Also if **enterprise modelling** is considered as the process of building models of the whole or part of the enterprise such as process models, data models, resource models etc. based on knowledge about the enterprise, previous models and/or reference models as well as domain ontologies and model representation language (Vernadat 1996), then **mining models** are directed to logically fit or overlap with **enterprise models**, except that they are obtained by knowledge discovery. Descriptive enterprise models are widely used and usually are based on diagrams described in an **Integrated Definition Language**

(IDEF) based on the *Structured Analysis and Design Technique* (SADT) (Toh 1999b, Toh and Harding 1999, Dorador and Young 2000, Aguilar-Savén 2004, Shen et al. 2004). Generally these models describe the business processes across the enterprise and the related modeling and integration framework enables a common understanding and analysis of the business activities. These models can be complemented and improved with *mining descriptive models* describing patterns in existing data about an enterprise's behaviour and past performances. The *mining predictive models* could be used to forecast enterprise model evolution and its future business behaviour as position on the market. These models may evaluate the initial enterprise model, its evolution during model execution and its achievement and forecast future business. Enterprise knowledge could be classified as follows:

- Knowledge about the past which is stable, voluminous and accurate;
- Knowledge about present which is unstable, compact and may be inaccurate;
- Knowledge about the future which is hypothetical.

Knowledge discovery and data mining are critical processes applied to existing enterprise databases in order to find new information, knowledge and patterns which show the future enterprise behaviour and improve its business performance. Besides because the mining models are based on enterprise application/systems they may cover a generic description for the business as well as production processes and product representations. Furthermore the knowledge view supports other areas of knowledge engineering and management, and it makes possible the distinction between tacit (implicit) and explicit knowledge (Polanyi 1966):

- *Tacit knowledge*: implicit, mental models and experiences of individuals.
- *Explicit knowledge*: formal models, rules and procedures.

The conventional model to turn data into information and further into knowledge is defined as follows: *data* =====> *information* =====> *knowledge* =====> *wisdom*

However one of the main goals of data mining is to gain new information and knowledge from databases. Therefore data mining may process the information embedded in the information view and the new information, knowledge and patterns could be captured in the same views.

In order to obtain a distinct separation, an additional view of the enterprise is suggested and this is called the mining view. The *mining view* should include the description of mining and new business, manufacturing and product models obtained through knowledge discovery, OLAP, data mining and other advanced data exploratory techniques.

The *data view* represents the abstract level of all data manipulated by different software systems into an enterprise information system. Therefore a data view has to be considered at the level of enterprise architecture and the definition of a high level enterprise data architecture should significantly support data mining. The data view exists as part of

ARIS and this has been used as a basis for the inclusion of additional views such as information, knowledge and mining views as depicted in figure 2.

Insert figure 2 about here

Generally, adding new views or extending the interpretation of existing views provides support for the development of applications based on knowledge and business intelligence solutions. However CIM-OSA, PERA, ARIS, GERAM etc. are high level and abstract reference architectures which do not offer implementation solutions especially for specific IT systems such as DM applications which process particular data stored in operational databases. It is therefore suggested that the implementation solutions should be based on an *application view* obtained through a unified perspective of an enterprise's applications such as supply chain, enterprise resource planning, and customer relationship management systems and support advanced software tools such as SAP, I2 Technologies, Manugistics etc. (Neaga 2003). The identified enterprise model is based on the enterprise's applications and their combination in order to demonstrate the application of KD&DM in several manufacturing and business areas. Moreover the OMG's CWM, which has been employed to design the common knowledge enterprise model, provides a clear manner to be linked with a generic representation of enterprise applications.

3.1. Outline of Zachman Framework for Enterprise Architecture

Zachman (1996, 1999) who is an internationally recognized expert on Enterprise Architecture and author of a Framework for Enterprise Architecture, defines a strong and logical connection between business processes, organization strategies and enterprise architectures. This approach also emphasizes that an enterprise must produce models in order to deliver systems implementations in the short-term and at the same time for the long-term, instantiate the architecture process in order to ensure on-going coherence of system implementations and to build an enterprise environment conducive to accommodating high rates of change. The Zachman Framework could also be defined as a conceptual methodology which shows how all of the specific architectures that an organization might define can be integrated into a comprehensive and coherent environment for enterprise systems. It is an analytic model or classification scheme that organizes descriptive representations. It does not describe an implementation process and is independent of specific guidelines (Frankel et al. 2003).

In summary, this framework has the following characteristics (Zachmann 1996):

- a. **Simplicity:** it is easy to understand and it is not technical, but purely logical. In its most elementary form, it has three perspectives: Owner, Designer, Builder, and three abstractions: Material, Function, Geometry. Anybody (technical or non-technical) can understand it.
- b. **Comprehension:** it addresses the enterprise as a whole. Any issues can be mapped against it to understand where they fit within the context of the enterprise as a whole.
- c. **Language supporting:** it helps to think about complex concepts and communication of them precisely with few and non-technical terms.

- d. Planning Tool: it helps to make better choices about the enterprise planning and its objectives. It is possible to find the best alternative in the context of a complex business with a range of alternatives.
- e. Problem-Solving: it enables working with abstractions, to simplify and to isolate simple variables without losing sense of the complexity of the enterprise as a whole.
- f. Neutrality: it is defined totally independently of tools or methodologies and therefore any tool or any methodology can be mapped within the framework.

Generally, compared to the Zachman framework any academic research project in the area of enterprise engineering is more complicated and in consequence there may be many limitations or restrictions in applying it for a specific enterprise. Furthermore the framework which also considers knowledge discovery and data mining becomes too restrictive and difficult to apply in practice, due to the many systems that are involved. The Zachman framework is very general and can oversimplify some enterprise issues such as its business performance and behaviour although it takes in consideration decision support systems, analytical processing and data exploration. However, the corresponding systems cannot be integrated industry-wide like the enterprise applications supported by the framework proposed within this paper. Generally, the expense of building a data warehouse in a given enterprise is substantial and may not be directed to a quick return on investment. Frankel et al. (2003) have elaborated a mapping methodology between the Zachman framework and Model-Driven Architecture (MDA) developed by OMG. The Common Warehouse MetaModel (CWM) defined at the heart of MDA is used to develop the common knowledge enterprise model as it is described in the next sections.

4. The Common Knowledge Enterprise Model

4.1 Introduction of Model-Driven Architecture and Common Warehouse MetaModel

Model-Driven Architecture (MDA) is the latest OMG initiative, developed in order to support enterprises and organizations to integrate new applications into existing systems. MDA is a middleware and it acts as a high-level abstract architecture based on UML methodology and existing profiles. At the heart of MDA is the already defined Meta-Object Facility(MOF), CORBA, XMI/XML and Common Warehouse Metamodel (CWM) (OMG 2001).

Common Warehouse MetaModel (CWM) defines a *generic model* that enables data exchange and sharing across databases or even data warehouses across enterprises (OMG 2000, Poole et al. 2002, 2003). It is a new open industry standard recently adopted by some companies such as Oracle, SAS and others which are progressing towards incorporating this standard in their implementation. CWM is also a *common metamodel* which should be independent of any specific data warehouse implementation, but which becomes domain specific in association with domain specifications such the common knowledge enterprise model presented in this paper. The metamodel is developing as a set of packages which describes metadata.

Metadata is defined as data describing data or information about data, and generally, it comprises a description of information structures and models (Poole et al. 2002, 2003). Using the CWM supported by MDA, an identified common knowledge enterprise model has been modeled and designed. This phase has been done using UML implemented in Rational Rose Enterprise Edition (Rational Co. 2000)

4.2 Enterprise Model Description

Insert figure 3 about here

The enterprise modeling and integration framework based on knowledge discovery and data mining which is shown in figure 3 provides the design, development and implementation of a *common knowledge enterprise model*. Figure 3 illustrates that the enterprise framework defines a unified environment for integration of knowledge discovery and data mining systems such as PolyAnalystTM, Clementine etc. and source libraries of programmes such as Weka, ArMiner etc. with enterprise systems such as SAP, I2 Technologies, PeopleSoft, JD Edwards etc.

The framework conforms to existing reference architectures for enterprise modeling and integration and it is based on OMG middleware definitions such as Common Warehouse MetaModel(CWM) and Model-Driven Architecture(MDA). The identified common knowledge enterprise model intensively exploits previous projects, models, standards and methodologies in both areas of enterprise engineering and data mining. The identified common knowledge enterprise model has been designed using the OMG's CWM described in UML and the structure of the main model is shown in figure 4. The model depicted in figure 4 is a combination of a main model called CWM-DM (OMG 2000) and the enterprise information models produced by the enterprise reference architectures and models previously described in section 3. Therefore an identified common knowledge enterprise model consists of a *MiningModel* having at the heart *MiningSettings*, *ApplicationInputSpecification*, which specifies the enterprise applications and systems. *MiningModelResult* represents the generated model that is output for the mining activities.

The class *SupervisedMiningModel* extends *MiningModel* to include supervised learning such as classification and regression. Hence, this class requires a *TargetAttribute* which provides the correspondance between *ApplicationAttribute* and the obtained *SupervisedMiningModel*. Supervised learning is applied to a data set which describes an Application identified by *ApplicationInputSpecification* and *ApplicationAttribute*, to obtain this model.

Insert figure 4 about here

The attribute *function* of *MiningModel* describes the DM function class (i.e. *AssociationRules*) whereas the attribute *algorithms* are used to specify the concrete algorithm (i.e. *decisionTree*). This model also contains classes corresponding to the

employed algorithms and learning techniques such as (OMG 2000, Poole et al. 2002, 2003):

- *StatisticsMiningModel*: statistical models;
- *AssociationRulesMiningModel*: association rule models;
- *SequentialMiningModel*: sequential analysis models;
- *SupervisedMiningModel* for supervised learning models;
- *ClusteringMiningModel*: clustering models.

EE_ApplicationSpecification Class describes the classes representing enterprise applications or combinations of these. This part of the model is not part of OMG's CWM but the link at the level of application modeling provides a high level of integration of different applications. This class incorporates *Customer Relationship Management (CRM) Class*, *SupplyChain (SCM) Class*, *Enterprise Resource Planning (ERP) Class*, product, manufacturing and marketing models usually described using a *Product Data Management (PDM)* system. *EE_ApplicationSpecification* class has also the role of eliminating the redundant information which could appear in the classes describing the enterprise systems and applications. Using a one to one multiplicity link the class *ApplicationInputSpecification* defines the set of input attributes for the mining model that are further used for enterprise modeling and integration and/or inter-enterprise communication in an extended enterprise.

The common knowledge enterprise model supports the development of standard collaborative information systems which are also platform-independent. Furthermore this paper demonstrates that if an information system follows reference architectures and models and also uses OMG standards such as UML, CORBA and MDA then standard information systems can be incorporated in its design, and can be flexibly integrated within the enterprise reference architectures. This model also allows standard knowledge discovery and data mining applications to be developed which adhere to RM-ODP framework as well as to DM standards developed by OMG and Data Mining Group (Neaga 2003).

4.3 Application of the model for extended enterprise

Extended enterprise is defined as a long-term co-operation and partnership based on information and knowledge exchange (Szegheo 1999) and the co-ordination of the manufacturing activities of collaborating independent enterprises and related suppliers (Jagdev and Browne 1998). The extended enterprise intensively uses communication and collaboration between manufacturing enterprises, and aims to achieve competitive advantage (Vernadat 1996, Harding et al. 1999, Szegheo 1999). ***Virtual Enterprise*** is the temporary alliance between enterprise systems using Internet and intranet technologies (Szegheo 1999). Jagdev and Browne (1998) and Zhang and Browne (1999) have comprehensively and comparatively analysed the extended and virtual enterprise issues and their associated concepts and paradigms. The model suggested is based on business processes modeling and integration (Shen et al. 2004)

Global competition and distributed resources make it necessary for the extended enterprise to create and use a framework that enables the association of product development, supply chain management activities and manufacturing strategies. The application of the common knowledge model to the extended enterprise is directed at solving the above issue.

For the extended enterprise, a non-overlapping combination of enterprise systems and applications is assumed, such as a supply chain management(SCM) system, a customer relationship management(CRM) system and an enterprise resource planning (ERP) system which may also support the business processes. For example a particular company may focus on implementing an ERP system, whilst others concentrate on SCM, CRM and other enterprise systems. The non-overlapping combination of enterprise systems represents the logical integration of dissimilar and complex applications which run in the same environment as the extended enterprise. Generally, these applications support different processes such as business and production and the relation between suppliers and customers within the extended enterprise. However these systems may generate redundant and identical data which needs to be pre-processed or aggregated in a data warehouse. Also some companies, particularly those with technical products may use Product Data Management (PDM) applications in order to improve their supply chains because a collaborative PDM system has the capabilities not only to define product components but also to share this design information along the supply chain.

In the same manner as for a company running an individual business, an extended enterprise could consider the strategies which are of vital importance in order to achieve its business objectives (Platts and Gregory 1991, Storey 1994, Wang and Bell 1994). Also a competitive strategy involves using business analysis to maximize the value and capabilities that distinguish an organization from its competitors (Porter 1980, 1998). Shen et al. (2004) have suggested a framework based on business processes modeling and integration across companies and units which may be part of an extended enterprise. This approach provides data and information flow descriptions. However there is not any specification regarding the huge amounts of data generated by applications and accumulated by databases which could be further explored.

In order to formulate and include the strategies for an extended enterprise it is necessary to define its objectives which may include aspects of quality, delivery, cost, flexibility and innovation as short-term objectives as well as long-term objectives. Therefore the common knowledge enterprise model has been completed with *The ExtendedEnterpriseStrategies* and *The GenericProductData* classes.

The ExtendedEnterpriseStrategies class and its associated sub-classes which link strategic management and operations are presented in figure 5. These classes logically describe several concepts such as JIT (Just-in-time) and TQM (Total Quality Management), LP (Lean Production), MRPII (Manufacturing Resource Planning),

ERP (Enterprise Resource Planning), FMS (Flexible Manufacturing Systems) etc. which capture the changes in the global economy and competitive market places.

Insert figure 5 about here

The GenericProductData class and its sub-classes presented in figure 6 describe the following items (Harding 1996, Dorador and Young 2000, Wortmann et al. 2001):

- CAD geometrical product model including feature-oriented product description;
- STEP (Standard for the Exchange of Product Model Data) and Express product neutral model class which describes the neutral data format for the representation and exchange of product data;
- Product Data Management (PDM) representation.

5. Development Issues

From the IT perspective an integrating infrastructure is a set of common services and functions available as middleware to all applications on different platforms of a distributed system (Vernadat 1996). **Object Management Group**(OMG) has developed **Common Object Request Broker Architecture**(CORBA) and the associated language Interface Definition Language(IDL) (Orfali et al. 1997, Orfali and Harkey 1998), and most recently **Model-Driven Architecture** (MDA) and **Common Warehouse MetaModel** (CWM).

Insert figure 6 about here

The defined classes of a prototype system have been generated and transferred in **Oracle-JDeveloper** which is Oracle's Java development tool for building, debugging and deploying Internet applications. Java source generation is based on the component specification rather than on the class specification. In order to realize Java source code, after diagrams for classes have been created, every class is assigned a valid Java component. In practice the following methods have been used (Rational Co. 2000):

1. Generating Java source from Class Diagram;
2. Generating Java source from a Component Diagram.

JDeveloper V3.2. includes **Java Data Base Connectivity**(JDBC) capability for an Oracle database, and automatic generation of Corba interfaces. It also includes Oracle Business Components for Java, **EXtensible Markup Language**(XML), powered application component framework that significantly simplifies the development, deployment and customization of multi-tier Java applications for the Internet. Furthermore **Enterprise JavaBeans**(EJB) offer the possibility to develop reusable enterprise components which include knowledge discovery and data mining components.

Some Application Programming Interfaces(API) have been proposed and these have the aim of making dissimilar software systems run in the same environment.

In order to leverage legacy databases, existing applications and data mining systems, a set of CORBA/IDL has been generated. Neaga (2003) includes some details about the implementation and the development of object wrappers using CORBA/IDL.

6. Advantages and Limits of the Common Knowledge Enterprise Model

The main advantages of the enterprise integration and modeling framework and of the common knowledge enterprise model described in this paper are as follows:

- The inclusion within the high level enterprise referenced architectures and models of the knowledge and mining views which provide a good potential to perform intelligent data exploration such as data mining on enterprise databases. Knowledge Discovery in Databases and data mining could significantly contribute to the improvement of the business performance of an enterprise, and facilitate the re-engineering and re-design of manufacturing systems as well as new product introduction, design and manufacture.
- Definition of a unified object-oriented framework for manufacturing, product, mining and knowledge models and associated support systems based on OMG's MDA and CWM.
- The identification of a common knowledge enterprise model which is fully designed using a new standard developed by OMG for data warehouse and mining as the CWM.
- The identified enterprise model has been partially implemented in JavaTM, Enterprise JavaBeans (EJB) and Corba/IDL.
- The possibilities of mapping the identified enterprise model to standards and enterprise generic models obtained based on Zachman framework and RM-ODP.

However, the approach presented in this paper has the following limitations:

- The combination of several systems which individually satisfy particular requirements may not provide the best overall solution (Toh and Harding 1999). The common knowledge enterprise model supported by an enterprise framework based on knowledge discovery has not included any selection methodologies of software systems as suggested by Toh and Harding (1999).
- Enterprise Resource Planning (ERP) software is the dominant strategic platform for supporting enterprise-wide business processes. However, it has been criticised for being inflexible and not meeting specific organisation and industry requirements. An alternative approach, best of breed (BoB), integrates components of a standard package and/or custom software. The BoB integration approach describes the specifications in order to integrate software systems especially for ERP applications developed by different software vendors (Light, 2001). BoB is not an integration environment, rather, it is a strategy that could provide flexible integrated enterprise solutions that are complementary to the the enterprise modeling and integration framework suggested in this paper.
- Business processes modeling and automation using the workflow techniques and associated tools including workflow business process discovery and mining should have a very good potential (Aguilar-Savén 2004, Shen et al. 2004). However, workflow methods are not supported by this framework which is focused on data and its intelligent exploration mainly in order to gain knowledge.

- Several enterprises especially from North America already use advanced IT support for advanced planning and scheduling systems based on the Supply Chain Organization Reference(SCOR) model (Stadtler and Kilger 2000). SCOR is not explicitly considered within this approach, but a particular system may use a SCOR model which arises from a different modeling approach.
- The common knowledge enterprise model could generate redundant data, information and files even though this is thought to have been eliminated.
- Web and text mining have not been considered even though they are very important for communication, collaboration and co-ordination within an extended enterprise.

Additional investigations such as multi-tier architectures applied to enterprise information systems are needed to support this approach. These are especially important because of the above mentioned complexities. Web architectures and applications should be investigated and adopted, such as .NET architecture supported by MDA, and Web services which are standard modular applications that can be published, located, and invoked across the Web. The .NET architecture and Web services are the actual implementation alternatives for several enterprise applications, but the methodologies to integrate legacy systems and databases have not been fully identified.

7. Concluding remarks

The modeling and integration framework presented in this paper provides a dynamic environment for enterprises and extended enterprises especially to accommodate the following complex and dissimilar software systems:

- Applications to support Customer Relationship Management(CRM), Supply Chain Management (SCM) and Enterprise Resource Planning (ERP) applications. Examples of these systems are SAP R3, JD Edwards, I2 Technologies, Baan, Oracle Applications, PeopleSoft, Manugistics etc.
- Knowledge discovery and data mining products such as PolyAnalystTM, Clementine, Weka, ArMiner etc.
- Manufacturing and products models stored in databases such as Oracle, Object Store, and associated database management systems and applications.

The framework is especially directed at enabling knowledge discovery and data mining activities to be encompassed within an enterprise's existing standardized and referenced architectures and models.

References

- ADAMI, C., 1998, *Introduction to A-Life*, (Berlin:Springer-Verlag)
- ADAMO, J.M., 2001, *Data Mining for Association Rules and Sequential Patterns*, (Berlin:Springer-Verlag)
- ADRIAANS, P., ZANTINGE, D., 1996, *Data Mining*, (London:Addison-Wesley)

- AGUIAR, M.W.C. and EDWARDS, J.M, 1999, Achieving manufacturing business integration through the combined formalisms of CIMOSA and Petri nets. *International Journal of Production Research*, 36, 1767-1786.
- AGUILAR-SAVÉN, R.S., 2004, Business process modelling: Review and framework. *International Journal of Production Economics*, 90, 129–149.
- BERNUS, P., NEMES, L. and WILLIAMS, T.,J. (eds), 1996, *Architectures for Enterprise Integration*, (London:Chapman & Hall)
- BRAMER, M.A.(editor),1999, *Knowledge Discovery and Data Mining*, (New York: The Institute of Electrical Engineers)
- BROWNE, J. and ZHANG, J. 1999, Extended and virtual enterprises - similarities and differences. *International Journal of Agile Management Systems*, 1(1), 30-36
- CHAPMAN, P., CLINTON, J., KERBER, R., KHABAZA, T., REINARTZ, T., SHEARER, C. and WIRTH, R., 1999, 2000, *CRISP-DM 1.0 Step-by-step data mining guide*, (New York: SPSS Inc.)
- DORADOR, J.,M. and YOUNG, R.I.M., 2000, Application of IDEF0, IDEF3 and UML methodologies in the creation of information models. *International. Journal of Computer Integrated Manufacturing*, 3(5), 430-445.
- EBECKEN,N.F.F (editor),1998,*Data Mining*, (London:Witt Press/Computational Mechanics Publication)
- EDWARDS, J.M., AGUIAR, M. W.C. and COUTTS, I.A., 1998, A top down and bottom up approach to manufacturing enterprise engineering using the function view. *International Journal Computer Integrated Manufacturing*, 11(4), 364 – 376.
- ESPRIT Consortium AMICE, 1993, *CIMOSA: Open System Architecture for CIM*, 2nd revised and extended edition, (Berlin:Springer Verlag)
- FAYYAD, U.M., PIATETSKY-SHAPIRO, G., SMYTH, P. and UTHURUSAMY, R. (eds), 1996, *Advances in Knowledge Discovery and Data Mining*, (New York: AAAI Press/The MIT Press)
- FRANKEL, D.S., HARMON, P., MUKERJI, J., ODELL, J., OWEN, M., RIVITT, P., ROSEN, M. and SOLEY, R.M., 2003, The Zachman Framework and the OMG's Model Driven Architecture, White paper, *Business Process Trends*, 1-14.
- FOX, S.M., 1996, Research in Enterprise Integration. Proceedings Artificial Intelligence and Manufacturing Research Planning Workshop, USA, pp. 40-52.
- GIACHETTI, R.E., 2004, A framework to review the information integration of the enterprise. *International Journal of Production Research*, 42, 1147-1166.
- HAN, J. and KAMBER, M., 2001, *Data Mining: Concepts and Techniques*, (New York:Morgan Kaufman)
- HARDING, J.A., 1996, A Knowledge Model to Support Concurrent Engineering Team Working, Ph.D. thesis, Wolfson School of Mechanical and Manufacturing Engineering, Loughborough University, Loughborough, UK.
- HARDING, J.A. and YU, B., 1999, Information-centred enterprise design supported by a factory data model and data warehouse. *Computers in Industry*, 40, 23-36.
- HARDING, J.A., YU, B., and POPPLEWELL, K., 1999,Information modelling: an integration of views of a manufacturing enterprise. *International Journal of Production Research*, 37(12), 2777-2792.
- HELBERG, C., 2002, *Data Mining with Confidence*, 2nd edition, (New York:SPSS Inc.)

- ISO/OSI, 1995, ISO Basic Reference Model for Open Systems Interconnection (ISO/OSI). International Standards Organisation.
- JAGDEV, H.S. and BROWNE, J. 1998, The extended enterprise – A context for manufacturing. *Journal of Production Planning & Control*, 9(3), 216-229.
- KLOSGEN, W. and ZYTKOW, J.M.(editors), 2002, *Handbook of Data Mining and Knowledge Discovery*, (Oxford: Oxford University Press)
- LIGHT, B., 2001, ERP and best of breed: a comparative analysis, *Business Process Management Journal*, 7(3), 216-224.
- MOLINA, A., 1995, A Manufacturing model to support data-driven applications for design and manufacture”, PhD Thesis, Wolfson School of Mechanical and Manufacturing Engineering, Loughborough University, Loughborough, UK.
- MOLINA, A. and BELL, R., 2002, Reference models for the computer aided support of simultaneous engineering. *International Journal of Computer Integrated Manufacturing*, 15(3), 2002, 193–213.
- NEAGA E.I. and HARDING J.A., 2002, A Review of Data Mining Techniques and Software Systems to Improve Business Performance in Extended Manufacturing Enterprises. *International Journal of Advanced Manufacturing Systems (IJAMS) Special Issue on Decision Engineering*, 5(1), 3-19.
- NEAGA, E.I., 2003, Framework for distributed knowledge discovery systems embedded in extended enterprise, PhD thesis, Wolfson School of Mechanical and Manufacturing Engineering, Loughborough University, Loughborough, UK.
- OBJECT MANAGEMENT GROUP, 2000, Common Warehouse Metamodel (CWM) Specification, Proposal to the OMG ADTF RFP: Common Warehouse Metadata Interchange (CWMI).
- OBJECT MANAGEMENT GROUP, 2001, Model-Driven Architecture (MDA), Document 2001-07-01, Architecture Board ORMSC.
- ORFALI, R., HARKEY, D. and EDWARDS, J., 1997, *Instant CORBA*, (New York:John Wiley & Sons Inc.)
- ORFALI, R. and HARKEY, D., 1998, *Client/server programming with Java and CORBA*, (New York:John Wiley & Sons, Inc.)
- PLATTS, K.W. and GREGORY, M.J., 1991, Manufacturing audit in the process of strategy formulation. *International Journal of Operations and Production Management*, 10(9), 5-26.
- POLANYI, M., 1966, *The tacit dimension*, (London: Routledge and Kegan Paul)
- POOLE, J., CHANG, D., TOLBERT, D. and MELLOR, D., 2002, *Common Warehouse Metamodel An Introduction to the Standard for Data Warehouse Integration*”, (OMG Press, London: John Wiley & Sons, Inc.)
- POOLE, J., CHANG, D., TOLBERT, D. and MELLOR, D., 2003, *Common Warehouse Metamodel Developer Guide*, (OMG Press, London: John Wiley & Sons, Inc.)
- PORTER, M.E. 1980, 1998, *Competitive Strategy-Techniques for Analysing Industries and Competitors* (London:Free Press Edition)
- RATIONAL Co., 2000, *Using Rose – Rational Rose^R*, (New York:Rational Software Corporation)
- ROLSTADAS, A. and ANDERSON, B., 2000, *Enterprise Modeling – Improving Global Industrial Competitiveness*, (London:Kluwer Academic Publishers).

- SCHEER, A. W. and KRUSE C., 1994, ARIS-framework and toolset: a comprehensive business process reengineering methodology. Proceedings of 3rd Int'l. Conference on Automation, Robotics and Computer Vision (ICARV' 94) Singapore, Invited session on enterprise integration, pp. 327-331.
- SHEN, H., WALL, B., ZAREMBA, M, CHEN, Y. and BROWNE, J., 2004, Integration of business modelling methods for enterprise information system analysis and user requirements gathering. *Computers in Industry*, 54(3). 307–323
- SOLEUNET, 1999, Data Mining and Decision Support for Business Competitiveness: A European Virtual Enterprise”, EU 5TH Framework programme.
- STOREY, J. (editor), 1994, *New Wave Manufacturing Strategies, Organizational and Human Resource Management Dimensions*, (London:Chapman Publishing Ltd.)
- STADTLER, H. and KILGER, C., 2000, *Supply Chain Management and Advanced Planning – Concepts, Model, Software, and Case Studies*, (Berlin: Springer-Verlag).
- SZEGHEO, O.,1999, Investigating the Extended Enterprise and the Virtual Enterprise. Proceedings of International Conference on Industrial Engineering and Production Management (IEPM'99), Glasgow, pp. 477-485.
- TOH, K.T.K., 1999a, The realization of reference enterprise modeling architectures. *International Journal of Computer Integrated Manufacturing*, 12, 403–417.
- TOH, K.T.K., 1999b, Modelling architectures: a review of their application in structured methods for information systems specification. *International Journal of Production Research*, 37, 1439-1458.
- TOH, K.T.H and HARDING, J.A., 1999, An enterprise modelling CASE tool and data schema requirements for the selection of software support. *International Journal of Production Research*, 37, 4079-5104.
- VERNADAT, F.B., 1996, *Enterprise Modeling and Integration: Principles and Applications* (London: Chapman & Hall)
- WANG, C., ZHANG, Y., SONG, G., YIN, C. and CHU, C., 2002, An integration architecture for process manufacturing systems. *International Journal of Computer Integrated Manufacturing*, 15(5), 413–426.
- WANG, W. and BELL, R., 1994, Progress, Challenges and Issues: A literature Review on Manufacturing Strategy”, internal report, Loughborough University, Department of Manufacturing Engineering.
- WILLIAMS, T.J., 1996, The needs of the field of integration. In Bernus, P., Nemes, L., Williams, T.,J. (eds), *Architectures for Enterprise Integration* (eds), Chapman & Hall.
- WORTMANN, J.C., HEGGE, H.M.H. and GOOSSENAERTS, J.B.M., 2001, Understanding enterprise modelling from product modeling. *Production Planning & Control*, 12(3), 234-244.
- ZACHMAN, J., 1996, The Framework for enterprise architecture: Background, Description and Utility, Zachman International, <http://www.zifa.com>.
- ZACHMAN, J., 1999, Enterprise Architecture: The Issue of the Century. Zachman Institute For Advancement, <http://www.zifa.com>.
- ZHAO, J., CHEUNG, W.M. and YOUNG, R.I.M., 1999, A consistent manufacturing data model to support virtual enterprises. *International Journal of Agile Management Systems*, 1(3), 150-158.

