Aligning Legacy Information Systems to Business Processes

Panos Kardasis & Peri Loucopoulos
Department of Computation Department
U.M.I.S.T.
P.O. Box 88,
M60 1QD Manchester, UK
{kardasis | pl} @co.umist.ac.uk

Abstract

Recent years have experienced a growth in demand for re-engineering legacy information systems. The complexity of a development endeavour leading to migration of a legacy system stresses the need for a systematic supporting approach. We argue in this paper that such an approach should facilitate (a) documentation of dependencies between business processes and supporting IS in a way that changes in the business level are reflected on system specifications and (b) quantification of effects of business changes on the associated migration from legacy systems so that alternative solutions can be systematically evaluated and the optimal solution can be chosen. In this paper we put forward an approach to meeting these two objectives based on the confluence of two technologies: Enterprise Knowledge Modelling and Knowledge Discovery in Data. The approach is demonstrated using examples from a project involving a banking application.

1 Introduction

Over the past decade, continuous challenges have been made to traditional business practices. Many institutions, companies and virtually all industries have been forced into reactive patterns of change in order to remain competitive. This effort has witnessed the disabling effects that the build-up of legacy systems has on such change. Immense size and criticality in the overall business operation, use of inappropriate and obsolete hardware, poor database services, lack of interface among applications and presence of unstructured and undocumented patches are only some of the typical characteristics of legacy systems [Brodie and Stonebraker 1996]. As an effect, migration from legacy environments is certainly not a trivial process, while it may become extremely expensive and time consuming.

Projects aiming at the replacement of legacy Information Systems (IS) by totally new ones tend to fail for a number of reasons:

- Specifications for the legacy systems rarely exist, relevant documentation is out-of-date or lost, and thus, the only source of information is the system code.
- Too much effort is spent on the analysis phase, which may never end, either because of the complexity of the system, or because of the ineffectiveness of the analysis approach.
- The replacement of the IS raises the need for business changes, which are rarely welcome. The fear of change in working practices, techniques and enabling

technologies contribute to enormous resistance. The large migration projects become even larger as everybody in the company is eventually found to be either involved or affected. This makes the situation uncontrollable and the project vulnerable to termination.

The above observations stress the need for a systematic approach to assist system migration. Such an approach should support:

- Understanding of the enterprise in terms of its operations and resources, in order to provide a solid background for analysing the system, and for assisting the co-ordination of business changes dictated by the migration project.
- Documentation of dependencies between business processes and supporting IS in a way that changes in the business level are reflected on system specifications, and vice versa.
- Quantification of effects of business changes on the associated migration from legacy systems so that system migration strategies can be systematically evaluated and the optimal solution can be chosen.

In this paper we put forward an approach to meeting these objectives based on the confluence of two technologies: *Enterprise Knowledge Modelling* and *Knowledge Discovery in Data*. The term 'enterprise knowledge modelling' refers to the set of techniques for describing the structure and business processes of an enterprise, its missions and objectives together with the way that these objectives may be *operationalised* onto system components [Bubenko, Rolland, et al 1994; Loucopoulos 1994; Loucopoulos and Katsouli 1992; Loucopoulos and Kavakli 1995; Rolland and Grosz 1994]. The term ' knowledge discovery in data' refers to correlations between data variables, identification of rules, and classifications implicitly contained in large amounts of corporate data [Matheus, Chan, et al 1993; Yoon and Kerschberg 1993].

Enterprise knowledge modelling provides the basis for developing models of current business processes and objectives for change, including changes to the business goals and business rules. Knowledge Discovery in Data is used for investigating the behaviour of the legacy IS in terms of its operational data and the way that such data is presently being used within the chosen business processes; it can also be used for identifying behavioural patterns which may give rise to new business processes [Dhar and Tuzhilin 1993].

The approach advocated in this paper is underpinned by three key activities:

1. Modelling of enterprise objectives, rules, and processes – for describing both the AS-IS and the TO-BE situations.
2. Analysis of legacy IS – for discovering the actual behaviour of the system against a set of defined business criteria.

3. Matching business knowledge models to results from analysis of legacy IS – for identifying the necessary changes to the IS (primarily) but also potential changes to the business processes themselves.

These three activities represent the backbone of the paper. The discussion is also based on practical grounds by considering an industrial application within the banking domain. Following a short introduction of the business application and of the background modelling approach (section 2), the paper discusses the modelling activity (section 3), the knowledge discovery activity (section 4) and the integration activity (section 5), using examples from the banking application in order to demonstrate the issues being discussed. Finally, the paper concludes with a number of observations (section 6).

2 The Banking Application

The examples presented in this paper are part of a large project aiming at enabling customer profiling in the banking sector. The part of the application in which this paper is confined is one particular business area namely the *marketing* of loans, hire purchase agreements, and preference accounts through a dense network of local sales representatives and external agents. The critical business functions deal with:

- Targeting of the customer base and contacting customer target groups through massive mailings
- Contacting individual customers in order to make offers and agreements
- Evaluating customer applications by considering both the profitability and the risk of the proposed agreements

These main functions are supported currently by a legacy IS, an abstract view of which is shown in the diagram of Figure 1.

The first process is supported by the 'TCS' system (Figure 1) which alerts sales representatives about customers that need to be called each time. Offers to existing customers are generally aligned to previous interaction between the customer and the bank, the history of which is provided accordingly by the system. However, the decision about the type of product to be offered and the agreement terms are left to the sales representative himself. The second process is supported by the component shown in Figure 1 as 'Application Systems'. This component is composed of many application programs that collectively deal with the scoring of customer requests according to customers' personal details, behaviour from previous transactions with the bank and references from external agencies. Finally, the 'Marketing Database' brings together the contents of all other database systems in order to facilitate decision making for massive mailing campaigns.

The majority of the systems supporting the current situation were developed and put to work within the past three decades or were occasionally appended to the existing systems in an undocumented, ad-hoc manner.

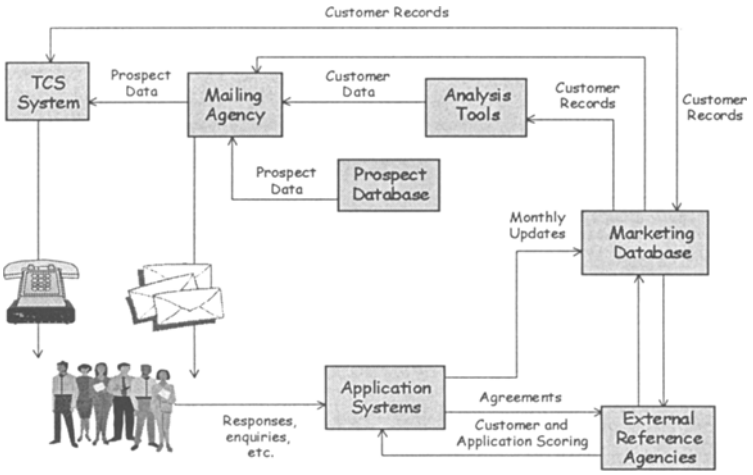

Figure 1: The current situation

The legacy infrastructure of the organisation consists of several databases, interfaces for inserting customer details and performing transactions of various types and applications for processing data and for accessing external sources of information. The heterogeneity of these legacy systems is a source of major problems including tremendous complexity, lack of sufficient documentation, co-existence of different functions serving the same purpose within the same environment, duplication of data and also confusion about which of these data are correct.

Several attempts for analysing the business in the past had revealed that the existing infrastructure constrained the effectiveness of business processes. However, drastic measures could not be taken. This is a well-known problem also met by many other large enterprises handling voluminous data records and bearing strong dependencies with their information systems [Brodie and Stonebraker 1996]. Typically, the full replacement of the legacy IS would probably raise the need for parallel changes in business processes; and in addition the data conversion for supporting the new processes might be infeasible within the time limits that the business can support being without its mission-critical IS.

The management of the particular organisation opted for the integration of enterprise knowledge modelling and knowledge discovery techniques. The main business objectives with respect to any business process improvement was to ensure retention of existing customers, efficiency in approaching new customers, maximum profitability of agreements and minimum contract risk through better customer understanding.

The modelling approach that was used in this business application was based on the Enterprise Knowledge Development (EKD) framework. The EKD approach brings together business process modelling, and business goals within a rationale framework.

Recent developments of EKD are reported in [Loucopoulos, Kavakli, et al 1997; Kavakli and Loucopoulos1998].

The EKD conceptual models are organised according to four viewpoints:

- The *enterprise goal* sub-model expresses the concepts involved in describing enterprise objectives, enterprise issues, and their causal relationships.
- The *enterprise actor role* sub-model expresses the concepts involved in modelling business processes in terms of the actors, their roles, interactions and activities in order for a business process to meet its objectives.
- The *enterprise object* sub-model expresses the concepts involved in modelling physical and informational objects for the functioning of business processes and activities of individual roles.
- The *enterprise rules* sub-model expresses the concepts involved in modelling the policy, and constraints, laid down by the enterprise in the way that processes, roles and resources may behave in the enterprise.

3 Enterprise Modelling for the Banking Application

It is often argued that enterprise knowledge modelling constitutes a 'top-down' conceptual modelling activity whereby domain experts articulate their views about existing and future business situations and requirements. This was indeed the case for the banking application, as banking personnel participated in the process of defining and validating the models for the main business functions introduced in section 2.

In addition, there was a significant amount of 'bottom up' modelling since much of the deliberation constituted the analysis of system functionalities. The study of the bank's support systems facilitated the understanding of business functions, regarding their purpose, and the way employees interact with each other, with customers and with the systems themselves. Much information was derived from examining existing IDEF0 documentation of system processes, interface screens and database contents. Nevertheless, it is our premise that this 'bottom-up' view very seldom is accurate since the actual functionality of the system is rarely that which is documented in specifications derived perhaps many years earlier. On the contrary, the knowledge discovery activity, which looks at the behaviour of the systems data, provides a safer understanding of the systems functionality (more about this in section 4).

The enterprise knowledge modelling activity for the banking application considered a number of interrelated components, each corresponding to the EKD sub-models. These were:

What is the purpose of each process?
Who are the actors and what roles do they play within each process?
What are the business objects required by the business system?
What are the business rules that dictate the behaviour of the business processes?

4.1 Defining business goals

Regarding the first question, the business goals that guide the functioning of the company were defined as shown in Table 1.

Table 1

Business Goal		Business Process
"Identify potential customers"	*a*	Targeting Customers
"Sell products to potential customers"	*b*	Contacting Customers
"Ensure maximum profitability" "Ensure low risk"	*c*	Evaluating Customer Applications
"Fulfil agreement obligations"	*d*	Conducting Pay-outs

The legacy information systems (shown in Figure 1) are support systems for the corresponding business processes presented in the previous table. A set of goals for change summarised in the goal graph of Figure 2 reflects the need for improvements both in business processes and their support systems.

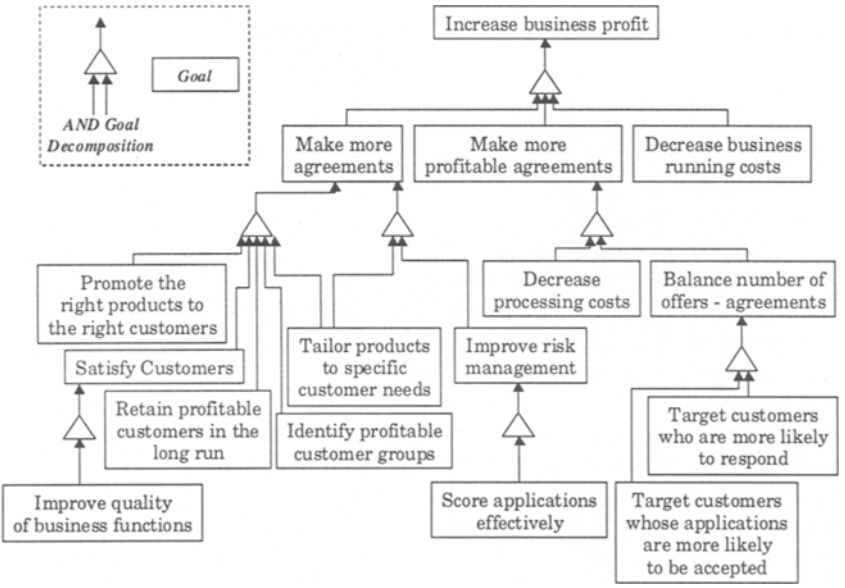

Figure 2: The bank's goals for change

4.2 Defining business roles

Given that the overall aim of the project was to achieve customer profiling for understanding customers better, and for customising products and services to their needs and potentials, we concentrated on processes *a*, *b* and *c* of Table 1 for which we identified the involved actors (Table 2):

Table 2

Business Process	Actors
Targeting Customers	Marketing manager, Database developer
Contacting Customers	Customer, Sales representative, Telemarketing support system (TCS)
Evaluating Customer Applications	Sales representative, Underwriter, External reference agency, TCS (Telemarketing facility for Central financial Services), AUF (Automatic Underwriting Facility)

Details about the roles that the aforementioned actors play in order to achieve the bank's business goals were derived from the study of the IDEF0 models. The elaboration of all the available information resulted in a number of Role-Activity diagrams [Ould 1995]. The portion presented here, reference the roles of process "Underwriting Customer Applications" which are depicted in a RAD-like notation:

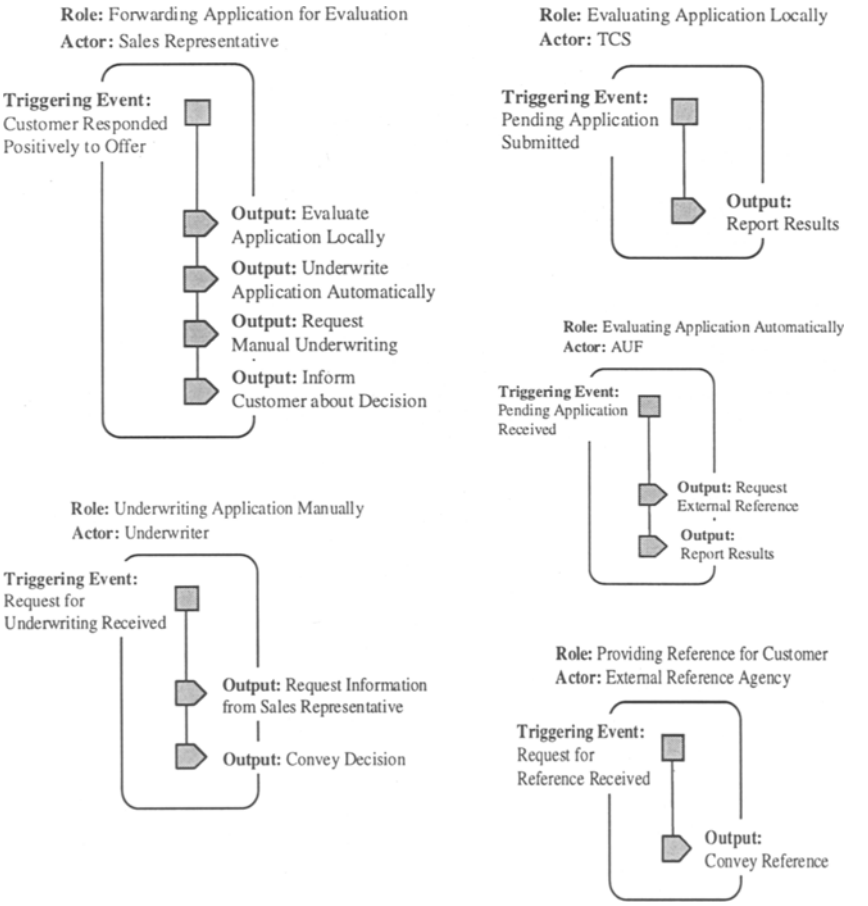

Figure 3: The role-activity diagrams

4.3 Defining business objects

The bank's business objects were mainly derived from various database tables summarised in Table 3.

Table 3

Table	Contents
Customer	Personal Data, Geographic and Demographic Details, Employment Information, Household Information, Financial Status.
Product	Product Description and Value, Repayment Method, Reason of Finance.
Contact Details	Details of Phone Calls and Mailings, Details about Customer's Response.

After identifying the principal business objects and their interrelations, we generated the business object model (Figure 4) referring to current enterprise informational entities. The latter have been associated logically, although they are derived from flat database tables, where all the attributes were interrelated through a common database key.

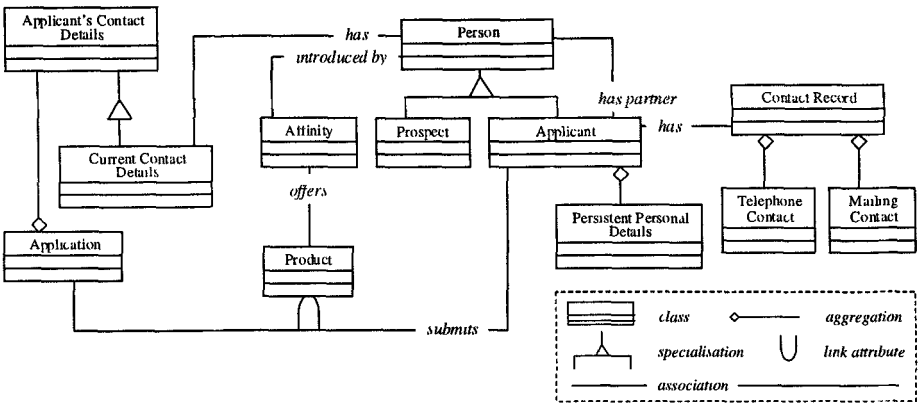

Figure 4: The main components of the business object model

4.4 Defining business rules

The dynamic aspect of business processes was demonstrated in terms of *role interaction* business rules, which are statements that control the co-ordination of activities within business roles. Activity co-ordination rules were represented as "WHEN... IF... THEN..." statements. The "WHEN..." part contains triggering events and controls, also referenced by one or more role-activity diagrams. The "IF..." part contains the preconditions for the invocation of an activity. Preconditions are logical expressions between object controls, i.e. comparison of object attributes to certain values. Finally, "THEN..." contains the activities to be initiated when the triggering event occurs and if all the preconditions are fulfilled.

All the identified role-interaction rules for the specific application fall into the following three categories:

Non-persistent guidelines of strategic planning and marketing decisions. The term "non-persistent" implies that decisions regarding the customer groups to be contacted and the types of products to be offered are regularly revised according to market tendencies and customers' demand.

Business rules for determining which individual customers need to be contacted and when. These rules are built into the telemarketing support system. The process of contacting customers by phone is also supported by rules regarding the types of products and services that can be offered each time. Other business rules deal with the specification of agreement terms (interest rates, duration of repayment, etc.). Given that sales representatives offer services on behalf of other financial corporations, they are also bound by different rules dealing with the cross-selling of products.

'Decline' and 'pend' rules applicable in underwriting customer applications. When customers do not fulfil the conditions of the 'decline' rule set, their applications are rejected immediately. When, the conditions of the 'pend' rule set are satisfied, the application is forwarded to the underwriter for manual evaluation. The underwriter also considers a number of implicit rules, in order to co-estimate the credibility of the customer, and his profitability for the bank.

Examples of modelling business rules according to the EKD textual and diagrammatic notations follow:

```
WHEN    Pending_Application_Submitted
IF      (Application.Behaviour_Score > y) AND
        Application.Behaviour_Score < z)
THEN    Request_Manual_Underwriting
```

4 Knowledge Discovery for the Banking Application

Knowledge Discovery in Data (KDD) has been defined as the non-trivial process of identifying valid, novel, potentially useful and ultimately understandable patterns in data. The KDD process usually involves data preparation, knowledge evaluation and refinement. The step of applying algorithms for extracting patterns from data is referred to as data mining. The existence of a data warehouse is a necessary "pre-processing" condition for developing knowledge discovery applications. Knowledge discovery usually starts with understanding the outcome variables, as well as the overall application universe (i.e. the data populated in the warehouse) [Tej Anand 1995]. In the case examined here, both steps were facilitated by the business object model presented in the previous section, which groups the available data attributes in an unambiguous and flexible way.

The mining exercise [Keane, Murray, et al 1997; Filippidou, Keane, et al 1998] was conducted upon a sample of the overall bank's data for loans and hire purchases. As it can be seen in Figure 4, the examined data constitute history of customer applications. Apart from current contact details (phone number and address) and history of contacts (dates of calls and mailings), all the other information describes the profile of the customer the time that he submitted his application for a certain product. Applications are related to customers':

Employment Details	Occupation, period with current employer, etc.
Household Information	Demographic information, type of occupancy, information about other family members, etc.
Financial Profile	Income, bank/building society, cards, etc.
Financial Status	Mortgages, credit cards and loans balance, standing expenses like rent, etc.

Other critical attributes like `Behaviour Score` for the customer and `Reason For Decline` for the application are directly assigned to `Application`. The `customer` is only related to his persistent personal details, be they his/her name and surname, sex and date of birth.

The outcome of data mining was a set of classification decision trees [ATTAR 1996], which represented the relation between one or more input data variables and a predefined output variable. The classification output variable was the `Reason For Decline`, reflecting whether an application has been accepted, rejected or just has not resulted in agreement, and the reasons why.

The information represented by the generated classification decision trees is equivalent to statements of the following structure:

```
α% of customers in (group χ for input variable κ) were accepted

α% of customers in (group χ for input variable κ and group ψ for
    input variable λ) were rejected (reason for decline ω)
```

From the generated results, influences of different attributes were determined and were used to inform the enterprise knowledge models and potentially to modify the systems and the business way of working.

5 Integration Between Enterprise Knowledge Models and Discovered Knowledge

The outcome of the activities described in sections 3 and 4 is:

(a) A set of specifications for the bank's business processes.

(b) Informal descriptions and models presenting the purpose, structure, functionality and constraints of their support IS.

(c) A number of statements reflecting the conclusions from the data mining exercise (customers classification and clustering of the customer base).

We claim that none of these views is sufficient for driving the transformation of the enterprise and/or the migration of their legacy IS. However, the integration of the gained knowledge (both *developed* and *discovered*) provides an overall picture of how the bank's business goals can be operationalised. Table 4 relates several suggestions for the improvement of the bank's legacy IS with the business goals that these improvements may fulfil.

Table 4

Goal 1: Promote the right products to the right customers	
	a. Automated support for selecting products to be offered to individual customers according to their personal profiles
	b. Association of customer personal profiles with relevant data mining extracts
Goal 2: Improve quality of business functions	
	a. Accessibility of unique customer profiles by all business functions
	b. Automatic propagation of business tactics along all the affected business functions
Goal 3: Retain profitable customers in the long-run	
	a. Automated estimation of life-time value for individual customers
	b. Facility of negotiating on rejected applications coming from 'good' customers
	c. Facility of re-examining rejected applications after a period of time
Goal 4: Tailor products to specific customer needs	
	a. Automated support for specifying agreement terms according to customer profiles
Goal 5: Identify profitable customer groups	
	a. Automated co-ordination of massive mailings
	b. Feeding of data mining extracts into mailing co-ordination systems
	c. Support for aligning the desired number of mailings with the acceptance probability of customer groups
	d. Clustering of customers according to their life-time value
Goal 6: Score applications effectively	
	a. Improvement of the algorithm for underwriting customer applications
Goal 7: Decrease processing costs	
	a. Replacement of the external customer scoring application with an internal one
	b. Effective management of frequency and timing for contacting individual customers
Goal 8: Contact individual customers who are more likely to respond	
	a. Feeding of data mining extracts into the telemarketing support system
Goal 9: Contact individual customers whose applications are more likely to be accepted	
	a. Feeding of customer scoring information into the telemarketing support system

Figure 5 provides a high-level view of the bank's business processes also considering the rule sets that are used in each case.

Currently, the 'decline' and 'pend' rule sets are the only explicitly expressed and documented sets of business rules within the bank's functions. The 'marketing strategy' represents approaches and tactics in organising massive mailings, being based on the analysis of the bank's marketing database contents. The 'algorithm of the telemarketing support system' refers to several behaviour characteristics of the 'TCS' system. Finally, the 'scoring policy' deals with the current method for estimating

customers' credit and behaviour scores. This high-level view, depicted graphically in Figure 5 is the current way of working. To demonstrate the impact of the discovered knowledge on re-designing and improving the current way of working there were three topics of relevance: (1) Correlation between input and output variables; (2) Validation of existing business and system functions; and (3) Management of 'rule' related knowledge.

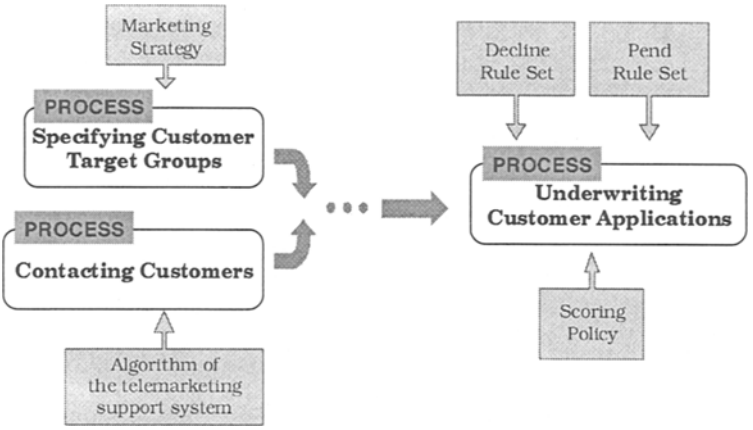

Figure 5: High-level view of the bank's processes

Correlation between input and output variables

The discovered knowledge resulted in statements of the following type: `"Group x of customers is related to an application acceptance probability z%"`. Decision-makers are potentially aware of the existence of such rules, due to their intuition and experience. However, it is very useful to move from fuzzy estimations to more explicit and quantified correlations like the ones described above. These new rule expressions can be applied in targeting customers more accurately, by aligning the number of mailings that the company can afford to send with the usual response rate of customers. Other rule statements of the type `"Group x of customers is related to an application acceptance probability z% for product y"` may contribute to the maximisation of customer response by aligning offers with customer needs. These new rules are grouped under the 'customer targeting rule set' (Figure 6), while their integration in the new IS will satisfy goals 1, 5, 8 and 9 of Table 4.

Similar clusterings may facilitate understanding of individual customers, so that maximum response to offers is ensured. A number of data mining extracts may result in business rules ('contacts co-ordination rules') for determining which is the best time for contacting a certain type of customer, and which are the most suitable products to be offered. In this way intelligence is added to the telemarketing support system being responsible for the co-ordination of telephone contacts (goals 1 and 8 in Table 4).

Validation of existing business and system functions

The knowledge discovery experiments showed that several operations (on business and system level) are not that much dependent on certain data attributes as previously thought. This observation raised the need for revisiting the corresponding processes and legacy support systems, and for improving them in case the problem is due to inaccurate input data, or erroneous algorithms for processing them, rather than to inaccurate data mining extracts. Such improvement efforts constitute an indicative example of cases where knowledge discovery practices are used for validating the current business and systems.

Figure 6 reflects changes of the bank's processes advocated by the previous observations:

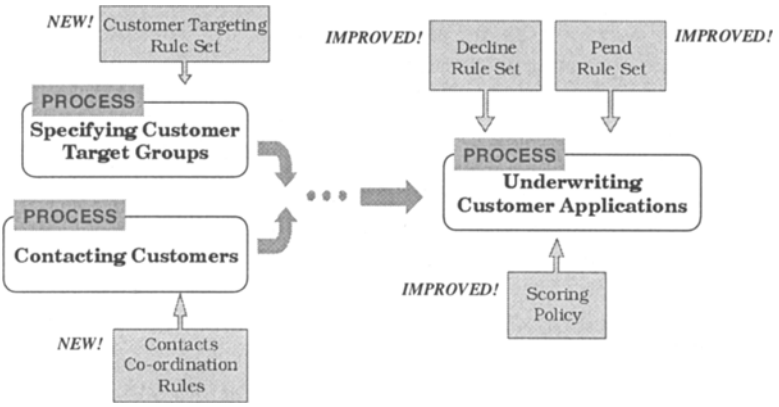

Figure 6: High-level view of the future situation

Management of rule-related knowledge

Business policies, regulations, guidelines, and also discovered knowledge expressions within a large enterprise can be considered as a complex net of interactions, given that they all make reference to the same business objects and activities. Appropriate tools can facilitate the management of this knowledge, by identifying the dependencies and conflicts between rule expressions and by propagating potential changes (according to newly discovered knowledge or new business strategic decisions) along the affected business processes, the people that need to be informed and involved and the systems that need to be modified or customised accordingly (goal 2 in Table 4).

6 Conclusions

Migration of legacy IS is becoming increasingly important as organisations face significant pressures for change. Such a migration however, is a complex undertaking, requiring the commitment of large resources with uncertain results at the outset. Therefore, understanding the relationship between the needs of the business domain and the capabilities of their support IS is of critical importance. Such an

understanding would lead to better planning with a more evaluative way of assessing the risk of the migration task.

In this paper we have put forward an approach to developing this understanding through the alignment of the business itself to its legacy IS. Central to this approach is the confluence of two techniques namely, enterprise knowledge modelling and knowledge discovery from data. The former is concerned with the development of models pertaining to business objectives, business processes and business rules whereas the latter is concerned with the development of models from the discovery of business behaviours from the existing IS. We have presented a way of utilising these techniques and demonstrated their utility in terms of a large banking application. By integrating the two sets of models it is possible to: (a) identify those parts of the legacy IS that require improvement to the extent that they will meet the stated objectives for change and (b) improve the business knowledge in terms of opportunities that may be available through the innovative exploitation of hidden knowledge.

7 Acknowledgements

The work reported in this paper has been partly supported by the commission of the European Union under the ESPRIT programme. The authors wish to acknowledge the assistance of Mr John Keane, Mrs Sofia Svinterikou and Mr Bob Scott in the insights of the data mining results.

8 References

ATTAR Software Limited. (1996) *XpertRule Profiler: Knowledge from Data, 1996.*

Brodie, M. and Stonebraker, M. (1996) Migrating Legacy Systems, Morgan Kaufmann Publishers Inc, San Francisco California, 1996.

Bubenko, J., Rolland, C., Loucopoulos, P. and de Antonellis, V. (1994) *Facilitating "Fuzzy to Formal" Requirements Modelling*, IEEE International Conference on Requirements Engineering, 1994.

Dhar, V. and Tuzhilin, A. (1993) *Abstract-Driven Pattern Discovery in Databases*, IEEE Transactions on Knowledge and Data Engineering, Vol. 5, No. 6, December, 1993, pp. 926-938.

Filippidou, D., Keane, J., Svinterikou, S., Murray, J., (1998) *Data Mining for Business Process Improvement: Using the Hyperbank Approach*, PADD98, 26-28 March 1998, London, U.K.

Kavakli, V. and Loucopoulos, P. (1998) *Goal Driven Business Analysis: An Application in Electricity Deregulation*, CAiSE*98, 8-12 June 1998, Pisa, Italy.

Keane, J., Murray, J., Scott, B. and Svinterikou, S. (1997) *Preliminary Analysis of Data Mining Results*, UMIST, Hyperbank WP3/T3.5/U/01, 1997.

Loucopoulos, P. (1994) *The F^3 (From Fuzzy to Formal) View on Requirements Engineering*, Ingénierie des Systèmes d'Information, Vol. 2, No. 6, 1994, pp. 639-655.

Loucopoulos, P. and Katsouli, E. (1992) *Modelling Business Rules in an Office Environment*, ACM SIGOIS, No. August, 1992.

Loucopoulos, P. and Kavakli, E. (1995) *Enterprise Modelling and the Teleological Approach to Requirements Engineering*, International Journal of Intelligent and Cooperative Information Systems, Vol. 4, No. 1, 1995, pp. 45-79.

Loucopoulos, P., Kavakli, V., Prekas, N., Rolland, C., Grosz, G. and Nurcan, S. (1997) *Using the EKD Approach - The Modelling Component*, UMIST, The ELEKTRA Project WP2/T2.1/UMIST/1, April 1997, 1997.

Ould, M.A., (1995) *Business Processes: Modelling and Analysis for Re-engineering and Improvement*, John Wiley & Sons Ltd, U.K., 1995.

Matheus, C.J., Chan, P.K. and Piatetsky-Shapiro, G. (1993) *Systems for Knowledge Discovery in Databases*, IEEE Transactions on Knowledge and Data Engineering, Vol. 5, No. 6, December, 1993, pp. 903-913.

Rolland, C. and Grosz, G. (1994) *A General Framework for Describing the Requirements Engineering Process*, IEEE International Conference on Men, Systems, and Cybernetics, IEEE Computer Society Press, San Antonio, Texas, 1994.

Tej Anand, A.T.G.I.S. (1995) *Commercial Knowledge Discovery Applications*, KDD95, Montreal, Canada, 1995.

Yoon, J.P. and Kerschberg, L. (1993) *A Framework for Knowledge Discovery and Evolution in Databases*, IEEE Transactions on Knowledge and Data Engineering, Vol. 5, No. 6, December, 1993, pp. 973-979.