DATABASES AND INFORMATION SYSTEMS INTEGRATION USING CALOPUS: A CASE STUDY

Prabin Kumar Patro, Pat Allen, Muthu Ramachandran, Robert Morgan-Vane, Stuart Bolton
15 Queen Square, Leeds, LS2 8AJ West Yorkshire, U.K

Keywords: Database Integration, Systems Integration, Workflow, Hybrid Integration, Enterprise Application Integration, Calopus.

Abstract: Effective, accurate and timely application integration is fundamental to the successful operation of today's organizations. The success of every business initiative relies heavily on the integration between existing heterogeneous applications and databases. For this reason, when companies look to improve productivity, reduce overall costs, or streamline business processes, integration should be at the heart of their plans. Integration improves exposure and, by extension, the value and quality of information to facilitate workflow and reduce business risk. It is an important element of the way that the organization's business process operates. Data integration technology is the key to pulling organization data together and delivering an information infrastructure that will meet strategic business intelligence initiatives. This information infrastructure consists of data warehouses, interfaces, workflows and data access tools. Integration solutions should utilize metadata to move or display disparate information, facilitate ongoing updates and reduced maintenance, provide access to a wide variety of data sources, and provide design and debugging options. In this paper we will discuss integration and a case study on Hybrid Enterprise Integration within a large University using the Calopus system.

1 INTRODUCTION

In today's competitive marketplace, it is essential to have an accurate, up-to-date, flexible environment to assess the business and make strategic decisions for the organization. The best way to do this is through an integrated environment. With the ever changing options available technically, businesses need to decide how to integrate their applications. To be successful, the organization must run its operations effectively and efficiently—which requires the ability to analyze operational performance. For an organization to thrive, or perhaps even survive, operations and analysis must work together and reinforce each other.

Without the whole business picture, it is difficult to make sound and dependable business decisions. That is because good decision making requires a complete and accurate view of data. Though the organization needs a complete view of operations, the data it needs often resides in a variety of application systems that do not necessarily all use the same database management system. These application systems may only contain current data values. They may not store prior data values needed to provide historical context and to discover trends and causal relationships.

Data integration allows an organization to consolidate the current data contained in its many operational or production systems and combine it with historical values. And the creation of a data warehouse (or, on a more limited scale, a single-subject data mart) facilitates access to this data. Collecting and consolidating the data needed to populate a data warehouse or data mart and periodically augmenting its content with new values while retaining the old is a practical application of data integration. In this paper we discuss the importance, technology and also present a case study of real-time hybrid integration implementation within a large University using the Calopus system.

Kumar Patro P., Allen P., Ramachandran M., Morgan-Vane R. and Bolton S. (2006).
DATABASES AND INFORMATION SYSTEMS INTEGRATION USING CALOPUS: A CASE STUDY.
In *Proceedings of the Eighth International Conference on Enterprise Information Systems - DISI*, pages 200-207
DOI: 10.5220/0002496902000207

2 IMPORTANCE OF DATA INTEGRATION

Database integration has various benefits for the organization. Database integration increases the number of experts viewing and manipulating the data by increasing the exposure of the data. It also helps the organization by detecting and correcting more errors in the data. The utilization of data in the workflow also increases and hence enhances the quality, trust and decreasing the business risk.

The process of integrating a particular category of data can be thought of as a progression. Initially the organization has only a single-source version of the data. Over time more sources become available. These multi-source versions are brought together until a single comprehensive system is created that relates these "variants" of the information and resolves the inconsistencies. (Steve Hawtin et.al, 2003)

Integrated data provides a framework that helps organization by delivering a complete view of a customer and standardizing business processes and data definitions. Integrated data helps to combine the current and past values from disparate sources in order to see the big picture. It helps to offload the processing burden on operational systems. At the same time this increases the effectiveness of its data access and analysis capabilities. (Steve Hawtin et.al, 2003)

3 RELATED WORK

Database integration is often divided into scheme integration, instance integration and application integration. Scheme integration reconciles schema elements (e.g., entities, relations, attributes, relationships) of heterogeneous database. (Kim W, Seo J, 1991) Instance integration matches tuples and attribute values. Attribute identification is the task of schema integration that deals with identifying matching attributes of heterogeneous databases. Entity identification (Lim E-P et.al, 1993) is the task of instance integration that matches tuples from two relations based on the values of their key attributes. Application integration involves storing the data an application manipulates in ways that other applications can easily access. (Ian Gorton et.al, 2004) Meta-data management has become a sophisticated endeavour. In the present world, nearly all components that comprise modern information technology, such as Computer Aided Software Engineering (CASE) tools, Enterprise Application Integration(EAI) environments, Extract/Transform/ Load (ETL) engines, Warehouses, EII, and Business Intelligence (BI), all contain a great deal of meta-data. Such meta-data often drives much of the tool's functionality. (John R Friedrich, 2005)

Database integration can take many forms. There are three main forms of integration: Extract Transform and Load (ETL), Enterprise Application Integration (EAI), and Enterprise Information Integration (EII).

ETL refers to a process of extracting data from source systems, transforming the data so it will integrate properly with data in the other source systems, and then loading it into the data warehouse. ETL simplifies the creation, maintenance and expansion of Data warehouses, data marts and operational data stores. ETL is either batch, near real-time, and sometimes real-time. (Surajit Chaudhuri et.al, 2004, CoreIntegration, 2004)

Enterprise Application Integration (EAI) combines separate applications into a cooperating federation of applications by placing a semantic layer on top of each application that is part of the EAI infrastructure. EAI is a business computing term for plans, methods, and tools aimed at modernizing, consolidating, and coordinating the overall computer functionality in an enterprise. (Jinyoul Lee, 2003) Typically, an enterprise has existing legacy applications and databases, and wants to continue to use them while adding or migrating to a new set of applications that takes advantage of Internet and other new technologies.

Previously, integration of different systems required rewriting codes on source and target systems, which in turn, consumed much time and money. Unlike traditional integration, EAI uses special middleware that serves as a bridge between different applications for system integration. All applications communicate using the common interface.

EII: Enterprise Information Integration

EII provides real-time access to aggregated information and an infrastructure for integrated enterprise data management. While the graphical EII data mapping tools are easy to use and speed the integration process, the information they capture is valuable corporate information required for enterprise data quality management. EII helps to capture the metadata to drive data transformations, is the same information required for enterprise information management. (Beth Gold-Bernstein, 2004).

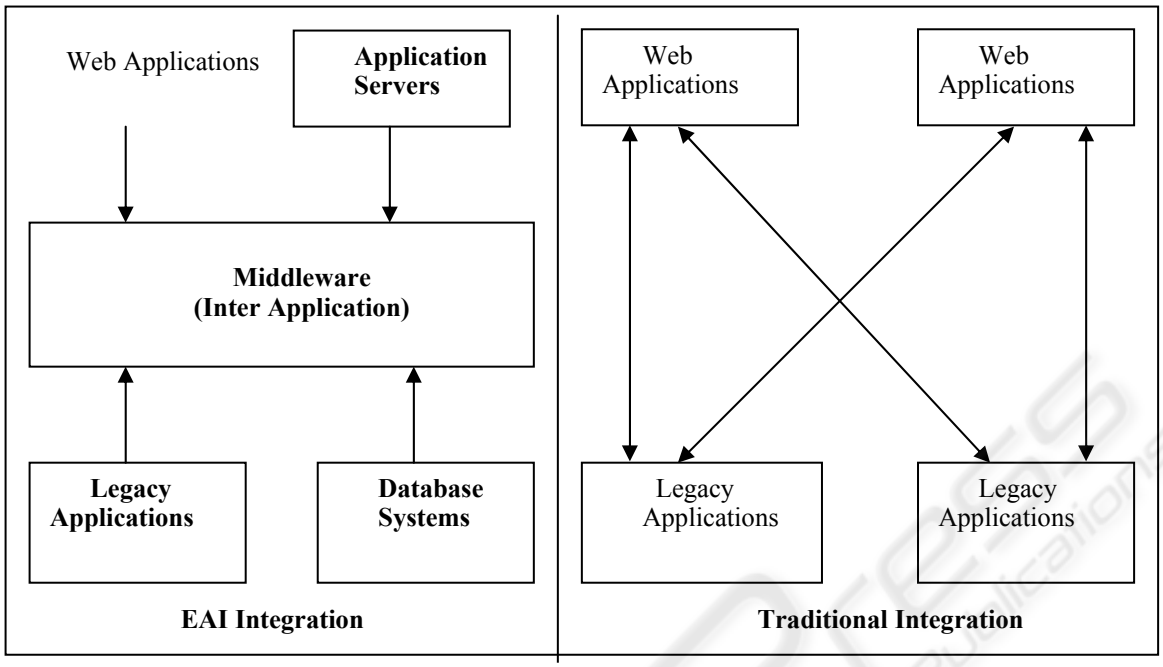

Figure 1: EAI Integration vs Traditional integration, Jinyoul Lee (2003).

4 INTEGRATION THROUGH CALOPUS

Calopus defines the structure in Metadata layer. It automates the process and uses a replication approach while performing Database Integration. Calopus integration solution supports all three methods of Database Integration ETL, EAI and EII with a codeless environment and hence can be called a hybrid solution. Calopus deploys ETL, EAI, and EII instantly and each can be called from the other.

Calopus Integration Strategy:

The following are the activities of Calopus ETL Interfacing Strategy:

- Logical datasets are defined for each source system and can be files, XML events or SQL Tables/Views.
- Each data set is frequently scanned for additions, deletions and changes against an internal control table of last known data
- Each difference can trigger qualified complex events such as file creation, SQL updates, emails, XML streams and LDAP
- Calopus defines the structure in a Metadata layer.

Apart from this, the Calopus Data Warehouse can contain a core set of frequently used records and data items which can be extended by configuration of Meta Data with little or no coding.

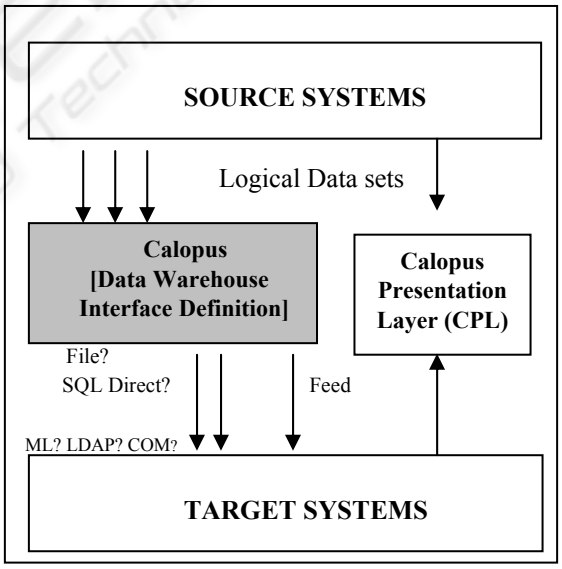

Figure 2: Calopus Integration Strategy.

The Calopus Data warehouse allows secure, qualified and role based access to transferred data and remote data. Calopus supports a hybrid approach by seamlessly deploying ETL, EAI and EII instantly within a codeless environment. The Calopus hybrid approach provides flexibility in allowing the data to reside its source system and

accessing it via e-forms and workflows, or transferring the information from a source system to a target system using an ETL approach. Calopus supports ETL (Extract/Transform/Load) by extracting data from a source system, transforming the data and loading it into a Data Warehouse or target system. Calopus employs an Interface Designer graphical tool to facilitate this. Calopus supports EAI through e-forms, menus and workflows. Calopus maintains EAI by having a common interface (Calopus Presentation Layer) which serves as a bridge between different applications for system integration. All applications communicate using this common interface (CPL). Calopus supports EII (Enterprise Information Integration) by capturing metadata. Calopus workflow is a combination of EAI and EII which helps an organization to easily and visually define job roles, processes and tasks. Apart from this, Calopus is a development environment that uses Meta Data to configure interfaces, workflows, forms applications, extranets and portals. This means that as the business changes, Calopus adapts to this change through its configuration, rather than through wholesale redevelopment of more application code. These are illustrated in the following case study.

5 CASE STUDY: BUSINESS SYSTEM INTEGRATION WITHIN A LARGE UNIVERSITY

In nineteen ninety-nine a UK University was identified that had started a six year rolling programme to review and replace all its key corporate systems. The existing systems were a mixture of bought in solutions and in-house developments. As part of this programme application interfacing and integration was clearly an issue, particularly once the decision was made to pursue a 'best of breed' policy.

5.1 Technical Issues

5.1.1 Earlier Interfacing Strategy at the University

There was a long-standing policy in the University that in order to maximise data consistency, information was entered a minimum number of times, preferably only once into an 'owning' system. Wherever possible mission critical data was passed from an owning application to other applications that needed this information via interfaces. This policy provides efficiency gains and it was rapidly agreed that the policy should continue.

The existing strategy had been to hand build interfaces on a demand model and to use the knowledge of the in-house systems to leverage efficient and accurate data exchange. Figure 3 shows the information flows/interfaces for a University administration system involving Student, Library, Finance and Marketing System. The network of interfaces had built up over a number of years as new systems had been deployed across the University. It was realized that the existing ad-hoc interfacing methods could not be used. The University implementation plan did not have sufficient time for each existing interface to be individually reengineered to meet new supply and demand criteria. A coherent and manageable solution was required.

5.1.2 Design Constraints

One major problem was the different processing rules required in each system. Every information feed seemed to have different requirements in terms of the information required.

5.1.3 Timely Availability of Information

Many of the interfaces related to student facilities such as IT suite accounts, library accounts, halls of residence etc. It was important that as soon as the student arrived in the University these facilities were available to them. A major design decision was whether the interfacing mechanism needed to work in 'real time' or whether the delay imposed by batch processing was acceptable. This decision was complicated by the highly variable amount of activity required by interfaces at different times of the year.

5.1.4 Disparate Data Destination

The University's corporate systems run on either Oracle or SQL Server databases. However the destinations for interface data are diverse. For example the student IT authentication is to a Novell tree, ATHENS is a remote 'black box', the student and staff card system runs Linux. Each application has its own distinct import mechanism with distinct data formats and field sizes. The interface mechanism was required to deal with any necessary data transposition and output format required.

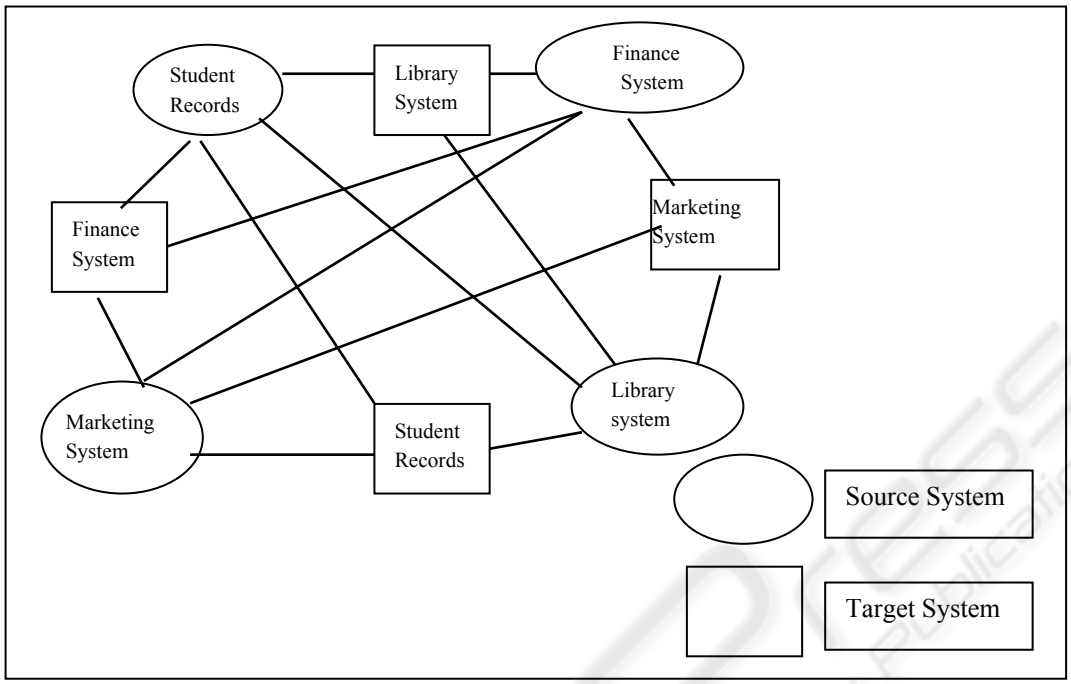

Figure 3: Historic Information flow and interfacing within the University.

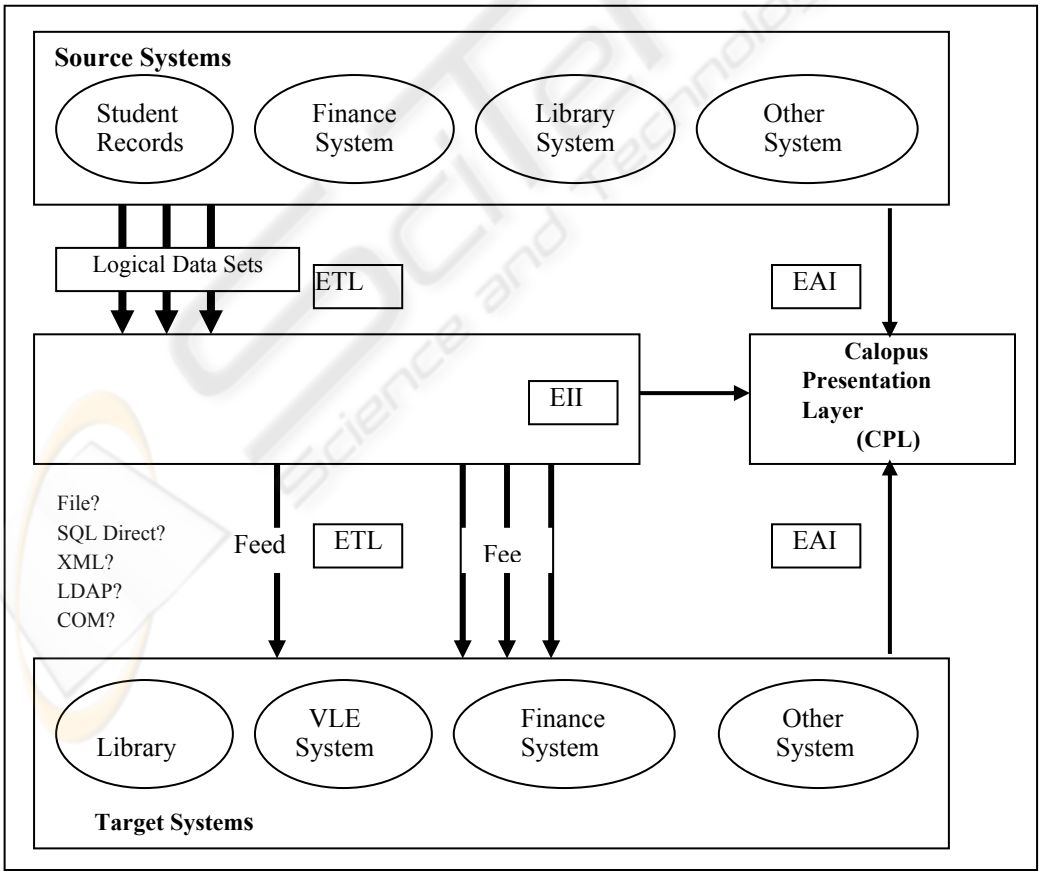

Figure 4: Calopus Interfacing Strategy.

5.1.5 Maintaining Consistent Information

Interfaces need to not only provide new information but also to provide updates to client systems when a required data item changes.

5.1.6 Develop or Purchase

This decision was principally resource driven. The University had the staff with the necessary skills to develop this mechanism. However these staffs were already committed to other aspects of the development programme. The Calopus product met the University's outline design and supported their strategic requirements. The University decided to implement Calopus for the Data Integration with in the University.

5.2 Calopus Product Overview and Its Technology

Calopus is a Web SDK built in the PL/SQL language to enhance the standard features of Oracle HTTP Server or 9i Application Server. It achieves this by providing an application layer (missing from the base Application Server software) with a rich set of core features on top of which business processes can be easily and quickly built. Calopus aims to provide a complete, coherent and secure solution for all of the "System" requirements of a business application or portal. When developing an Application or Portal, the development team does not need to devote resources to such issues as Session Management, Role Based Security, Data Auditing, Menu Systems, Content Management, Corporate Branding, Display Preferences, Data Integration or File based data loading.

5.3 Using the Calopus SDK to Solve the Integration Problems at the University

To maximize development speed, it was first decided to adopt a hybrid approach of two possible strategies. Namely:

- Implement a coordinated sequence of batch interface jobs across database links using the job scheduling features within Calopus;
- Implement a coordinated sequence of batch interface jobs writing the output to file for subsequent upload to the target systems.

The University already has a number of tried and tested file upload mechanisms in place at a significant number of their target systems. It therefore made sense to utilise these programs and concentrate on customising a sequence of data extraction, manipulation and file creation jobs. Calopus is a PL/SQL system, and therefore can perform the complete sequence of job scheduling, Data Selection, Transformation and File Creation within the same set of stored procedures. It can also be configured to seamlessly import Data from a "Flat File" rather than directly by a SELECT statement. This ability to code the whole solution in one programming language and within only one database transaction creates a seamless solution with little or no platform dependencies.

As time progressed, the University moved towards more dynamic interfaces and Calopus has been enhanced to support these changing interfaces. In order for the University to be able to administer, configure and further develop their Interface Solutions, a Calopus Web Extranet Application was deployed from the structure of the Transformation Database Schema. This allowed a role secured view of the interface data and its historical changes and also any set-up information required by Calopus to design the interfaces. The Extranet was defined to operate with different views of the Interface information for different University departments.

As the interfacing mechanisms were already coded within the Calopus SDK framework, the development times and costs of creating more than ten interconnecting Interface solutions at the University were greatly reduced in comparison with a ground up development. Indeed, Calopus allowed a single skilled Oracle developer to configure each of all of the University's interfaces within one working day of the release of signed off technical requirements documentation.

While implementing Calopus for the University the source and targets systems are identified. Figure 4 shows the information flows/interfaces for the University administration system involving Student, finance, Library and Marketing system through Calopus. As can be observed in Figure 4 **(**For understanding please compare this with the earlier Interfacing strategy of University shown in Figure 3**)**, the Calopus Interfacing strategy applies a common approach for interfacing all systems with in the University for reducing the system maintenance.

6 RESULTS

The project was initially reviewed post go-live for the student record system in 2002. A series of

informal reviews have followed as other applications were launched and integrated.

Positive Outcomes:

- Technically the project has been a success and accurate data is correctly passed between applications. Best estimates currently are that about 97% of student records are passed through the interface mechanism without any issues.
- The design was correct and the product has proved robust. Despite some misgivings a metadata driven model has proved both practical and efficient.
- The solution was designed to be scalable and its use has continued to grow as the programme of rolling replacement of key corporate systems has continued over the last three and a half years.

Issues:

- In the first year, the user community was under prepared for the discipline required for all the interfaces to trigger correctly. Consequently too much data was entered but not passed through the interface either because it failed validation or was not entered in a timely fashion. This was addressed by a programme of education and by the user community experiencing direct and significant advantages to carrying out the process correctly.
- The complexities of the interdependencies were underestimated and a more complete set of Standard Operating Procedures were required at both the business and technical level. The introduction of these has been a long but successful project with key business processes now better and more widely understood than at any previous time.

Overall the project has been a significant success. Information sharing could not have been as flexible or as complete without some type of standardised interface mechanism. Calopus has fulfilled most of the requirements of the University in this area.

7 CONCLUSION AND PROPOSAL FOR FUTURE WORK

Database integration plays a very important role in modern generation for industries. Effective data integration is an essential requirement for any organization. Calopus has been a cost effective, powerful and quick tool for creating dynamic web applications. It has been found that the development effort and cost are minimised. Calopus has been installed and successfully running in eight universities in U.K.

Flexibility to changing business requirements is key to providing effective IT solutions that can stand up to the tests of time. Calopus takes the approach that business processes should be the key driver for the functionality that IT systems deliver. The Business Model has great importance in the current Software Industry. Calopus supports Business model through its Enterprise Application Integration (EAI) and Enterprise Information Integration (EII). This allows a business analyst to visually define job roles, processes and tasks in the form of Workflow Diagrams and at the same time, enhance these models with application configuration information. This enhanced model is then interpreted into a working and fully featured prototype system automatically by Calopus. As the model changes so does the prototype in real time. In our experience and understanding, it is vital for business model to be visible not only to systems architects, but also to end users of the system. End users visualize a system more easily when presented with a working prototype that they can access, use and change easily. This prototype should accurately reflect the business model and change with it.

It may be possible for an end user to understand a business model enough to say that it truly represents their business. However, this model is then interpreted by IT specialists into a working system or delivered by package selection. Quite often the clarity of the business model is then lost in the technicalities of traditional systems development or some processes are not supported within a selected package solution. With the Calopus EAI solution, this does not have to be the case. Continual research is being carried out on this aspect. The aim is to create a framework where a business analyst can produce a fully featured prototype system as the details of the requirements emerge. The prototype will be self documenting through its use of designers and META Data and deployment is instant so that the requirement can be checked against the prototype without a “generative” phase- closure between requirements and prototype is bound to

follow. Calopus is very close to this aim of creating the "Model Deployed Application Framework". The Calopus data integration product is aiming to provide more code less environment and at the same time enhancing the tool with new integration methods as they arrive.

REFERENCES

Kim W, Seo J (1991) Classifying schematic and data heterogeneity in multidatabase systems. IEEE Comput 24(12):12–18

Lim E-P, Srivastava J, Prabhakar S, Richardson J (1993) Entity identification in database integration. In: Abstracts of the ninth international conference on data engineering, Vienna, pp 294– 301

Ian Gorton,Anna Liu (2004), Architectures and technologies for Enterprise Application Integration, National ICT and Microsoft, IEEE, pp. 726-727, 26th International Conference on Software Engineering (ICSE'04), 2004.

Surajit Chaudhuri, Umeshwar Dayal(1997), An Overview of Data Warehousing and OLAP Technology, Microsoft Research and Hewlett-Packard Labs, Volume 26 , Issue 1, Pages: 65 - 74

Jinyoul Lee (2003), Enterprise Integration with ERP and EAI, Vol.46, No.2, Communications of the ACM.

John R Friedrich, II (2005), Meta-Data Version And Configuration Management In Multi-Vendor Environments, Meta-Integration Technologies, California

Steve Hawtin, Najib Abusalbi,Lester Bayne (2003), 'Data Integration: The solution spectrum', Schlumberger Information Solutions, Houston, Texas.

Cortney Claiborne, Darren Cunningham, Davythe Dicochea, Erin O'Malley, Philip On (2004), 'Data Integration: The key to effective Decisions', MAS Strategies, California

Beth old-Bernstein (2004) "*Enterprise Information Integration, What was old is New again*",DMReview, USA, Retrieved September 21, 2005, from , http://www.dmreview.com/portals/portal.cfm?topicId=230290

CoreIntegration (2004), *Data Integration,* Core Integration Partners Inc., 2004, Retrieved September 20,2005 from [http://www.coreintegration.com/solutions/di.asp]