University of South Australia

The Next Generation of Data Warehouse:
The Story So Far,
Getting UniSA RQF-Ready

Authors:

Tara Hemingway, Manager: Business Intelligence

Tony Dalwood, Manager: Information Architecture and Management

Abstract
2005 saw the commencement of the Stage 1 design of a high performance enterprise data warehouse (edw) that is dynamic, will stand up to change, is scalable and flexible to accommodate both current and known future data streams and will deliver enduring value to the University of South Australia (UniSA).

Driven by the business, in partnership with IT, the project requires its team to work across organisational boundaries into discrete business areas. To date, the project has delivered:
• The foundations of a high-performance, flexible data warehouse architecture.
• Standardised and conformed student and research activity data and reference data ("one version of the truth").
• The foundations of a data quality framework which addresses both intrinsic (data validity, reliability and accuracy) and extrinsic dimensions that relate to the business needs of data consumers (accessibility, timeliness, relevance and conciseness; metadata).
• Better quality reports and cubes in UniSA's management reporting environment, UniSAinfo Reporting.
• Noticeable improvement in speed of complex reports with data from different source systems.
• Support infrastructure and processes to build user understanding of the new products.

A significant benefit of this project has been in the shift of effort for users with a greater emphasis on analysis rather than data gathering. Strategic benefits from Stage 1 have already been realised in this way. The integration of Research Activity has seen the production of a suite of cubes and reports for the foreshadowed national Research Quality Framework (RQF) derived from the edw and delivered to users through UniSAinfo Reporting. UniSA is now a RQF-ready institution.

This presentation will cover the journey that is the UniSA edw; where the major stars are two unlikely areas working closely together with limited resources and little support from the top. The paper highlights the frustrations, the workarounds and what have become the obvious critical success factors to making this project deliver what it has promised.

1.0 Introduction

An *enterprise data warehouse* (edw) is a data store designed to produce a single, comprehensive view of data that an organisation accumulates during its course of operations.

In 2005 the University of South Australia (UniSA) commenced a *Stage 1* build of a high performance edw. This edw is dynamic, will stand up to change, is scalable and flexible to accommodate both current and known future data

streams and will deliver enduring value to the University within an increasingly complex and competitive environment.

Driven by the business, in partnership with IT, the project requires its team to work across organisational boundaries into discrete business areas. To date, the project has delivered:

- The foundations of a high-performance, flexible data warehouse architecture.
- Standardised and conformed *Research Activity* and reference data.
- Better quality reports and cubes in UniSA's management reporting environment, *UniSAinfo Reporting,* that address the requirements of the Research Quality Framework (RQF).
- Noticeable improvement in the speed of complex reports using data from different source systems.
- Support infrastructure and processes to build user understanding of the new products.
- Shift of effort for users with a greater emphasis on analysis rather than data gathering.
- A new data quality framework, which has enhanced the validity, reliability, accuracy, accessibility, timeliness, relevance and conciseness of the data incorporated into the edw.

This paper analyses the *Stage 1: Research Activity* experience and explores the frustrations, the workarounds and what have become the obvious critical success factors, for UniSA, in making this project deliver what it promised.

We will discuss three key areas of critical success factors – People, Business and Technology - with reference to literature on "good practice" and the "real" UniSA experience. This reflective approach, with the benefit of retrospection, will assist the reader to gain some understanding of the complexity of data warehousing projects and how the presence of certain key factors will mitigate risk and, most importantly, manage for success.

2.0 UniSA's Business Intelligence (BI) System

Negash and Gray (2003) state that BI systems combine data gathering, data storage, and knowledge management with analytical tools to present complex and competitive information to planners and decision-makers. Implicit in this definition is the view that BI systems provide actionable information delivered at the right time, right place and in the right form to assist decision-makers. The objective is to improve the timeliness and quality of the input to the decision process, hence facilitating the work of both knowledge workers and managers (Dresner et al 2002).

Crucial to the UniSA BI system was the creation of a coordinated and consolidated business intelligence environment – both from an organisational and functional perspective. Importantly, this environment needed to

encourage users to meet their BI needs (data access, analysis and reporting) through self-service.

Consequently, in 2004, a number of Cognos products were implemented across the University (accessible to all 2000+ UniSA staff) to create *UniSAinfo Reporting*, a web based management reporting environment. This environment enables users to generate simple or complex information queries tailored to their local area, using data from the human resources, finance, research, student (domestic and international), marketing, project quality and evaluation corporate systems. The data are available as "canned" reports and multi-dimensional cubes, which the user can "slice and dice". Refer to **Appendix 1** for an overview of UniSA's BI system.

At UniSA we are committed to deploying the proper mix of BI solutions – both institution-wide and niche targeted, with no real commitment to any one vendor. However the critical objective of the BI system is to enhance the future corporate performance of UniSA and harnessing our data assets in a solid, yet flexible infrastructure is the foundation.

2.1 Why an enterprise data warehouse?

The quantity of data assets at UniSA has continued to grow at an exponential rate making it increasingly difficult to manage the volume held, but even more difficult to extract for strategic purposes. Clubb (2004, p.1) stated that "research shows the amount of information in the world doubles every three years and there is an increasing pressure on companies to extract value from it for their operations". This pressure arises because the data explosion is being accompanied by a concomitant growth in the number of high-impact strategic and tactical decisions required (Dresner et al 2002).

Furthermore, organisations are moving away from the traditional data-centric approach to BI to a process-centric one (Arcplan 2006, p.3). Process-centric BI is directly embedded into the underlying business processes, enabling managers to:

- Monitor end-to-end business processes in real-time.
- Analyse the corresponding impact on operational results and strategic outcomes.
- Correctively intervene into operational business processes and correct the corporate strategy, if necessary.

Consequently, a completely flexible process-centric BI system requires the tying together of disparate data sources. Organisations are now challenged with collecting and integrating all available data into a single data store – the data warehouse – to maintain a transparent view of the data and therefore empower managers to make decisions with confidence based on that data. Watson, Goodhue and Wixom (2002, p.493) state that a data warehouse is "a critical enabler for a major shift in an organisation's strategy".

Table 1 illustrates the architectures that are being adopted by UniSA. These attributes of the edw will ensure the business is delivered maximum insight through the deployment of a process-centric BI system, with the edw as its foundation.

Table 1: *Comparison of Traditional & Emerging Architecture (Knox 2002, p.1)*

Traditional Architectures		Architectures for the Future
By default	→	By design
IT-driven	→	Business-driven
Application focus	→	Infrastructure focus
Cost savings and control	→	Revenue-generation & efficiency
Buy lists	→	Alternatives

Rigid architectures optimize current processes, but stifle creativity

Flexible architectures stimulate innovation and growth

3.0 Stage 1: Research Activity

From the outset, business requirements have driven the design of the edw, not technology. The project team saw the edw as a tool for solving users' need for strategic information. Consequently, the most critical part of the project start-up was focusing on what information was needed. To determine this, a series of 25 workshops (170 academic and professional staff) were held across the University from February to May 2005. At each workshop, participants were asked what data were required to enable the University to conduct its core business, and assist individuals to meet the responsibilities of their positions. The data scoping workshops identified 38 themes and many gaps and priorities for required data.

The value gained from this type of activity cannot be underestimated, despite its resource intensive nature:

- The business is involved in the project from the start and begins to develop an understanding of the edw.
- The final outcomes of the workshops enable the project team to prioritise the data and determine the roll-out plan for the edw.

Consequently, the BI team in Planning and Assurance Services (PAS) identified up to four stages of data integration, with each stage estimated to take one calendar year. It was agreed that *Stage 1* would include reference

files, student bio/demographic data, program and course enrolments, student finance and funding data.

Research Activity (AVCC Benchmarking/ Publications/ Income/ Higher Degree Research) data also was added to this first stage due to the:

- the appointment to the University of a “data-driven” Pro Vice Chancellor: Research and Innovation;
- the foreshadowed Research Quality Framework (RQF); and
- the decision to begin the edw with a reasonably self-contained area of data.

This paper focuses on this first part of *Stage 1*, the *Research Activity* data which saw the integration of the following:

- **Research Activity (RA)**
    - Proof of Concept - AVCC External Benchmarking data (Publications/Income/Completions)
    - DEST Staff Aggregated data (external source)
    - Research Publications data
    - Research Income data
    - Higher Degree Research Student data
    - Higher Degree Research Student Supervision data
    - Staff Classification and Occupancy data
    - Supported Researchers data
    - ARC-RFCD Benchmarking data

The outcomes so far have been significant:

- standardised and conformed Research Activity and reference data.
- standardised Organisational Codes.
- a suite of reports to help UniSA monitor its RQF performance and make strategic decisions about research capacity and impact.
- the capacity to report across discrete business areas; standardising and conforming Research Activity data in the edw has enabled reporting across Human Resources, Finance, Research and Student systems and significantly reduced the time involved to do this.
- a shift from data gathering to analysis with the production of quality reporting products and an associated improvement in end user understanding of data.

So how did we get there? The following sections will analyse and discuss the impact on the project of the critical success factors - People, Business and Technology.

4.0 People Factors

When planning an edw project, quite some effort is necessary to manage the people factors: the type of project team, communication structures, planning

processes and project management. A data warehousing project involves organisational change, so robust support structures can ensure a successful implementation.

4.1 The project team: balancing the business and technical

A data warehouse implementation requires commitment from across the organisation, which begins with the establishment of a functional team, accommodating the appropriate business and technical skills, to lead the project.

At UniSA a cross-functional, appropriately skilled team was established comprising “business” staff from Planning and Assurance Services (PAS), the University’s key unit responsible for corporate performance management, and technologists from the IT unit, known as Information Strategy and Technology Services (ISTS). This approach enabled the design to be determined by the business and ensure that the needs of the BI system were a driver of the physical data architecture. This collaborative approach, which has been typically applied to other BI initiatives at UniSA, enables different team members to take a lead role at different stages, since each project phase requires different skills and expertise.

Table 2 details the UniSA established project team membership, excluding those representatives from the broader business.

Table 2: *UniSA established edw project team, 2005*

Business Commitment	Technical Commitment
Business Project Manager (0.3 FTE)	Technical Project Manager (0.4 FTE)
Business Analyst: edw (1.0 FTE)	Information Technologist: edw Architect/Development Team Leader (1.0 FTE)
Business Analyst: Information Systems (0.6 FTE)	Information Technologist: Analyst Programmer (1.0 FTE)
Business Analyst: Data Management (0.4 FTE)	Information Technologist: Analyst Programmer (0.5 FTE)
Business Analyst: Reporting Services (0.4 FTE)	Information Technologist: Analyst Programmer (0.5 FTE)
	Data Base Administration Support (As reqrd)

Skills, experience and knowledge are important attributes of the edw project team members, yet our experience has shown that attitude, team cohesiveness, passion for the edw effort, and strong commitment are equally important, if not more so. These traits are particularly important in moments of crises and conflict, which occur frequently over the life of the project.

4.3 Involving the broader business

The iterative nature of a data warehousing project means that there is no project team "standard" for a data warehousing effort. A core project team is vital, however as Clarry (2001, p.3) states, "successful data warehousing [efforts] come in a variety of shapes and sizes…but what is most consistent among the successful [ones]….is their ability to create adaptive structures that incorporate business participation and continue focusing on delivering business value". The benefit of adaptive structures is they support rapid change yet provide stability for individuals and the data warehousing effort. The focus is to build competencies in key skills and then structure specific task-focused projects by deploying those skills as needed. Adaptive organisations are characterised by clear, but changing, roles and responsibilities (Clarry, 2001). In addition, they rely on formal communication that follows organisational hierarchies but also cultivate informal communication laterally for the benefit of the project.

So what does this mean for an edw project? It assumes that the business, beyond the core team members, is involved. Their participation might be a temporary, task-focused data warehousing structure and their time commitment might be variable but they are seen as an integral part of the project.

At UniSA, the intention was to involve the broader business (in this case, Research and Innovation Services (RIS)) throughout *Stage 1: Research Activity*. However, this was not the reality. The new Pro Vice Chancellor: Research and Innovation required RIS to work on a number of new initiatives within a tight timeframe, to ensure UniSA was RQF-ready. Also, RIS staff were not always available at the critical points of the project which required their input, causing delays (especially those associated with source system data quality issues).

In hindsight, the RIS data custodian was actually required to be part of the project team across the life of *State 1: Research Activity*. This would have engendered an understanding of the "real" value of the edw to their work; ensured that the "core team" understood the latest priority business objectives of the research area and enabled us to better manage potential conflicts of our resource demands and their competing priorities. Interestingly enough, RIS involvement increased once they realised the edw was crucial to the delivery of one of their new PVC-driven initiatives.

The lesson learned here is that business sponsorship, commitment and their participation provide the initial and ongoing focus for delivering business value. Clarry (Wells 2005, p.6) summarises what makes the partnership effective for the benefit of the project:

- The creation of joint responsibility for the project and its outcomes.
- Ensuring everyone takes an active part in the problem resolution, both the core project team and the wider business.
- Involving the team members and business early and often in the project in a meaningful way.
- Allowing the team to be successful early, collectively and build from there.

Clarry (2001) advocates that a good data warehousing team would spend 70% of its energy on tasks and 30% on building relationships. Investing time in communication, participation and appreciation ensures that all individuals involved will be committed both to the success of the project and the success of each other.

4.4 Roles and responsibilities

Frank Buytendijk, Vice President, Gartner Inc points out that "most data warehouse project plans focus entirely on technical aspects of initial implementation and change is thereafter seen as a production issue. This does not reflect reality. Enterprises need to adopt a methodology where both business and IT align themselves during the entire life cycle of the data warehouse".

While most well-intentioned project teams would endeavour to heed this advice it is not always the case. Data warehousing projects constantly present new and varied difficulties, challenging project team members with competing

priorities and concomitant workload crises. It is then easy to adopt a business versus IT, or them and us, mentality.

Figure 1 was used in UniSA's initial business case to identify the responsibilities of the two areas of the project team and areas of essential cooperation. The diagram assisted with the establishment of the project team. However, due to the ever-changing nature of the edw, clarifying responsibilities on a regular basis – including overlaps and gaps - became important.

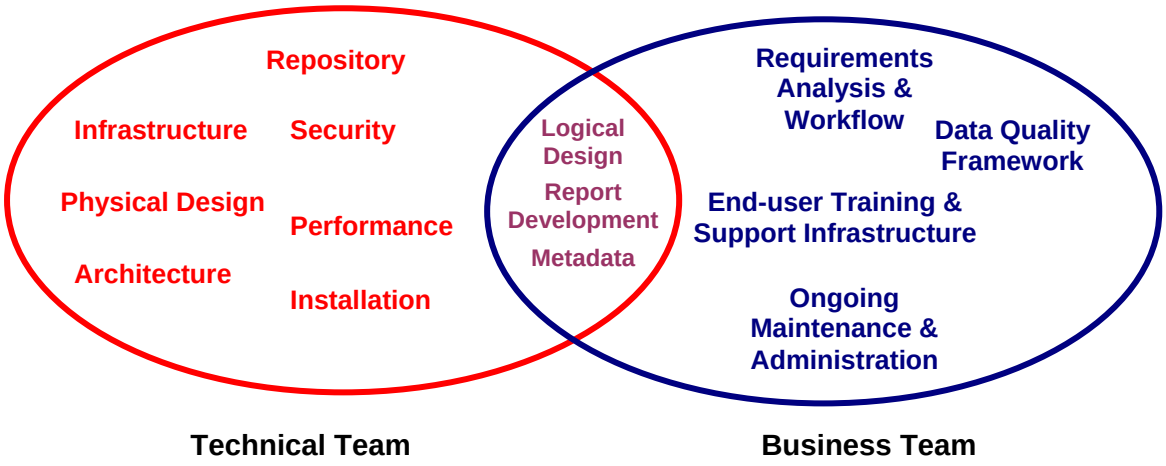

Figure 1: *Data Warehouse Implementation Team (adapted from Dresner et al, 2002, p.29)*

4.5 Project management

Project management for a data warehouse is very different from most other applications (Pohl 2006). These projects are characterised to be organic and ever changing. In addition, although there are phases with start and end dates, the data warehouse itself never reaches an end state until a decision is made to terminate it. These characteristics make data warehouse project management both unique and challenging.

The literature advocates that utilising a proven methodology for managing the design, development and deployment of a data warehousing initiative is essential (Furey 2005). UniSA adopted a "system development lifecycle" approach (refer **Figure 2**) and experimented with tools to facilitate this – such as *MS Project* and *SharePoint*.

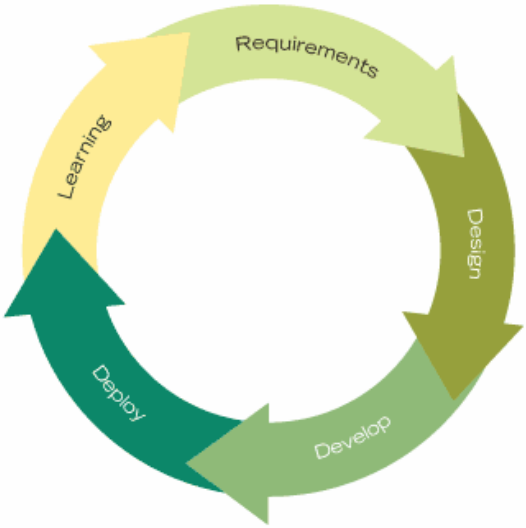

Figure 2: *System Development Lifecycle*

However, it is not the application of a project management methodology that is most challenging but rather project status monitoring and communication. Data warehousing projects require the coordination of numerous resources and tasks that must be integrated at the right time to ensure success. Furey (2005, p.2) states that “project status monitoring and communication are proven techniques for achieving the objective, as well as for measuring performance, deploying a quality solution and assessing stakeholder satisfaction”.

Finding the right balance of face-to-face meetings, electronic communication (email/discussion boards/portals), documentation (issues and testing registers) and reporting to governance structures is learned as the project progresses. The UniSA project team has favoured more communication within the team rather than less and has established both formal (project team meetings) and informal mechanisms (*SharePoint* site) to manage this.

5.0 Business Factors

Data warehouse proponents maintain that the culture and requirements of an organisation will determine whether an edw is the most viable solution (Ponniah 2001, p.64). These organisational features will provide the foundation of support for your project and contribute to its success.

5.1 Support from the top

The cornerstone to an edw project is securing full support of senior management from day one (Ponniah 2001). This high-level support provides critical approval and assistance in two ways; it:

- Provides budgetary commitment for the costs to develop, implement and maintain the project.

- Establishes senior leadership presence and commitment within organisational change management functions such as communications, education and compliance.

The latter point is particularly important; it acknowledges that data warehousing involves numerous stakeholders within an organisation, establishing it as a top-down corporate initiative. The literature states that a "sponsor from the highest levels of management is important to keep focus and be available to arbitrate and mediate when conflicting requirements arise" (Ponniah 2001, p.67).

At UniSA the drive to have an edw was initiated bottom-up and funding for *Stage 1* was secured by presenting a soundly rationalised business case to the University's IT capital funding group. However, this early success has since struggled to secure senior management support. This can be attributed to a number of factors:

- A general lack of understanding about data warehousing and business intelligence (a generational issue that is improving)
- A misconception that producing reports simply requires the throwing together of available data – "Why does UniSA need an edw?" (we are victims of our own "good service" to senior management).
- An inability to realise the significant strategic value that an initiative like this can deliver.

"A data warehouse without the full support of top management and without a strong and enthusiastic executive sponsor is doomed to failure from the outset" (Ponniah 2001, p.86). Recognising this pessimistic view, the project team took a different tact by securing high-level endorsement of particular phases of the project. The appointment of a "data-driven" Pro Vice Chancellor: Research and Innovation Services presented itself as an opportunity to the project during *Stage 1: Research Activity*. The PVC has certainly become an advocate for the project and whilst no additional resources have been offered, her advocacy has enabled us to establish a higher profile at all levels within this portfolio.

UniSA has adopted this approach with *Stage 2.* Support has been secured from the PVC: Academic for the *Student Evaluation* data phase and the Executive Director: International and Development for the *Student Application* data phase. This embeds the outcomes of the edw within the key business deliverables of each senior manager's portfolio. For example, a corporate priority project in 2007 for the PVC: Academic is the mining of student evaluation data.

5.2 Marketing the edw

Justifying the cost of a data warehousing project is not easy. In fact, real benefits may not be known until after the data warehouse is built and in full use.

Typically the benefits are measured in terms of process improvement and general efficiency rather than hard, tangible bottom-line figures. These benefits include greater business knowledge, improved business processes and more effective partnerships. Intangibles are difficult to quantify. They are, however, important sources of business value, particularly in regard to the use of BI. Traditional methods of calculating value perform poorly when assessing the impact of BI, as many of the benefits are strategic and consequently not easily quantifiable. This is well-documented in the literature (Gibson, 2005).

Specifically, a properly designed data warehouse provides a complete, accurate and easily accessible perspective of the organisation. It provides the university with "total value opportunity" and a foundation for high-impact, strategic BI. Many of the benefits of a warehouse stem from its ability to greatly reduce the amount of time needed to get information. Viewing the warehouse as a time-compression device, UniSA can develop a framework to assess the sources of business value. The framework is documented in **Figure 3**.

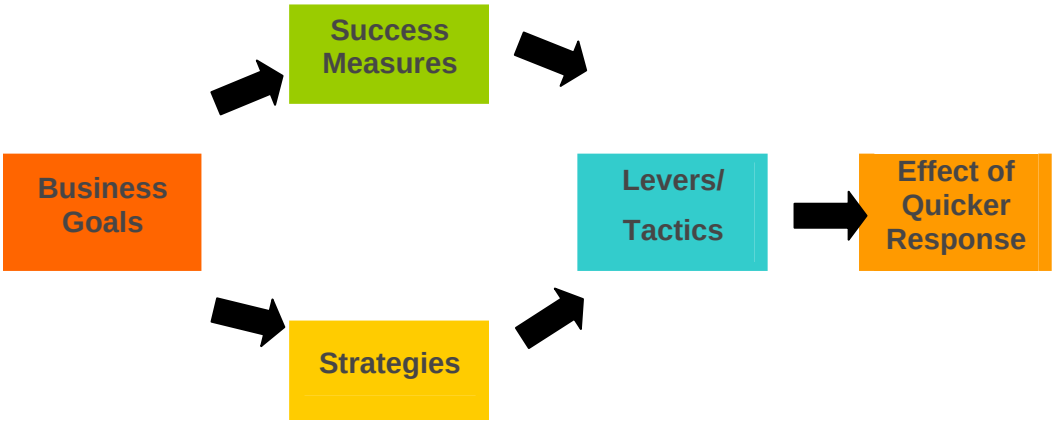

Figure 3: *Analysing the business value of a data warehouse investment*

However, how is this message conveyed to the sponsors and other key stakeholders?

Establishing a marketing strategy for the data warehousing project can assist to convey the value and benefits that will be realised to the broader business, as well as manage expectations. Howson (Wells 2005, p.5) states that marketing plans for BI applications provide ongoing dialogue with the business and "enables the team to develop solutions that are more aligned with business goals and user needs". She recommends the team to "engage sponsors and key users to promote their BI business successes in a variety of communication media – for example, internal newsletters, staff meetings, company intranet" (Wells 2005, p.5). Raising awareness of data warehousing capability can assist others in the organisation to determine how they might also benefit.

Much of the marketing effort for the UniSA edw has resided with the Business Project Manager and focused on "roadshows" to key stakeholder groups.

Whilst these were well-received, the message would have had greater impact, particularly on the senior management group, if the roadshows were delivered by a senior sponsor (in this case, the PVC: Research and Innovation). It is tempting to show stakeholders the complexities and inner workings of the data warehouse. This gives credit to the project team and demonstrates the effort required to achieve the outcomes. However, stakeholders are interested in the tangibles of the data warehouse; the things that deliver real benefits to them, for example new and better reports.

On the flipside, increasing visibility of the edw project and its outcomes may heighten the risk of bad PR. This has been the experience at UniSA. Delays in delivery and poor data quality (a result of lack of integrity in business processes around data input) have almost quashed the team's good work and associated marketing effort.

6.0 Technology Factors

The most significant technical factors influencing the edw project at UniSA can be categorised as occurring in the conceptual phase; project initiation and architectural design aspects of the project.

6.1 Conceptual Phase

While the literature advocates marketing to the business, selling the idea to IT is just as important. As earlier stated, UniSA already has a self-service management reporting environment that provides business units with an end-user view of their operational system data. To some organisations this may sound like Utopia. This environment performs the function of an ad hoc reporting environment for business areas not yet covered by the edw. It can be thought of as a number of "data marts", although we use this term somewhat liberally.

Why is there a need for an edw? The answer lies in the historic definition of a data warehouse. Bill Inmon (Ponniah 2001, p.43), considered the father of data warehousing, provides the following definition "a data warehouse is a subject oriented, integrated, non-volatile, and time variant collection of data in support of management's decisions". **Table 3** summarises the components of this definition and how, through the edw, they will deliver better reporting to UniSA.

Table 3: *UniSA Reporting Goals*

Data Type	UniSA Requirement
Subject Oriented	• Existing reporting environment is built on a siloed foundation, containing attributes of data within that particular operational system. • The corporate view is spread across numerous reporting "schemas" - not consistent, verified or integrated. • The edw is "subject oriented" – e.g., all student data needed to be consolidated into one component known as the "Student" subject and shared by all other reporting areas.
Integrated	• UniSA's siloed "data marts" operate independently with separate extracts and reporting schemas as their own entities. • Sharing of data across schemas does not exist with reporting restricted to only data within the operational system the reporting environment was built for. • The edw will provide "cross-functional" reporting.
Time Variant	• An operational system supports day to day operations. • A reporting environment built on this data contains only "current" data sets. • Data analysis often requires a view of data "as it was" in order to compare it to today's values.
Non-Volatile	• A data warehousing requirement is the non-volatility of data – e.g., the business requires data snapshots at a point in time (typically every 24 hours) and the functionality to track changes of that data as new versions of these records.

6.2 Project Initiation

"Data warehousing is really a simple concept: Take all the data you already have in the organization, clean and transform it, and then provide useful strategic information. What could be simpler than that?" (Ponniah 2001, p.39)

This quote asserts that data warehousing and the processes around it are relatively straightforward for an experienced IT unit, especially one that has developed and implemented many systems in a complex environment. Consequently it is not surprising that the technical journey at UniSA began with an air of confidence and optimism. However, it wasn't long before the realisation struck that this was a naïve view of the project.

There are considerable differences in the implementation of an edw which must be understood at the project initiation stage. This section covers the issues that had the biggest impact within UniSA.

An environment, not a product

Ponniah (2001, p.14) states that "a data warehouse is not a single software or hardware product you purchase [or develop] to provide strategic information. It is, rather, a computing environment where users can find strategic

information, an environment where users are put directly in touch with the data they need to make better decisions. It is a user-centric environment". This highlights that building a data warehouse is ultimately about creating a reporting environment. This is important to understand from the project outset and differs from a typical application development project. Requirements are not necessarily solid and tangible. The business knows what data they want to analyse but do not necessarily know what they want to report it by, and the attributes they might need. A common view within the project at UniSA is that the business would like to report on "everything by anything". The business did actively engage in refining detailed requirements for each data stage, however this statement reflects the desire for flexibility, and the fact that not even they can predict the reporting needs of the end user (nor what the Vice Chancellor is focussing on next month!).

Where to start?

UniSA has adopted an approach recommended by Ralph Kimball which is a hybrid accommodating both an enterprise perspective but implementing in small incremental stages like a data mart. The steps in this approach are:

1. Plan and define requirements at the overall corporate level.
2. Create a surrounding architecture for a complete warehouse.
3. Conform and standardize the data content.
4. Implement the data warehouse as a series of "super marts", one at a time.

The crux of this approach is to plan big however implement small, and iteratively.

As discussed in section **3.0 Stage 1: Research Activity** the project team engaged the entire University to determine UniSA's reporting requirements. These were then identified in order of priority and in the context of corporate priorities. *Research Activity* and *Student* were determined to be the first two "super marts" to be developed.

Resourcing

Underestimating the roles and responsibilities of a data warehouse project was probably our first project flaw. **Table 4** identifies what some consider "good practice" (Ponniah 2001, p.100) but in reality, only the projects with the biggest of budgets could provide for this type of resourcing.

Table 4: *Roles of a data warehousing effort.*

Executive Sponsor	Data Access Developer
Project Director	Data Quality Analyst
Project Manager	Data Warehouse Tester
User Representative Manager	Maintenance Developer
Data Warehouse Administrator	Data Provision Specialist
Organizational Change Manager	Business Analyst
Database Administrator	System Administrator
Metadata Manager	Data Migration Specialist
Business Requirements Analyst	Data Grooming Specialist
Data Warehouse Architect	Data Mart Leader
Training Leader	Web Master
Technical Writer	Data Modeler
Tools Specialist	Security Architect
Vendor Relations Specialist	

From a technical perspective, the UniSA edw project commenced with 1.5 FTE resources, a technical manager and an analyst programmer, with Database Administration services available upon request. With this in mind, the above roles had to be incorporated into available resources, both technical and business, and timelines adjusted due to the serial nature of work. As a lesson learnt, any data warehouse project commencing must carefully examine the roles mentioned above; determine the need for each (low to high priority), and ensure that there are sufficient technical resources to meet those roles critical to the success of the project.

Technical skills

“IT professionals responsible for building data warehouses need to revise their mindsets about building applications. They have to understand that a data warehouse is not a one-size-fits-all proposition; they must get a clear understanding of the extraction of data from source systems, data transformations, data staging, data warehouse architecture, infrastructure, and the various methods of information delivery” (Ponniah 2001, preface).

The project at UniSA had one mandate and that was to use the *Oracle RDBMS* as the database technology to support the edw. *Oracle* was the dominant database behind most of UniSA’s corporate and student systems and therefore was the logical choice. This also meant that the IT unit had a number of analyst/programmers skilled in *Oracle* design and development. A key learning, however, was that this experience did not necessarily mean an easy transference of skill to data warehouse design and development.

Most data warehouses, like the one at UniSA, use a dimensional modelling technique that is unique to this type of environment. Unfortunately, time required to learn modelling techniques was not considered in the initial project planning stage. Experience here is the key. Although progress was made from reading key texts and training courses, UniSA also supplemented this by engaging data warehousing experts. This approach has enabled the technical team to build understanding through the application of “real” examples from the scope of the project.

Once the design phase was achieved, the next step into new development ground was the Extraction, Transformation and Load (ETL) process. A data warehouse involves moving considerable amounts of data from source to target systems. Additionally, some complex transformations can be required when manipulating data to conform to the target schemas. There are hundreds of tools in the marketplace to assist this process however the project’s budget will have a significant effect on the options feasibly available. UniSA had the fortune of owning a number of *Developer Suite* licenses that enabled the use of *Oracle Warehouse Builder*. Whilst at the lower end of the price range this tool has proven to be, at the very least, comparable to the high end products. Obviously the introduction of a new development tool, albeit an *Oracle* one, would have had some training time overhead. Also tool, or no tool, a development framework had to be set with standards and procedures developed. These tasks at the outset must be considered from both a time and skills perspective ensuring they are factored into the overall project plan, realistically.

6.3 Architectural Phase

The architecture provides the overall framework for developing and deploying your data warehouse - it is a comprehensive blueprint. It defines the standards, measurements, general design, and support techniques for the edw. The architecture can be separated into three areas - data acquisition; data storage and information delivery.

Data acquisition

This area concerns the source systems that ultimately feed the data warehouse, and the methods with which to extract data. As mentioned previously, many of the corporate source systems at UniSA are *Oracle* based and that allowed commonly used extraction methods (via *Oracle* snapshots and *Procedural SQL* modules) to drive the data acquisition process. The main concern here, however, is not the technical extract solution, but rather the knowledge of the corporate system itself – that is, where the data is, and in what form it is stored.

In the first stage of the project the Research data was stored within a third party vendor application. Both the business and technical team members were involved in the extraction process meetings; however this approach proved to be somewhat cumbersome. Data custodians (source system owners) must be seen as critical members of the edw project team with

defined project plan tasks in order to ensure the timeliness of data to the project.

Other non *Oracle* data sources included *Excel* spreadsheet data and *SQL Server* data sources. A standard of creating .csv files for non *Oracle* data was developed and implemented in *Stage 1*. It is envisaged that as more *SQL Server* data is incorporated, a more seamless integration to SQL server will be explored.

It is worth noting that the *Oracle Warehouse Builder* tool implemented by UniSA does provide mechanisms for seamless extraction from non *Oracle* data sources including *Excel* and *SQL Server*.

Data storage

Given that an *Oracle* was the mandated data repository, the project did not need to evaluate differing technologies for data storage. The main decision was concerned with how to best support the data flows from source, to staging, through to the target warehouse.

UniSA's previously implemented siloed reporting environment had already accumulated much of the required data into schemas within an Operational Data Store (ODS). This area proved to be a useful source of data from which to determine differences between snapshots of the data without impacting the source system itself. The diagram below indicates this architecture.

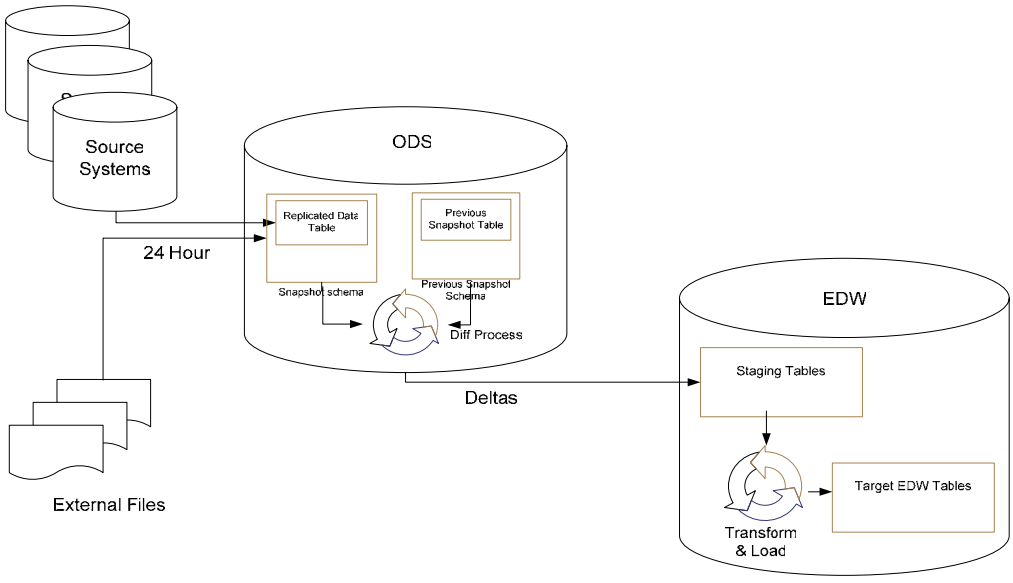

Figure 4: *Data Flow Architecture of a Data Warehouse*

Data is extracted from source systems on a 24 hour basis from both data repositories and external files (spreadsheets). This data is then placed into schemas within the ODS to support online staff and student portals. The edw has now added a process to this architecture to also store the previous night's

snapshot enabling a differences process to be undertaken. These “deltas” are then pushed through to the edw staging tables and processed by the edw Transform and Load process.

To date this has proved effective in managing the ETL process and facilitates distinct checkpoints in the process before pushing data through to the next stage.

Information delivery

Information delivery is the pay off for all the effort put into requirements, design and development. The end user experience must be in line with the ultimate goals of the data warehouse. These being:

- breadth of reporting;
- high performance;
- accuracy; and
- ease of use.

The first three points are achieved through sound requirements gathering, design and data quality processes throughout the project. The last point is a somewhat less tangible goal and can be a challenge to achieve.

Another mandate within the project was that the data warehouse support the existing *Cognos* end user tools of *Impromptu* and *Powerplay.* These tools are already in place throughout the University and considerable investment and training have already occurred. This meant that the project did not need to evaluate end user reporting tools. What it did mean for the project was that the “end user” experience was also constrained within the capabilities of these tools.

Due to its expansiveness of scope a data warehouse can result in overwhelming the user. Now a user’s ad hoc reporting includes the possibility of either reporting on current or historic data. Common data areas are now conformed for the organisation and shared across reporting areas. This can negate the effort put into simplifying data schemas and therefore seemingly end up creating inherent complexities.

Within the *Cognos* tools suite, the I*mpromptu* catalog is the report developers window into the edw. Whether it is *Cognos*, or some other reporting suite, considerable effort must be invested into configuring items, folders and relationships for ease of use.

The development of catalog standards is a must so that at the very least all data areas of the edw are consistent. Standards developed at UniSA include the use of:

- Business names for all items and folders
- Aliases for commonly used attributes across folders
- Grouping of reference data and discrete business area data

- Clear representation of current versus historic items

Another key to successful information delivery is the use of pre-calculated summaries and pre-canned reports. If possible, it is advised to push as much calculation and data derivation into the overnight ETL process. This has a dual purpose of allowing the front end reporting experience to perform well, as well as remove a component of risk in user error when implementing business calculations rules in the reporting tool. *Oracle* provides the concept of "materialized views" that allows creation of static views that can include complex calculations or commonly generated result sets based on complex queries. Incorporating a materialized view layer also allows some risk mitigation for the back end ETL process. If anything fails during the overnight refresh process, end user reporting is not affected as it is accessing these views that are static data sets. This gives the data warehouse support team crucial time in diagnosing ETL failures.

7.0 Is the UniSA edw project a success?

This paper has endeavoured to identify some critical factors – People, Business and Technology – that need to be considered to ensure a smooth implementation of a data warehousing project. So at the end of the journey, how do you know if the project has been successful?

Pohl (2006, p.4) says that a data warehouse implementation is successful when:

1. Users are constantly knocking on your door for more information that the data warehouse currently contains. This tells you the word is spreading and that stakeholders have faith in the edw as a solution to their business objectives.
2. Stakeholders are talking about the edw and it is often referred to as a potential source of data on key corporate projects.
3. The data warehouse becomes the heartbeat of the business, where decisions are made from the data intelligence it provides.

Research by Ariyachandra and Watson (2006) identified two major success constructs as indicative of a data warehouse's success: product measures and development measures. **Table 3** details the component parts of each of these.

Table 3: *Data Warehousing Success Metrics*

Success Construct	
Product Measures	
Information Quality	*The data warehouse provides accurate, complete and consistent information.*
System Quality	*The data warehouse is flexible, scalable and able to integrate data.*
Individual Impacts	*Users are able to quickly and easily access data; think about, ask questions and explore issues in new ways; and improve their decision-making because the data warehouse and BI.*
Organisational Impacts	*The data warehouse and BI meet the business requirements; facilitate the use of BI; support the establishment of strategic business objectives; enable improvements in business processes; lead to high quantifiable ROI; and improve communication and cooperation across organisational units.*
Development Measures	
Development Cost	*The cost of developing and maintaining the data warehouse is appropriate.*
Development Time	*The time to develop the initial version of the data warehouse is appropriate.*

While it is too early to say that the UniSA edw project is succeeding, many of the success factors that Pohl (2006) and Ariyachandra and Watson (2006) have identified are emerging and it is the likelihood of these that will ensure funding is secured for *Stage 2*. We are optimistic the journey traveled for *Stage 2* will be smoother, having learnt our lessons in *Stage 1* and that the edw will have the profile it deserves within the university. Watch this space!

References

Ariyachandra T & Watson H (2006). Benchamrks for BI and Data Warehousing Success. *DM Review*. Retrieved 10 January 2007 from http://www.dmreview.com/portals/portalarticle.cfm?articleId=1044330&topicId=230055

Clarry M (2001). Ten Mistakes to Avoid When Building a Data Warehouse Team. *TDWI Paper*. Retrieved 17 October 2006 from http://www.tdwi.org

Clubb G (2004). Picking Punters: Business intelligence uses disparate data to boost the bottom line. (Article by Ian Grayson, November 2) *The Australian, IT Business*: 1-6

Dresner H, Linden A, Buytendijk F, Friedman T, Strange K, Knox M & Camm M (2002). The Business Intelligence Competency Center: An Essential Business Strategy. *Gartner Strategic Analysis Report*, *R-15-2248*: 1-36. Retrieved 6 October 2003 from http://www4.gartner.com/research

Furey T (2005). Data Warehouse Project Management. *DM Review.* Retrieved 10 January from http://www.dmreview.com/portals/portalarticle.cfm?articleId=1026059&topicId=230055

Gibson, M (2005). *The Evaluation of Business Intelligence: A Case Study in a Major Financial Institution*. 16th Australasian Conference on Information Systems, 29 Nov-2 Dec

Knox M (2002). Adopting a Flexible Approach to Enterprise Architecture. *Gartner Research Note, COM 17-7431*: 1-2. Retrieved 4 October 2004 from http://www4.gartner.com/research

Negash S & Gray P (2003). Business Intelligence. *Ninth Americas Conference on Information Systems*

Pohl K (2006). Data Warehouse Project Management. *DM Review.* Retrieved 10 January 2007 from http://www.dmreview.com/portals/portalarticle.cfm?articleId=1048521&topicId=230055

Ponniah P (2001). *Data Warehousing Fundamentals: A Comprehensive Guide for IT Professionals.* Wiley & Sons Inc, London

Watson HJ, Goodhue DL & Wixom BH (2002). The benefits of data warehousing: why some organisations realise exceptional payoffs. *Information and Management, 39*: 491-502

Wells D (2005). Bringing Business and IT Together: Key Elements of an Action Plan. *What Works: Volume 20.* Retrieved 10 January 2007 from http//www.tdwi.org/Publications/WhatWorks/display.aspx?id=7733

Appendix 1: UniSA's BI System