Report and Recommendations

Decision Support Infrastructure at the University of Kansas

June 26, 2000
(Revised September 2000)

Knowledge Management Infrastructure Definition Task Force:

Cheryl Pace (chair)
Ryan Cherland
Ron Spangler
Bob Underwood
Beth Warner

Table of Contents

EXECUTIVE SUMMARY **1**

INTRODUCTION **5**

DEFINITIONS / ACRONYMS **7**

GUIDING PRINCIPLES **10**

SCOPE OF THE KU DATA WAREHOUSE **13**

OBJECTIVE 13
SCOPE STATEMENT 13
A PHASED APPROACH 14
ACCESS 14

MANAGEMENT / SUPPORT CONSIDERATIONS **15**

LIFECYCLE MANAGEMENT 15
STAFFING ROLES / FUNCTIONS / PRIORITIES 16
Management Roles *16*
Requirements Development Roles *17*
Database Development Roles *18*
Application Development Roles *20*
Other *21*
PROJECT PLAN OUTLINE AND ROLE RESPONSIBILITIES 22

TECHNICAL ARCHITECTURE **23**

TYPICAL DATA FLOW 23
METADATA DRIVEN 23
MAJOR ARCHITECTURAL ELEMENTS 25

TOOLS REQUIREMENTS **29**

DATA EXTRACTION, CLEANSING / QUALITY, AND STAGING 29
DATA WAREHOUSE LAYER 29
DATA FEED / DATA MINING / INDEXING 29
DATA MART LAYER 30
PRESENTATION, QUERY, AND ANALYSIS (END-USER APPLICATIONS) 30
METADATA REPOSITORY LAYER 30
WAREHOUSE MANAGEMENT LAYER 30
APPLICATION MESSAGING TRANSPORT LAYER 31

RECOMMENDATIONS **32**

FIGURE 1: THE BUSINESS DIMENSIONAL LIFECYCLE 15
FIGURE 2: HIGH LEVEL WAREHOUSE TECHNICAL ARCHITECTURE 23
FIGURE 3: KU ENTERPRISE DATA ARCHITECTURE 26

TABLE 1: STAFFING ROLES AND FUNCTIONS 34

APPENDIX A: Knowledge Management Infrastructure Definition Task Force

APPENDIX B: Sample Project Plan Outline and Role Responsibilities

APPENDIX C: Dublin Core Metadata Element Set

BIBLIOGRAPHIC RESOURCES

Executive Summary

Like many, the University of Kansas has depended for many years on a variety of mainframe-based and stand-alone systems for daily operations and access to historical information. In the mid-90s, the University moved to integrate administrative information systems using the PeopleSoft software suite. During implementation, there was a growing awareness of the need to preserve data of historical importance from the legacy systems as well as provide easy access to current information resources.

Several projects have helped better define a coordinated information management strategy for the University as well as responded to immediate access and reporting demands. As user expectations and the demands on existing systems and resources have grown, it is evident that the University must clearly recognize its various data resources as institutional assets and aspire to the next level of coordinated management of these critical information resources.

To better define the basic infrastructure for such a resource management program, the Knowledge Management Infrastructure Definition Task Force was created in February 2000 and charged with:

- Defining the guiding principles of building and maintaining the KU KM infrastructure and outlining the business rules involved
- Using the Orr "Enterprise Data Architecture" model as appropriate, define the elements of the technical architecture to be used for the KU KM infrastructure, including
  - Database architecture / data repository types
  - Data model outline(s)
  - Metadata requirements / repositories
  - Management requirements / routines
- Defining the requirements for
  - Data extraction, cleansing and staging tools
  - Data warehouse builder tools
  - Database(s)
  - Presentation, query, and analysis tools

During its discussions, the Task Force developed several fundamental tenets for developing a robust Decision Support Infrastructure for the University. Chief among these are the following.

Informed decision making and knowledge sharing are critical to an institution's success. Providing an enterprise-wide view of data on which to base future directions can help achieve that success. The Task Force has established ten *Guiding Principles* for building and maintaining a University decision support infrastructure:

1) Improve stakeholder satisfaction by increasing the effectiveness and efficiency of University decision support processes
2) Align decision support efforts with the University's business processes
3) Identify and prioritize subject areas for University data
4) Separate reporting from production
5) Define the scope of the data included

6) Establish the architecture of the data warehouse
7) Provide universal access / interface
8) Incorporate tools
9) Emphasize quality
10) Provide a management focus

University Data are institutional assets and are held by the University to support its fundamental instructional, research, and public service missions. While a complete Decision Support Infrastructure will eventually include additional components, the first priority should be the implementation of a robust data warehouse environment to consolidate and preserve these assets.

The data warehouse is not intended to take the place of existing transactional systems, but to complement them. As such, the data warehouse should contain only the mission-critical data needed to support management decision-making functions. At a minimum, the KU data warehouse will provide an integrated, subject-oriented set of databases for tactical and strategic information retrieval based on relevant data from key transactional systems.

The Task Force has identified the primary categories, or Subject Areas, of critical data which are defined as Students, Financial, Human Resources, Facilities / Equipment, Research, and Service. Data from historical, existing, or future transaction systems should be defined and grouped by these Subject Areas.

Because of the volume of data, level of user community expectations, and limited personnel resources available, it is critical for University leadership to balance the scope and timeline of the initial project(s) against the available resources. As a first step, it is necessary to prioritize the Subject Areas to provide practical direction for the program. In addition, a prioritized, phased approach to improving access, i.e. reports and reporting tools, is important. If the process is managed properly, the benefits will outweigh the time and costs involved.

Staff resources will be critical to a successful implementation effort. There are many roles in a data warehouse project – and there is seldom a one-to-one relationship between data warehouse roles and individual people. The size of the data warehouse project dictates the number of team members needed at various phases of development. Although the number of people on the project may vary, the roles filled remain constant for any data-warehousing project. Each of these roles is critical to the success of the project.

Of equal importance is the establishment of a basic architecture for the data warehouse and the setting and enforcement of standards for working within that architecture. The Task Force developed a modified version of Ken Orr's Enterprise Data Architecture™ to identify the critical components, systems, and resources for KU's Enterprise Data Architecture. The Task Force also drew heavily on a number of other reference sources to further define the elements comprising the Enterprise Data Architecture, staffing role descriptions, and project steps.

This report is the result of the Task Force's discussions on defining the basic infrastructure needed to support a successful decision support effort for the University of Kansas. It is intended as the starting point for a broader University effort and should act as a guide for the definition and implementation of the initial project(s) in data warehousing. Task Force **Recommendations** for defining and implementing such an effort are as follows:

1) Adopt the *Guiding Principles* statement as the basis for any central Decision Support project for the University.
2) Adopt the KU Enterprise Data Architecture model.
    a. Recognize this as a University-wide, cooperative venture with concurrent, parallel efforts occurring on the Lawrence and KUMC campuses with joint coordination between these efforts.
    b. Adopt a phased implementation approach with centrally prioritized tasks to be addressed.
3) Begin the Project Planning and Business Requirements Definition steps of the Business Dimensional Lifecycle for the University's Data Warehousing initiative. Some of these steps are:
    a. Adopt the Scope statement
        i) Approve the definition of University Data
        ii) Approve the criteria for data to be included in the data warehouse
    b. Identify the initial problem statements which provide the business case for the data warehousing initiative
    c. Identify and prioritize projects which address these problem statements.
4) Based on the KU Enterprise Data Architecture Model and Staffing Roles requirements, evaluate the following to determine what is already in place, under development or missing:
    a. Relevant systems
        i) Identify existing operational systems that contain data that needs to be in the data warehouse
        ii) Identify existing non-operational systems that contain data tha needs to be in the data warehouse
        iii) Identify existing external systems that contain data that needs to be in the data warehouse
        iv) Is there any data needed for which there is not a source?
    b. Staff
        i) Identify available staff to fill the roles in the table below
        ii) People can fill multiple roles
        iii) A role may be filled by more than one person
        iv) Not all roles are needed during all phases
        v) Are there any roles that can't be filled by existing staff?
    c. Tools
        i) Identify those tools that the university owns (this includes servers, databases, software, etc.)
        ii) Descriptions of the tools needed can be found in the section on *Tools Requirements*
        iii) Are there any toos we don't have?
    d. Projects

i) What projects currently exist that either affect or may be affected by the data warehouse initiative?
ii) Determine whether a project needs to be realigned with the data warehouse initiative
iii) Make the necessary changes in the project.

5) Based on the results of Recommendations #3 and #4, charge a task force/leadership group to
   a. Select the first project for inclusion in the data warehouse
   b. Define scope of the project
   c. Define existing systems that are a part of this project
   d. Identify the dtat in these systems that belongs in the data warehouse
   e. Continue with development following the Business Dimensional Lifecycle and the KU Enterprise Data Architecture model.
6) Based on the outcome of the Project Planning and Business Requirements Definition phases, prioritize the staffing requirements for implementation and recognize that several roles will require full-time, dedicated staff.
7) Adopt and promulgate policies and procedures directed toward ensuring the quality of data in the decision support system (Guiding Principle #9).
8) Adopt and promulgate policies and procedures directed toward ensuring stable and secure access to the data warehouse for appropriate University personnel (Guiding Principle #7).
9) Define a decision-making and prioritization mechanism for coordinating and resolving issues such as non-standard project implementations, task and resource prioritization, metadata definition conflicts, etc.
10) In order to facilitate University awareness and continued close cooperation, implement various education and communications mechanisms for disseminating information about the KU Decision Support Infrastructure including:
    a. Reconvene the Information Management Planning Retreat participants to present implementation plans and gather feedback.
    b. Establish a website for centralized dissemination of Decision Support Infrastructure documents and news.
    c. Establish a discussion group (e.g. Listserv) for proactive information dissemination and discussion.

Introduction

Data – Information – Knowledge. Organizing resources along this progression is today's equivalent of scaling Mount Everest for the typical organization: expensive, labor intensive, requiring a coordinated team effort with a shared vision, fraught with potential danger but – what a view from the top!

During the '60s, the introduction of computers held out the promise of regaining control over the information of increasingly complex and distributed organizations. Over four decades later, that promise has still been only partially realized. As computing systems have evolved, changes have been driven more by technological advances than by a well designed, coordinated management-objectives approach, often resulting in a distributed computing environment with dis-integrated information resources. Like many organizations, the University of Kansas has experienced these situations.

Like many, the University of Kansas depended on a variety of mainframe-based and stand-alone systems for daily operations as well as access to historical information. In addition to systems for administrative data, a number of other systems for managing space, faculty data, student information, etc. were also developed. In the mid-90s, the University made a series of strategic decisions to integrate administrative information systems using the PeopleSoft software suite. As those decisions were implemented and new systems put in place, there was a growing awareness of the need to preserve data of historical importance from the legacy systems as well as provide easy access to current information resources.

Several steps have since been taken to attempt to better define a coordinated information management strategy for the University as well as respond to immediate access and reporting demands. These have generally been well-received efforts that have helped define the organizational needs and issues of a University-level information management program. However, as user expectations and the demands on existing systems and resources have grown, it has become clear that the University must aspire to the next level of coordinated management of its critical information resources.

In February 1999, an Information Management Planning Retreat was held that included major stakeholders from across the University to begin to build a shared vision of what that next level should be. During that retreat, a number of issues were identified as needing to be addressed in a coordinated fashion. These issues included:

- The timeliness of current data for tactical decision-making is critical
- Data requirements across the campus are complex and not always readily addressed by pre-defined 'printed' reports
- Analytical and reporting resources are limited
- Data is often duplicated across systems and not well integrated; there is no single place to go for all data collections, making it difficult to know exactly what is collected where
- Data definitions differ between systems, change over time, and are not well documented
- Data quality and integrity is not well controlled

- Users need education to understand what data is available, what it means, and how to legitimately use it
- Long-term archival procedures are not well defined which is threatening loss of historical data resources
- Current staffing resources are inadequate

As a result of the retreat, a consensus was reached that the solution to these issues was to create a coordinated information management program for the University and the University Steering Committee for Information Management (USCIM) was formed. The USCIM determined that the first step in implementing this program would be the creation of a robust data warehouse system. Further discussions led to the realization that while there was a shared view of the problems, there was not a consensus on what shape the data warehouse should assume. To better define the basic architecture for the program, the USCIM created the Knowledge Management Infrastructure Definition Task Force (KMIDTF) in February 2000 and charged it with:

- *Defining the guiding principles of building and maintaining the KU KM infrastructure and outlining the business rules involved*
- *Using the Orr "Enterprise Data Architecture" model as appropriate, defining the elements of the technical architecture to be used for the KU KM infrastructure, including*
  - *Database architecture / data repository types*
  - *Data model outline(s)*
  - *Metadata requirements / repositories*
  - *Management requirements / routines*
- *Defining the requirements for*
  - *Data extraction, cleansing and staging tools*
  - *Data warehouse builder tools*
  - *Database(s)*
  - *Presentation, query, and analysis tools*

(The full charge can be found in Appendix A.)

This report is the result of the Task Force's discussions on defining the basic infrastructure needed to support a successful knowledge management effort for the University of Kansas. It is intended as the starting point for a broader University effort and should act as a guide for the definition and implementation of the initial project(s) in data warehousing.

Definitions / Acronyms

Before going further, it is important to have a shared understanding of various information and data warehouse concepts. Definitions and acronym expansions for the most critical terms / concepts as used in this report are:

AAU – American Association of Universities

AAUDE – American Association of Universities Data Exchange

AMCAS – American Medical College Admission System

Architecture
The goal of the data warehouse architecture is to provide a single, integrated data warehouse meeting business information needs. It establishes the technical framework and procedures for the data warehouse at an enterprise level. Components of the architecture include: source data layer; data staging and quality layer; core DW layer; data feed / mining / indexing layer; data mart layer; presentation / desktop access layer; metadata repository layer; warehouse management layer; and application messaging transport layer.

CRINC FS – Center for Research Financials (Web-based reporting via DEMIS)

Data mart
A data mart is a subset of highly summarized data from the data warehouse optimized for end-user access to facilitate data analysis and designed to provide answers to a distinct set of questions.

Data steward(s)
The University Executive Officer(s), or their delegate(s), responsible for defining, setting, and enforcing policy for University Data.

Data stewardship
The process of defining, setting, and enforcing policy for University Data.

Data warehouse (DW)
Data warehousing is the process of creating an architected information management solution to enable analytical and information processing at the enterprise level. A data warehouse is a complete and consistent store of data obtained from a variety of sources and made available to end users in a way they can understand and use in a business context.

DBA – Database administration / administrator

DEMIS – Departmental Executive Management Information System

E/R – Entity Relationship

ERD – Entity Relationship Diagram

ETL – Extract, Transformation, Loading

FSKU – Financials System (KU-Lawrence, PeopleSoft database, ver. 6.10)

HRIS – Human Resources/Payroll Information System (KU-Lawrence, PeopleSoft database, ver. 4.1)

HRSA – Human Resources/Payroll (KU-Lawrence) and Student Administration system (KU, PeopleSoft ver. 7.5)

Information
Information is an organized collection of facts in a meaningful context.

Infrastructure
There is both an operational and a physical infrastructure that support the data warehouse architecture. The operational infrastructure provides the policies, procedures, roles, responsibilities and system software that define and guide the use, management and configuration of the data warehouse. The physical infrastructure includes the hardware, network, operating system, and other core physical components.

IPEDS – Integrated Postsecondary Education Data System

Knowledge
Knowledge is the ability to process data and information and turn it into effective action.

Knowledge management (KM)
KM is a conscious strategy of getting the right knowledge to the right people at the right time. It promotes a collaborative and integrated approach to the creation, capture, organization, access and use of an enterprise's information assets. KM has two primary business aspects: 1) treating the knowledge component of business activities as an explicit concern of business reflected in strategy, policy, and practice at all levels of the organization, and 2) making a direct connection between an organization's intellectual assets and positive business results.

KUFS – KU Financial System (Web-based reporting via DEMIS)

KUMC Fin – KU Medical Center Financial database

KUMCFS – KUMC Financial System (Web-based reporting via DEMIS)

MANOVA – Multivariate Analysis of Variance

MCHR – Med Center Human Resources/Payroll system (PeopleSoft database, ver. 4.1 and ver. 7.5)

Metadata
Metadata describes the meaning and structure of data, as well as how it is created, accessed and used. There are three basic types of metadata: 1) business metadata describes the business definition and rules of the data, 2) process metadata describes where, when and how data is placed into the warehouse, and 3) technical metadata describes the physical location, format, and structure of data resource.

ODBC – Open Data Base Connectivity

Operational data store (ODS)
The ODS is a subject-oriented, integrated, updateable collection of data designed and structured to allow better integration among operational data used for immediate inquiry and reporting. It reflects, at any point, the current operational state of its subject matter and enhances performance by removing the processing load from the transactional system.

OLAP – On-Line Analytical Processing

OLTP – On-Line Transaction Processing

QA – Quality Assurance

SCH MART – Student Credit Hour reporting data mart

SRIS – Student Records Information System

Star Schema
Star schema is a description of the physical database design supporting OLAP needs within the data warehouse. The database tables are arranged around a central fact table and related dimensions radially distributed around the fact table.

TQM – Total Quality Management

University Data
Data that satisfies the *Guiding Principles* and/or meets the following criteria:

- It is relevant to the planning, managing, operating, controlling, or auditing administrative functions of an administrative or academic unit
- It is generally referenced or required for use by more than one organizational unit
- It is shared among primary information system applications across a campus of the University; the compatibility of such data is a key requirement to integrating other data from these applications
- It is included in an official university administrative report
- It is used to derive an element that meets the criteria above

USCIM – University Steering Committee for Information Management

XML – Extensible Markup Language

Guiding Principles

Informed decision making and knowledge sharing are critical to an enterprise's success. Providing an enterprise-wide view of data to base future directions on can help achieve this. Quality of data and alignment with business goals are also keys to success. Gartner Group states that most re-engineering initiatives will fail because of lack of attention to information quality in the data, in the architecture, and in the business processes that produce or consume the data. In order to prevent this failure, the Task Force has established ten guiding principles for building and maintaining the University of Kansas Decision Support Infrastructure and Architecture. These Principles are:

1) Improve stakeholder satisfaction by increasing the effectiveness and efficiency of University decision support processes
2) Align decision support efforts with the University's business processes
3) Identify and prioritize subject areas for University data
4) Separate reporting from production
5) Define the scope of the data included
6) Establish the architecture of the data warehouse
7) Provide universal access / interface
8) Incorporate tools
9) Emphasize quality
10) Provide a management focus

Detailed Principles Statements:

1) Improve stakeholder satisfaction by increasing the effectiveness and efficiency of decision support University processes

a. Reduce or eliminate duplication of *efforts* to collect, verify, store, and maintain data used by multiple reporting systems.
b. Create a knowledge workplace that empowers all users to be their own analyst.
c. Provide information to the user at the time they need it to make a business decision.
d. Operational data will be transformed into decision support data.

2) Align decision support efforts with the University's business processes

a. The primary dimensions will be those of the University's business.
b. The Data Warehouse will be adaptive to changing environments - policies, information needs, and technology

3) Identify and prioritize Subject Areas for University Data

a. At a very global level, it is essential to identify the primary categories of data, which are important to the University.
b. These will be called subject areas and will be organized around functional or cross-functional subject areas.
c. Because of the volume of data, it will also be important to prioritize the subject areas to provide practical direction.

4) Separate reporting from production

a. The data warehouse will be a physically separate system in order to protect production transaction systems from the potential impact of numerous or complex ad hoc queries.
b. The data warehouse will integrate data from production transactional systems into commonly defined “reporting-friendly” table structures.

5) Define the scope of data included

a. The data warehouse is where we assemble and publish data, which originates in University transaction systems.
b. The data warehouse will also provide access to data not found in transaction systems, including summary data, historical data, metadata, and external data.
c. The data warehouse is a series of data layers, much like geologic layers, made up of static snapshots of the Online Transaction Processing systems over time.
d. The data warehouse will include a metadata repository that contains all the information about the data and processes used to populate and access its contents.

6) Establish the architecture of the data warehouse

a. The data warehouse will be open, extendable and designed for reporting and presentation.
b. Data in the warehouse can be separated and combined by all primary measures for the University (the classic 'slice and dice' requirement).
c. The data warehouse will be designed using a structure to facilitate drilling either *down* or *across* subject areas.

7) Provide universal access / interface

a. Stable and secure access to the data warehouse will be provided to all University personnel that have a business need for it.
b. Staff will be able to connect to the data warehouse through a World Wide Web browser.
c. The data warehouse will provide the ability to extract data into common file formats for end user access.

8) Incorporate tools

a. The data warehouse is not just data, but also a set of tools to *query, analyze,* and *present* information. It will include relational database and analytical software.
b. Tools will be very easy to use.
c. Queries or reports can be run with default options with minimal button clicks.
d. Simple queries or reports will run in less than five seconds.

9) Emphasize quality

a. Quality data is critical to University operations; poor quality data will not be published in the data warehouse.
b. Data in the warehouse will be consistent.
c. Definitions will be available for all publishable elements in the data warehouse and will document changes over time.
d. Use standard processes as a tool to reduce costs from poor information quality.
e. Provide continuous process improvement in all systems across the University.
f. Use scientific methods to influence outcome quality and measurement that proves we are on track.

10) Provide a management focus

a. Define a mechanism for arbitration of conflicting data definitions.
b. Provide "context-sensitive" training to users of the data warehouse.
c. Encourage informal training avenues such as user groups, online discussion groups, etc. managed by the functional areas.
d. Encourage practical data stewardship policies and practices with a “One University” view.

Scope of the KU Data Warehouse

Objective

A primary focus of today's information environment is to move new technologies and strategic decision making into the hands of the decision makers without compromising the integrity, security, or performance of day-to-day operational processing. The ultimate objective of the KU data warehouse is to give the University community direct access to integrated information from multiple sources that has been cleansed and stored in one location and made available for University planning and decision making.

Scope Statement

University Data are institutional assets and are held by the University to support its fundamental instructional, research, and public service missions. While all data captured about university assets through its various transactional systems are resources of the University, not all data are appropriate for the data warehouse. The data warehouse is not intended to take the place of existing transactional systems, but to complement them. As such, the data warehouse should contain only the mission-critical data needed to support management decision-making functions. A data element should be considered University Data, and therefore part of the data warehouse, if it satisfies one or more of the definitions found in the *Guiding Principles* and / or meets the following criteria:

- It is relevant to the planning, managing, operating, controlling, or auditing administrative functions of an administrative or academic unit
- It is generally referenced or required for use by more than one organizational unit. Data elements used internally by a single department or office typically are not considered University Data
- It is shared among primary information system applications across a campus of the University; the compatibility of such data is a key requirement to integrating other data from these applications
- It is included in an official university administrative report
- It is used to derive an element that meets the criteria above

As an aid to defining the boundaries of the initial data warehouse, it is essential to identify the primary categories of data, which are important to the University. These categories will be called Subject Areas and will initially be defined as:

- Students (enrollment, courses, financial aid)
- Financial
- Human resources (faculty, staff, student employees)
- Facilities / Equipment
- Research
- Service

Data from historical, existing, or future online transaction systems will be segmented and grouped by these subject areas. At a minimum, the KU data warehouse will

provide an integrated, subject-oriented set of databases for tactical and strategic information retrieval based on relevant data from the following transactional systems:

- PeopleSoft administrative systems
    - Budget
    - Financials
    - Human Resources
    - Student
- SRIS
- Faculty personnel system(s)
- Equipment, facilities, and space systems
- Academic data systems (i.e. School of Business faculty and student systems)
- Prospective student systems

Additional systems will be investigated for inclusion as the project progresses.

Data from external and non-operational systems may also be included in the KU data warehouse if they meet the criteria above and fall into one of the defined subject areas.

The KU data warehouse will be structured to support management analysis of university performance indicators by providing summary and historical data.

A Phased Approach

Because of the volume of data and limited personnel resources available, it will be critical for University leadership to prioritize the Subject Areas to provide practical direction for the program. In addition, a prioritized, phased approach to improving access, i.e. reports and reporting tools, is important.

Access

The collection of integrated institutional data in the KU data warehouse can be accessed by authorized KU administrators, business managers, service providers and institutional researchers for the purpose of performing analysis, producing ad hoc queries and reports, and maintaining data subsets as a function of fulfilling their official duties with the University. Access rights and limitations will be defined by University policy. Relevant University policy, and state and federal laws govern access by non-University personnel.

Management / Support Considerations

While a complete Decision Support Infrastructure will eventually include additional components, the first priority should be the implementation of a robust data warehouse environment. Building the data warehouse takes a great deal of effort to plan, design, develop, and implement properly. It is a labor-intensive, iterative process that takes experienced individuals to satisfy evolving institutional needs. The time and cost involved must be evaluated against the potential business benefits of having data immediately available to key decision makers. As with any project, it is critical to balance the scope and timeline against the available resources. If the process is managed properly, the benefits will outweigh the time and costs.

Lifecycle Management

The Business Dimensional Lifecycle diagram below illustrates the general flow that occurs during a data warehouse implementation. It identifies high-level task sequencing and highlights the activities that should be happening concurrently. It does not attempt to show a timeline for any given project – the focus is on sequencing and concurrency, not absolute timelines.

The Lifecycle model is most effective when used to implement projects of a manageable scope. It is nearly impossible to tackle everything at once. However, it is extremely possible to design, develop and deploy a data warehouse environment through multiple iterations using the Lifecycle model.

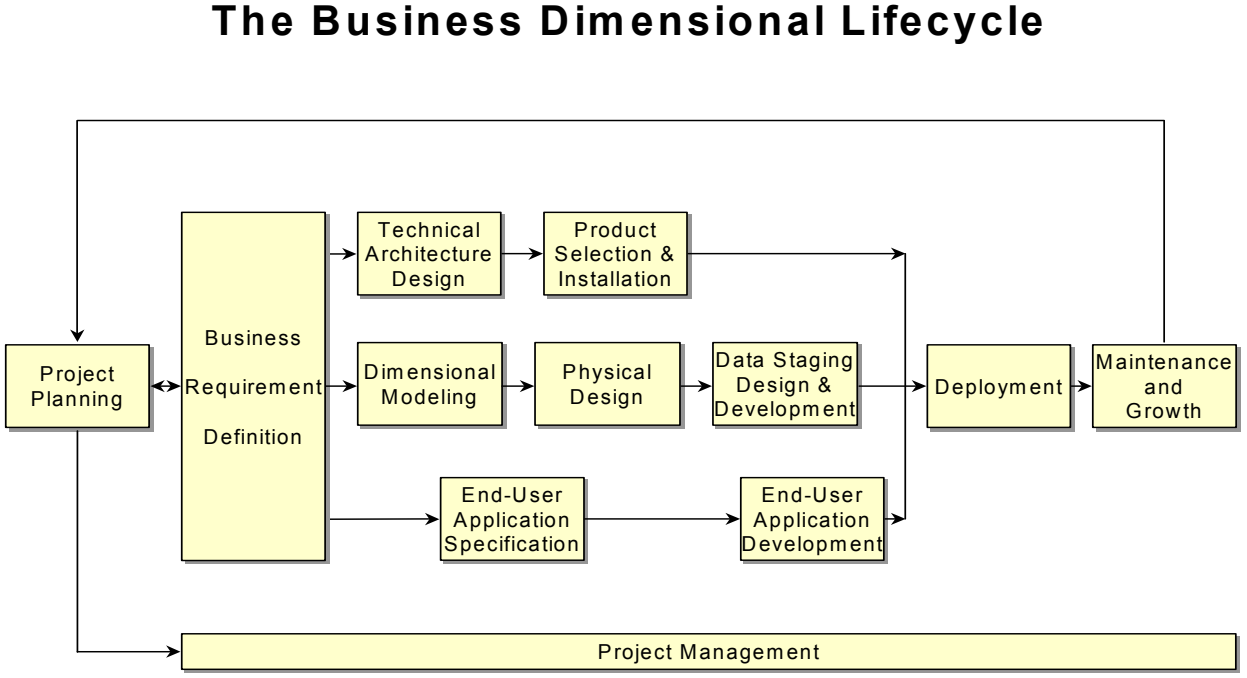


Figure 1: The Business Dimensional Lifecycle

Kimball, Ralph, et al, The Data Warehouse Lifecycle Toolkit, New York, NY:Wiley, 1998, pg. 33

Staffing Roles / Functions / Priorities

There are many roles in a data warehouse project – and there is seldom a one-to-one relationship between data warehouse roles and individual people. These roles fall into several broad categories including: *Management Roles, Requirements Development Roles, Database Development Roles, Application Development Roles,* and *Other*. The team must have a balance of skills across a variety of roles with individual staff members wearing several hats at once.

The size of the data warehouse project dictates the number of team members needed at various phases of development. Although the number of people on the project may vary, the roles filled remain constant for any data warehousing project. It is important to focus initially on the roles to be assigned before choosing people to fill them. It is also important to recognize that each of these roles is critical to the success of the project.

Management Roles

User sponsorship

User sponsors are the business owners of the project. Their role is to help make and then support key project scope decisions. Some organizations have a user steering team to make decisions for the user community at large. The user sponsor(s) will work closely with the technical sponsor to gain support for the project and recommend priorities.

Technical sponsorship

The IT executive sponsor role is to work closely with the user sponsor(s) to ensure a joint success. The technical sponsor also works closely with the program manager to assist with dispute resolution, especially when the disputes pertain to resources. Primary responsibilities of the technical sponsor include:

- educating top management on the application and impacts of decision support system(s) / data warehouses
- gaining economic support for the project
- identifying and prioritizing applications
- defining budgets and schedules
- working with the project manager to keep the project on track
- monitoring industry trends and identifying emerging technologies that should be adopted
- participating in user groups and industry associations

Project management

The project manager is responsible for day-to-day management of project tasks and activities including resource coordination, status tracking, and communication of project progress and issues, working closely with the user project lead. Project managers should possess a broad knowledge of technology and system development in order to comprehend the full lifecycle. The project manager role is responsible for balancing the mechanics of project management

with the needs of the people working on the project. The project manager responsibilities include:

- create, manage, and adjust project plans
- define overall architecture and set standards
- evaluate and select hardware platforms
- evaluate and select networking facilities
- evaluate and select middleware
- document test plans

Technical lead

The role of technical lead is to ensure that all the assignments fit together and lead toward a successful implementation. This role tracks details, issues, product problems, interface issues, etc. through to implementation and deployment.

User project lead(s)

The user project lead(s) work with the project manager on a day-to-day basis, jointly monitoring project progress and communicating to the rest of the organization. The user project lead(s) should have a solid understanding of the business requirements for any project in which they are involved.

Requirements Development Roles

User requirements analysis

The user requirements analyst is responsible for leading the user requirements definition activities and then representing those requirements as the dimensional model (table schema) is developed. The user requirements analyst role may be held by either an IT resource who is extremely user-centric and knowledgeable about the organization, or by a user resource with a solid technical foundation. The project manager or the user project lead, especially for smaller projects, may hold this role.

During the scope phase of the project, the user requirements analyst's role is to collect, consolidate, organize, and prioritize the business needs and problems the user community presents. The objective is to create a set of requirements, which ensure that the data warehouse accomplishes its original intent when it's deployed. The role of the user requirements analyst is to ensure that the data warehouse is a success in the business sense, not just a technical success.

Technical / security architecture

This role is responsible for the design of the technical infrastructure and security strategy to support the data warehouse providing the overall cohesiveness to ensure that the components will fit together. This role must represent the special needs and risks of the decision support system to the rest of IT. Other responsibilities can include:

- administering user access and security
- monitoring security and responding to incidents

Data stewardship

This role may be assigned to specific senior organization officials, or delegated data stewards, who have planning and policy-level responsibilities for data in their functional areas. The Data Stewards, as a group, are responsible for recommending policies, and establishing procedures and guidelines for organization-wide data administration activities. Data Stewards, as individuals, have management responsibilities for defined segments of the institutional database.

The Data Steward(s) is / are responsible for gaining organizational agreement on common definitions for conformed warehouse dimensions and facts, then publishing and reinforcing these definitions. Other primary responsibilities can include:

- metadata modeling
- identifying data sources
- monitoring data quality
- searching out causes of incompatibility between the warehouse and other systems
- assisting users in finding the correct information
- verifying that the data in the warehouse and reports produced by the warehouse are in agreement with financial reports and other enterprise information

Metadata modeling

Metadata modeling is the process of determining the metadata element requirements and their collection, organization, and maintenance processes. This role is involved in developing / acquiring the warehouse's metadata management system and information directories.

Quality assurance analysis

The quality assurance (QA) analyst ensures that the data loaded into the warehouse is accurate. This role identifies potential data errors and resolves them. This role has a significant workload during the initial data load to ensure that the data staging process is accurate and continues throughout the life of the project and a data changes. The QA analyst must determine whether variables and parameters in the tools are used correctly, how exception handling will work when data errors occur, and all QA tasks necessary for application development.

Database Development Roles

Data modeling and conceptual / logical database design

The data modeler is responsible for performing detailed data analysis and developing the dimensional data model. Knowledge about existing organizational data models is extremely valuable, as is a good understanding of the business rules to aid in meeting user requirements. This role is closely aligned with the data warehousing architect role, and may be responsible for developing an overall data architecture strategy.

Source data analysis

The source data analyst is responsible for reviewing data in source and legacy systems prior to the extraction and transformation process to determine what is really in the systems and what discrepancies may exist.

Decision support system architect

The role of chief architect is not the same as the role of the technical leader. The architect performs architectural functions, based on the convergence of user needs, the state of the art of data warehousing, and computing technology, and an organization's internal standards and guidelines. The architect has to take a broad look at how technology and products can be used to meet user objectives or solve business problems. An architect should look at a number of different approaches and architectures, then recommend one that's most sensible for the environment. Primary responsibilities can include:

- defining overall architectures and setting standards
- evaluating and selecting hardware platforms
- evaluating and selecting networking facilities
- evaluating and selecting database management software
- evaluating data transformation and cleansing software
- evaluating middleware
- evaluating reporting, query, analysis and mining software
- educating users on data warehousing capabilities
- monitoring industry trends and identifying emerging technologies that should be adopted
- designing closed-loop applications (where warehouse data and analyses are used in transaction processing systems)

Database administration and physical database design

The function of applying formal guidelines and tools to manage the organization's information resource is referred to as database administration. The DBA translates the conceptual / dimensional model into physical table structures. The DBA also determines initial aggregation, partitioning, and indexing strategies. The DBA is often responsible for day-to-day operational support of the database, ensuring data integrity, database availability, and performance. This role can be split into design and production roles. The DBA typically performs these tasks:

- creates the initial database schema (physical structure, e.g. Star Schema)
- evaluates and selects database management software
- modifies the database schema as required throughout development
- runs load scripts to handle initial population of the database with either test data or real data, and runs scripts to reload the database with new data (the data warehouse restocking processes)
- monitors query and database performance, and query repetitions
- tunes the database for performance by analyzing where response-time problems occur and how the database structure can be modified to make it run faster
- performs backup and restore operations as necessary

- administers user access and security
- monitors database capacity
- creates proactive monitoring and preventive action systems to avoid outages

Data staging system design / data movement and middleware specialist (ETL)
The data staging system designer is responsible for the end-to-end design of the production process to extract, transform, and load the data in preparation for the data warehouse. Primary responsibilities can include:

- evaluating and selecting data transformation / cleansing software and middleware
- developing methods of ensuring incompatibilities are systematically eliminated
- scheduling data updates and report generation
- developing monitoring systems to ensure data has been updated as expected
- developing and / or acquiring metadata management systems and information directories
- developing Internet and Intranet delivery applications
- testing and correcting applications
- designing closed-loop applications

Data staging programming
Programmers are needed to construct and automate the data staging extract, transformation, and load processes under the direction of the data staging system designer. Primary responsibilities can include:

- developing data cleansing programs and implementing acquired data re-engineering software
- identifying data sources
- programming data acquisition and transformation processes
- developing and documenting test plans
- automating the load process
- maintaining and updating acquisition and loading
- developing monitoring systems to ensure data has been updated as expected
- searching for causes of incompatibility between the warehouse and other systems
- coding and testing closed-loop applications

Application Development Roles

End user applications development / Front-end tools / Presentation system design
Equally as important as the database are the desktop tools used for reporting and querying, online analytical processing, or data mining. This role creates and maintains these types of end user applications, typically using off-the-shelf data

access software. The end user application developer is also responsible for loading the data access tool metadata. The tools specialist must be able to:

- determine which of several different implementation strategies makes the most sense in a specific environment and why
- evaluate reporting, query, analysis, and mining software
- follow design and specification guidelines to ensure that whatever is implemented is correct
- develop, test, and document applications
- convert existing reporting applications to use the data warehouse
- develop decision support applications
- develop new periodic reporting applications

Data warehouse education

End users must be educated on the system capabilities, data content, the prebuilt end user applications, and the data access tools. This role typically develops the initial education course materials, as well as delivers the education on an ongoing basis. Responsibility for user education can be split between the IT and user communities – an IT resource teaches tool mechanics and the user representative covers the data and end user application topics.

Technical support (software & hardware)

Specialists in these roles may be focused on mainframe systems, client / server systems, networking, etc. These specialists are involved in early stages of the warehouse to perform resource and capacity planning. During product selection, they ensure compatibility with the existing technical environment. Once technology has been selected, they are involved in the installation and configuration of the new components. They also provide ongoing production support.

Help desk

Personnel in this role are knowledgeable in the basics of various end user tools, data update schedules, metadata systems, etc. and are able to respond to or refer user questions / problem reports as appropriate.

Data warehouse auditing

Primary responsibilities include:

- auditing the processes used to acquire and transform the data
- verifying that data in the warehouse and reports produced by the warehouse are in agreement with financial reports and other enterprise information assumed to be accurate
- monitoring data warehouse quality

Other

Legal and policy issues for accessing and using data

Project Plan Outline and Role Responsibilities

For a more detailed view of a sample project plan with responsibilities assigned to specific roles, see Appendix B.

Technical Architecture

Development of an infrastructure for knowledge management or decision support will require the integration of information technologies, business procedures, and business data. The technical architecture provides the overall road map for accomplishing this task. As noted earlier in the Business Dimensional Lifecycle, implementation is an on-going iterative task. In addition to business requirements, two other factors need to be considered: the current technical environment and planned strategic technical directions. This section outlines the typical data flow, metadata, the architecture model and its major elements including both data stores and processes.

Typical Data Flow

The most straightforward explanation of the data flow is that Data flows from the source systems (HRIS, FSKU, etc.) through a data staging area then to the data warehouse where users can access it through various query services. This simple concept is illustrated by figure 8.2 from The Data Warehouse Lifecycle Toolkit, 1998, by Ralph Kimball, a recognized expert in the field.

High Level Warehouse Technical Architecture

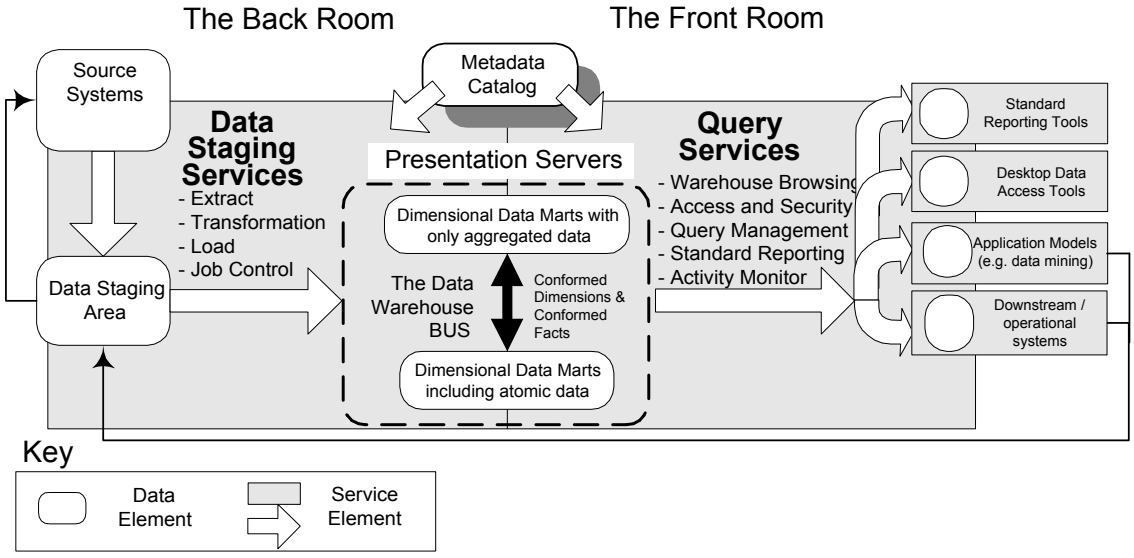


Figure 2: High Level Warehouse Technical Architecture

Metadata Driven

When the data, '3.75', comes across the screen, how do you know what it means? Is it a student's semester grade point average (GPA) or their cumulative GPA? Or is it the number of FTE positions in a small department? Or a purchase amount?

Obviously, a name or label for the field would help greatly. Sometimes though, there are several closely related fields, such as GPA. If the label is not enough to

distinguish the difference or if the field name is not familiar, the next step is finding a definition for the field. Although the name and the definition may satisfy some information needs, many other situations require more. Where did this data come from? When was it extracted? Has it since been updated? Is it the most current data available? What reports use this data? If it reflects a partial year, how is it annualized? The answers to these questions are provided by additional data about the field, or data about data, also known as metadata.

Corporate Information Designs, Inc provides one of the most succinct explanations. "Nearly every data warehousing expert will agree that metadata management is the key to data warehousing success. A metadata repository contains the organization's business models and rules, data view definitions, data usage model, report dictionary, user profiles, conceptual, logical, and physical data models, source file data dictionaries, data element descriptions, and data conversion rules. In effect, the metadata repository provides the context for the data that resides in the data warehouse. *Without this context, the data warehouse becomes a pit of meaningless scraps of data.* Without metadata, without the context in which to view data, we run the risk of misapplying it."

Metadata is no different from other business data; it too must be integrated across different tools, different systems, application upgrades and replacements. Meeting this need will require that metadata conform with standards being developed by the Meta Data Coalition (MDC; http://www.mdcinfo.org) and Object Management Group (OMG; http://www.omg.org). These are the primary organizations that are working together to set standards for metadata. Notable companies have signed on with each of these organizations:

> The MDC Council, founded in 1995 includes: Commercial Financial Services, Inc., ETI, MICROSOFT, NCR, PLATINUM Technology, Inc., and Sybase. The Technical Subcommittee currently consists of: PricewaterhouseCoopers, NCR, ETI, MICROSOFT, Sybase, Cognos, SAS, CFS, PLATINUM Technology, Inc., Mastersoft International (MSI), and Prudential.
>
> Eleven companies formed the OMG in April 1989, including: 3Com Corporation, American Airlines, Canon, Inc., Data General, Hewlett-Packard, Philips Telecommunications N.V., Sun Microsystems and Unisys Corporation. In October 1989, the OMG began independent operations as a not-for-profit corporation. Through the OMG's commitment to developing technically excellent, commercially viable and vendor independent specifications for the software industry, the consortium now includes over 800 members.

The University must also comply with the Kansas Statewide Technical Architecture (KSTA) available online at: http://da.state.ks.us/ITEC. The KSTA also refers to the Dublin Core Metadata Initiative, available online at: http://purl.org/DC. That initiative proposes a minimum standard metadata set found in Appendix C.

Major Architectural Elements

The overall road map for the development of the technical architecture is best expressed as a diagram. The KU Enterprise Architecture model on the next page illustrates a high level view of the essential components and the relationships necessary to establish the infrastructure for decision support. Following the model graphic are descriptions of each of the major elements, organized in two categories: data stores and services, and functions. Certain technologies that can support the architecture are already in place at the University. It will be important to include such support in the plans for implementing this architecture.

KU Enterprise Data Architecture

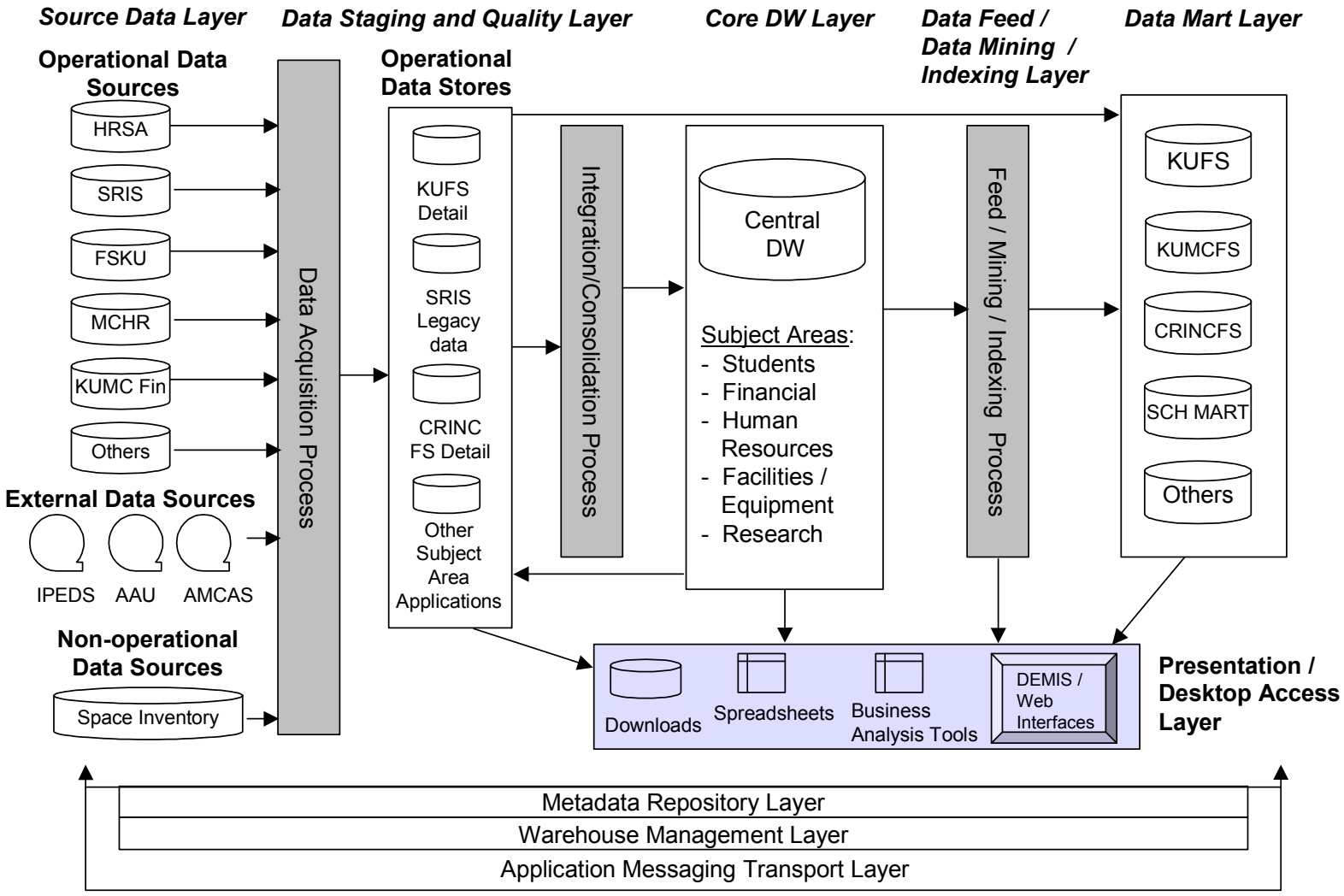


Figure 3: KU Enterprise Data Architecture

Data Stores

Source Data Layer
The data stores that are the original sources for the data warehouse. Includes an operational data layer (PeopleSoft transactional systems, etc.), an external data layer (AAUDE data exchange items, IPEDS data, etc.), and a non-operational data layer (Space Inventory, Senior Survey data, etc.)

Core DW Layer
The data stores, which can be considered the core of the data warehouse. Uses a relational database management system and includes detailed, summarized and customized views.

Data Mart Layer
The reporting tables that are created from the Core DW layer stores. This includes multi-dimensional databases and operational data stores such as those used by KUFS, and the student credit hour queries found in DEMIS.

Metadata Repository Layer
The reference data store which links the other data stores and functions used throughout the University, including the data warehouse.

Other Data Stores
Temporary and permanent data stores which are used by the warehouse systems to provide various services and functions.

Services and Functions

Metadata Repository Layer
Through gathering technical, informational, and knowledge data from the transactional systems and its own data stores, the metadata repository helps in the maintenance, coordination, and publication of the data warehouse systems.

- Provides the context for the data: e.g., the meaning of FTE in student records compared to payroll records.
- Some metadata, such as definitions are readily seen. Yet, there is other metadata behind the scenes at a system level, which is necessary for the warehouse operation.
- Data elements that are most used throughout University systems need to be prioritized and identified as 'University Data.' Definitions for these elements should be the first collected.
- Metadata for core elements must match across systems to support integration and linkage of University systems. This effort requires coordination and support from the various transactional system administrators and the warehouse staff.

Example: DEPTID in the PeopleSoft HR, Financials, and Student systems does not always carry the same meaning within a given system (DEPTID in PS_DEPT_TBL vs. PS_ORG_TBL) or between systems.

Data Acquisition Process
The tools and processes for extracting data from the University's online transactional systems as well as any external data and bringing the data into the staging and quality layer.

Data Staging and Quality Layer
These are processes to prepare data from transactional systems. Data transformations, aggregations, and quality checks occur during this stage.

Core DW Layer
The resulting data stores, which provide the detailed structures, that allow integration of the data. These are the raw materials.

Data Feed / Data Mining / Indexing Layer
Provides data for data marts and data garages. Builds data cubes for multi-dimensional analyses. Indexes data for optimization of access. These are the sub-components assembled from the raw materials.

Data Mart Layer
Provides reporting tables that allow multi-dimensional analyses, drill-downs, multi-dimensional manipulations, and fast access. These are the finished goods assembled from the sub-components.

Presentation / Desktop Access Layer
Provides the end user and the end user's desktop tools with applications to browse, access, query, retrieve and investigate information from the warehouse. Includes basic reports and the ability to extract information into desktop tools as well as the more sophisticated presentations such as data visualization, data mining, and business modeling.

Warehouse Management Layer
These processes provide the scheduling, execution, subscription, and monitoring of the data warehouse services.

Application Messaging Transport Layer
These services provide the inter-process communications between each layer.

Tools Requirements

The KU Enterprise Data Architecture model includes several different processes. There will undoubtedly be overlap among the tools needed to perform all of these tasks. All tools within the architecture must be compatible. The easiest way to insure this compatibility is to require that they comply with commonly accepted standards. The need to meet such standards, along with the complexity of the requirements, implies the use of commercially developed tools. Using such tools to develop the data store components of the architecture can maximize university resources. The tools requirements are organized in the following section by the processes listed in the model of the KU Enterprise Data Architecture.

Data extraction, cleansing / quality, and staging

Tools in this process need to have the ability to:

- Allow scheduling of jobs for automated production runs
- Read Oracle and other ODBC compliant databases
- Read flat files, both local and remote
- Perform edit checks, ranging from simple to complex, even drawing upon data from other applications as necessary
- Apply transformations to data according to business rules
- Write custom transformation algorithms to portray complex business rules
- Integrate with metadata and data mining tools
- Output cleansed data in denormalized reporting formats (ODS) for use by power users
- Replicate data sets
- Manage multiple copies and versions of data sets
- Summarize data for input to the data warehouse

Data warehouse layer

Tools in this layer need to have the ability to:

- Provide administrative services for the warehouse
- Manage the metadata across the full scope of the architecture
- Draw data values from various sources to populate the data warehouse tables, including: ODS, operational data applications, external data, and non-operational data
- Process simple queries
- Communicate with Internet servers

Data feed / Data mining / Indexing

Tools in this process need to have the ability to:

- Draw data from the data warehouse to populate data marts
- Build data cubes for multidimensional analysis
- Index data for bit-mapped indexing
- Support high speed analysis of detail data

- Support pattern identification algorithms for 'automatic' analysis or indexing

Data mart layer

Tools in this layer need to have the ability to:

- Support multidimensional manipulation and analysis
- Support drill-down and drill-across through the data
- Provide fast access to the data

Presentation, query, and analysis (end-user applications)

Tools in this layer need to have the ability to:

- Output pages in a variety of formats, including: HTML, rich text format (RTF), PDF, and XML
- Output data in formats to support standard desktop tools, including MS Excel, MS Access, and SAS; and in a variety of formats including tab delimited, comma delimited (CSV), and text (ASCII)
- Select or filter subsets of data for export to desktop computers in database formats
- Support simple queries and analyses from desktop computers, including listings, counts, sums, means, and pivot tables
- Provide users with easy access to canned queries
- Support statistical analyses from desktop computers, including TQM control charts, MANOVAs, non-linear regressions, and discriminant analysis
- Provide highly structured data tailored for modeling for “power users”
- Support sophisticated “what-if” analysis
- Support visualization of multidimensional analysis
- Locate and navigate data through a data directory
- Drill-down from high-level to detail data
- Provide flexible reporting options for canned or custom queries
- Support a variety of users from the casual user to the power user

Metadata Repository Layer

Tools in this layer need to have the ability to:

- Create and maintain models of metadata, including ERDs, process models, and dimensional models
- Store and manage metadata
- Update metadata
- Extract metadata
- Coordinate metadata

Warehouse Management Layer

Tools in this layer need to have the ability to:

- Schedule required tasks
- Execute required tasks

- Support subscription to notification services
- Monitor and manage processing of queries

Application Messaging Transport Layer

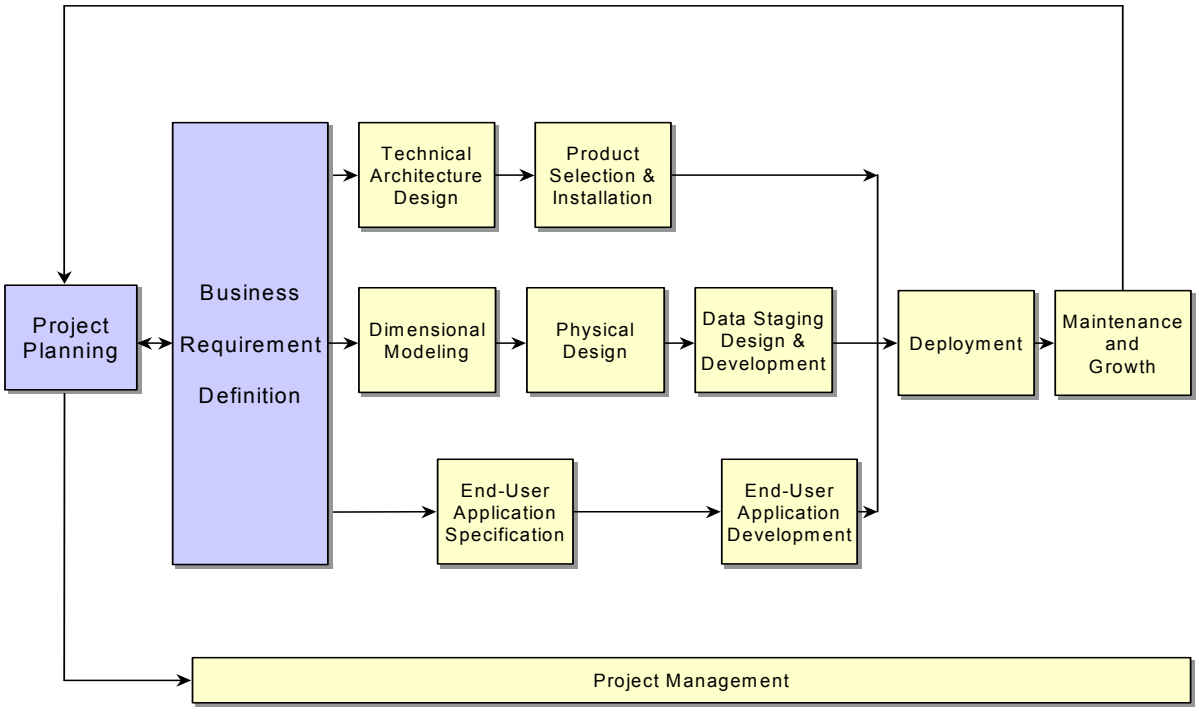
Tools in this layer need to have the ability to:

- Connect various data and tasks throughout the enterprise architecture model
- Store and forward data, rules, and instructions as needed by all parts of the model
- Provide and manage the bandwidth needed for communication among all parts of the model

Recommendations

As an outcome of our discussions and based on information included in this report, the Knowledge Management Technical Infrastructure Definition Task Force recommends the following actions be taken in order to facilitate the establishment of a robust Decision Support Infrastructure for the University of Kansas:

1) Adopt the *Guiding Principles* statement as the basis for any central Decision Support project for the University.
2) Adopt the KU Enterprise Data Architecture model.
   a. Recognize this as a University-wide, cooperative venture with concurrent, parallel efforts occurring on the Lawrence and KUMC campuses with joint coordination between these efforts.
   b. Adopt a phased implementation approach with centrally prioritized tasks to be addressed.
3) Begin the Project Planning and Business Requirements Definition steps of the Business Dimensional Lifecycle for the University's Data Warehousing initiative. Some of these steps are:
   a. Adopt the Scope statement
      i) Approve the definition of University Data
      ii) Approve the criteria for data to be included in the data warehouse
   b. Identify the initial problem statements which provide the business case for the data warehousing initiative
   c. Identify and prioritize projects which address these problem statements..



4) Based on the KU Enterprise Data Architecture Model and Staffing Roles requirements, evaluate the following to determine what is already in place, under development or missing:
   a. Relevant systems
      i) Identify existing operational systems that contain data that needs to be in the data warehouse
      ii) Identify existing non-operational systems that contain data tha needs to be in the data warehouse
      iii) Identify existing external systems that contain data that needs to be in the data warehouse
      iv) Is there any data needed for which there is not a source?
   b. Staff
      i) Identify available staff to fill the roles in the table below
      ii) People can fill multiple roles
      iii) A role may be filled by more than one person
      iv) Not all roles are needed during all phases
      v) Are there any roles that can't be filled by existing staff?
   c. Tools
      i) Identify those tools that the university owns (this includes servers, databases, software, etc.)
      ii) Descriptions of the tools needed can be found in the section on *Tools Requirements*
      iii) Are there any toos we don't have?
   d. Projects
      i) What projects currently exist that either affect or may be affected by the data warehouse initiative?
      ii) Determine whether a project needs to be realigned with the data warehouse initiative
      iii) Make the necessary changes in the project.

Table 1: Staffing Roles and Functions

Staffing Roles / Functions	
Area	**Role**
Management Roles	User sponsorship
Management Roles	Technical sponsorship
Management Roles	Project management
Management Roles	Technical lead(s)
Management Roles	User project lead(s)
Requirements Development Roles	User requirements analysis
Requirements Development Roles	Technical / security architecture
Requirements Development Roles	Data stewardship / Metadata modeling
Requirements Development Roles	Quality assurance analysis
Database Development Roles	Data modeling and conceptual/logical database design
Database Development Roles	Source data analysis
Database Development Roles	Decision support system architecture
Database Development Roles	Database administration and physical database design
Database Development Roles	Data staging system design / data movement and middleware specialist (ETL)
Database Development Roles	Data staging programming
Application Development Roles	End user applications development/Front-end tools/Presentation system design
Application Development Roles	Data warehouse education
Application Development Roles	Technical support (software & hardware)
Application Development Roles	Help desk
Application Development Roles	Data warehouse auditing
Other	Legal and policy issues for accessing and using data

5) Based on the results of Recommendations #3 and #4, charge a task force/leadership group to
   a. Select the first project for inclusion in the data warehouse
   b. Define scope of the project
   c. Define existing systems that are a part of this project
   d. Identify the dtat in these systems that belongs in the data warehouse
   e. Continue with development following the Business Dimensional Lifecycle and the KU Enterprise Data Architecture model
6) Based on the outcome of the Project Planning and Business Requirements Definition phases, prioritize the staffing requirements for implementation and recognize that several roles will require full-time, dedicated staff.
7) Adopt and promulgate policies and procedures directed toward ensuring the quality of data in the decision support system (Guiding Principle #9).
8) Adopt and promulgate policies and procedures directed toward ensuring stable and secure access to the data warehouse for appropriate University personnel (Guiding Principle #7).

9) Define a decision-making and prioritization mechanism for coordinating and resolving issues such as non-standard project implementations, task and resource prioritization, metadata definition conflicts, etc.
10) In order to facilitate University awareness and continued close cooperation, implement various education and communications mechanisms for information dissemination including:
    a. Reconvene the Information Management Planning Retreat participants to present implementation plans and gather feedback.
    b. Establish a website for centralized dissemination of Decision Support Infrastructure documents and news.
    c. Establish a discussion group (e.g. Listserv) for proactive information dissemination and discussion.

Appendix A: Knowledge Management Infrastructure Definition Task Force

Established by: University Steering Committee for Information Management, Co-Chairs: J. Bingham, M. Goodyear, D. Teeter

Members: R. Cherland, C. Pace (chair), B. Underwood, B. Warner

Effective Dates: February 18, 2000 to May 18, 2000

Background:

The University Steering Committee for Information Management (USCIM) provides overall guidance and policy development for the implementation of information management at the University of Kansas. On January 7, 2000, the USCIM members attended a workshop presented by Ken Orr, of The Ken Orr Institute, on *An Introduction to Data Warehousing and Information Management,* to give committee members a common contextual framework on enterprise-wide information management techniques.

Following the Orr workshop, the USCIM co-chairs have determined that the next step is to use the background information provided by Orr to develop technical and process models for a KU knowledge management infrastructure.

Charge:

The Knowledge Management Infrastructure Definition Task Force is charged with:

- Defining the guiding principles of building and maintaining the KU KM infrastructure and outlining the business rules involved
- Using the Orr "Enterprise Data Architecture" model as appropriate, defining the elements of the technical architecture to be used for the KU KM infrastructure, including
    - Database architecture / data repository types
    - Data model outline(s)
    - Metadata requirements / repositories
    - Management requirements / routines
- Defining the requirements for
    - Data extraction, cleansing and staging tools
    - Data warehouse builder tools
    - Database(s)
    - Presentation, query, and analysis tools

The Task Force will present their report and recommendations, in writing, to the USCIM by May 18, 2000.

Potential Future Steps:

Following completion of the Task Force report, and discussion by the USCIM, potential next steps for the Task Force include:

- Identifying the initial data sources that will populate the KU KM infrastructure and define a process for identifying additional sources over the next 12 months
- Defining the roles involved in developing, maintaining and using the KM infrastructure
- Recommending next steps in implementing the KU KM infrastructure model
- Software recommendations

2/18/00

Appendix B: Sample Project Plan Outline and Role Responsibilities

Insert link here

Appendix C: Dublin Core Metadata Element Set

Dublin Core Metadata Element Set, Version 1.1: Reference Description

Status of this document: This is a Dublin Core Metadata Initiative recommendation. Publication as a recommendation signifies that the specifications are stable and are supported for adoption by the Dublin Core community.

This document is the reference description, version 1.1 of the Dublin Core Metadata Element Set. This document supersedes the Dublin Core Metadata Element Set, version 1.0. See the Dublin Core Home Page (http://purl.org/dc/) for further information about the workshops, reports, working group papers, projects, and new developments concerning the Dublin Core Metadata Element set.

The document summarizes the updated definitions for the Dublin Core metadata elements as originally defined in [RFC2413]. These new definitions will be officially known as Version 1.1.

The definitions utilize a formal standard for the description of metadata elements. This formalization helps to improve consistency with other metadata communities and enhances the clarity, scope, and internal consistency of the Dublin Core metadata element definitions.

Each Dublin Core element is defined using a set of ten attributes from the ISO/IEC 11179 [ISO11179] standard for the description of data elements. These include:

Name - The label assigned to the data element
Identifier - The unique identifier assigned to the data element
Version - The version of the data element
Registration Authority - The entity authorized to register the data element
Language - The language in which the data element is specified
Definition - A statement that clearly represents the concept and essential nature of the data element
Obligation - Indicates if the data element is required to always or sometimes be present (contain a value)
Datatype - Indicates the type of data that can be represented in the value of the data element
Maximum Occurrence - Indicates any limit to the repeatability of the data element
Comment - A remark concerning the application of the data element

Fortunately, six of the above ten attributes are common to all the Dublin Core elements. These are, with their respective values:

Version:	1.1
Registration Authority:	Dublin Core Metadata Initiative
Language:	en
Obligation:	Optional
Datatype:	Character String
Maximum Occurrence:	Unlimited

The above attributes will not be repeated in the below definitions, however, they do represent part of the formal element definitions.

The definitions provided here include both the conceptual and representational form of the Dublin Core elements. The Definition attribute captures the semantic concept and the Datatype and Comment attributes capture the data representation.

Each Dublin Core definition refers to the resource being described. A resource is defined in [RFC2396] as "anything that has identity". For the purposes of Dublin Core metadata, a resource will typically be an information or service resource, but may be applied more broadly.

Element: Title

Name: Title
Identifier: Title
Definition: A name given to the resource.
Comment: Typically, a Title will be a name by which the resource is formally known.

Element: Creator

Name: Creator
Identifier: Creator
Definition: An entity primarily responsible for making the content of the resource.
Comment: Examples of a Creator include a person, an organization, or a service. Typically, the name of a Creator should be used to indicate the entity.

Element: Subject

Name: Subject and Keywords
Identifier: Subject
Definition: The topic of the content of the resource.
Comment: Typically, a Subject will be expressed as keywords, key phrases or classification codes that describe a topic of the resource. Recommended best practice is to select a value from a controlled vocabulary or formal classification scheme.

Element: Description

Name: Description
Identifier: Description
Definition: An account of the content of the resource.
Comment: Description may include but is not limited to: an abstract, table of contents, reference to a graphical representation of content or a free-text account of the content.

Element: Publisher

Name: Publisher
Identifier: Publisher
Definition: An entity responsible for making the resource available
Comment: Examples of a Publisher include a person, an organization, or a service. Typically, the name of a Publisher should be used to indicate the entity.

Element: Contributor

Name: Contributor
Identifier: Contributor
Definition: An entity responsible for making contributions to the content of the resource.
Comment: Examples of a Contributor include a person, an organization, or a service. Typically, the name of a Contributor should be used to indicate the entity.

Element: Date

Name: Date
Identifier: Date
Definition: A date associated with an event in the life cycle of the resource.
Comment: Typically, Date will be associated with the creation or availability of the resource. Recommended best practice for encoding the date value is defined in a profile of ISO 8601 [W3CDTF] and follows the YYYY-MM-DD format.

Element: Type

Name: Resource Type
Identifier: Type
Definition: The nature or genre of the content of the resource.
Comment: Type includes terms describing general categories, functions, genres, or aggregation levels for content. Recommended best practice is to select a value from a controlled vocabulary (for example, the working draft list of Dublin Core Types [DCT1]). To describe the physical or digital manifestation of the resource, use the FORMAT element.

Element: Format

Name: Format
Identifier: Format
Definition: The physical or digital manifestation of the resource.
Comment: Typically, Format may include the media-type or dimensions of the resource. Format may be used to determine the software, hardware or other equipment needed to display or operate the resource. Examples of dimensions include size and duration.
Recommended best practice is to select a value from a controlled vocabulary (for example, the list of Internet Media Types [MIME] defining computer media formats).

Element: Identifier

Name: Resource Identifier
Identifier: Identifier
Definition: An unambiguous reference to the resource within a given context.
Comment: Recommended best practice is to identify the resource by means of a string or number conforming to a formal identification system.
Example formal identification systems include the Uniform Resource Identifier (URI) (including the Uniform Resource Locator (URL)), the Digital Object Identifier (DOI) and the International Standard Book Number (ISBN).

Element: Source

Name: Source
Identifier: Source
Definition: A Reference to a resource from which the present resource is derived.
Comment: The present resource may be derived from the Source resource in whole or in part. Recommended best practice is to reference the resource by means of a string or number conforming to a formal identification system.

Element: Language

Name: Language
Identifier: Language
Definition: A language of the intellectual content of the resource.
Comment: Recommended best practice for the values of the Language element is defined by RFC 1766 [RFC1766] which includes a two-letter Language Code (taken from the ISO 639 standard [ISO639]), followed optionally, by a two-letter Country Code (taken from the ISO 3166 standard [ISO3166]). For example, 'en' for English, 'fr' for French, or 'en-uk' for English used in the United Kingdom.

Element: Relation

Name: Relation
Identifier: Relation
Definition: A reference to a related resource.
Comment: Recommended best practice is to reference the resource by means of a string or number conforming to a formal identification system.

Element: Coverage

Name: Coverage
Identifier: Coverage
Definition: The extent or scope of the content of the resource.
Comment: Coverage will typically include spatial location (a place name or geographic coordinates), temporal period (a period label, date, or date range) or jurisdiction (such as a named administrative entity).
Recommended best practice is to select a value from a controlled vocabulary (for example, the Thesaurus of Geographic Names [TGN]) and that, where appropriate, named places or time periods be used in preference to numeric identifiers such as sets of coordinates or date ranges.

Element: Rights

Name: Rights Management
Identifier: Rights
Definition: Information about rights held in and over the resource.
Comment: Typically, a Rights element will contain a rights management statement for the resource, or reference a service providing such information. Rights information often encompasses Intellectual Property Rights (IPR), Copyright, and various Property Rights. If the Rights element is absent, no assumptions can be made about the status of these and other rights with respect to the resource.

Bibliographic Resources

Agosta, Lou, The Essential Guide to Data Warehousing, Prentice Hall PRT, Upper Saddle River, NJ (2000)

Applehans, Wayne, Globe, Alden, and Laugero, Greg, Managing Knowledge: a Practical Web-Based Approach, Addison Wesley Longman, Reading, MA (1999)

Barquin, Ramon C. and Edelstein, Herb, Planning And Designing the Data Warehouse, Prentice Hall PTR, Upper Saddle River, NJ (1997)

Data Warehousing Institute, The, *Data Warehousing Fundamentals: A Roadmap to Success*
http://www.dw-institute.com/

Devlin, Barry, Data Warehouse: from Architecture to Implementation, Addison Wesley Longman, Reading, MA (1997)

Dublin Core Metadata Initiative, *Dublin Core Metadata Initiative* [Online serial], (June 1999)
http://purl.org/DC

Dyche, Jill, e-Data, Addison Wesley Longman, Reading, MA (2000)

English, Larry P., Improving the Data Warehouse and Business Information Quality: Methods for Reducing Costs and Increasing Profits, John Wiley & Sons, New York, NY (1999)

Gartner Group
http://www.gartner.com
http://www.gartner.com/public/static/home/ourservices/use/dm/home.html (Data Management)
http://www.gartner.com/public/static/home/ourservices/use/kms/home.html (Knowledge Management)

Inmon, William H., Building the Data Warehouse, John Wiley & Sons, New York (1996)

Inmon, William H., Data Stores, Data Warehousing, and the Zachman Framework: Managing Enterprise Knowledge, McGraw-Hill, New York, NY (1997)

Kansas Statewide Technical Architecture, The, Version 7.0, Kansas Information Technology Executive Council (ITEC), Topeka, KS (April 2000)

Kimball, Ralph, et al, The Data Warehouse Lifecycle Toolkit, Wiley, New York, NY (1998)

Kimball, Ralph, The Data Warehouse Toolkit: Practical Techniques for Building Dimensional Data Warehouses, John Wiley & Sons, New York, NY (1996)

Laudon, Kenneth C., and Laudon, Jane P. Essentials of Management Information Systems, 3rd ed., Prentice Hall, Upper Saddle River, NJ (1999)

Mattison, Rob, Web Warehousing and Knowledge Management, McGraw-Hill, New York (1999)

META Group
http://www.meta.com

Meta Data Coalition, *Meta Data Coalition: driving the evolution of metadata interchange format standards* [Online serial], (June 1999)
http://www.mdcinfo.com

Meyer, Don and Cannon, Casey, Building a Better Data Warehouse, Prentice Hall PTR, Upper Saddle River, NJ (1998)

Object Management Group, *Object Management Group Home Page*, [Online serial], (June 1999)
http://www.omg.org

Orr, Ken, *An Introduction to Data Warehousing and Information Management. Workshop,* presented to the USCIM at the University of Kansas, Lawrence, KS, January 7, 2000

Simon, Alan R., Data Warehousing for Dummies, IDG Books Worldwide, Inc., New York, NY (1997)