.

"Information is the lifeblood of any research university. In order for the Institute to achieve its commitments to the Rensselaer Plan, it must make strategic use of its information resources and infrastructure."

.

Executive Summary

This document is a proposal for building a data warehouse architecture that will consolidate and transform data into useful information for the purpose of decision-making and for establishing a new function that offers a broad array of decision-support services to all units at Rensselaer.

Executives and decision-makers often need information to chronicle the past, describe current circumstances, and anticipate the future. Presently, decision-makers across the Institute rely on hard copy reports to provide information. Typically, any request for information is forwarded to the operational areas of the Institute, which provide hard copy reports reflecting the data gathered in their functional area. To analyze and transform data into useful information, decision-makers and their staff have to manually re-enter the non-integrated data into their own mini-systems. This type of operation fosters the, "You can't see the forest for the trees" syndrome, and the executives are either drowning in too much data with no option to analyze it or too little data, which means they are back to square one and must request additional information. Often executives receive multiple, conflicting information or information that is based on incomplete assumptions about the types of analysis required.

In addition to accurate and integrated information, executives and decision-makers across the Institute need assistance in defining, analyzing, and interpreting institutional data to support strategic planning and policy-making processes at Rensselaer. Presently, each unit and division across the Institute is struggling on its own to conduct data analysis and interpret results obtained from non-integrated data sources. There is no institute-wide function that facilitates the flow of accurate, timely information and assists all levels of management in defining issues, selecting research designs, obtaining information, and interpreting results.

To demonstrate the viability of data warehouse architecture, a prototype was developed within the former Administrative Information Services unit (AIS), now Integrated Administrative Computing Solutions (IACS) The data warehouse prototype was able to demonstrate that not only is the proposed system technologically feasible, but that it is of immediate interest and use to Rensselaer executives and decision-makers.

A data warehouse architecture will be built to facilitate better access to information by a broader population of campus users who are not familiar with where data is stored or how it's coded in the operational systems. The data warehouse would contain consistent, cleaned, transformed, and summarized information derived from the Banner Finance and Student Information Systems and other operational systems.

The data warehouse architecture alone will not address all existing issues; it has to be combined with a broad array of decision-support services accessible to all units at Rensselaer. The primary focus of this service will be to facilitate the collection, analysis, and interpretation of Institutional data and provide this information to support-planning and decision-making personnel. The newly established function within the former Computing and Information Services (CIS), now the Division of the Chief Information Officer (DotCIO) backed by data warehouse-enabling technology, will deliver a full, comprehensive service to Rensselaer.

1.0 Introduction

In December 1999, President Shirley Ann Jackson chartered the Rensselaer Improvements Through the Use of Information Technology Task Force (RensIT2 Task Force), a representative group of key decision-makers at Rensselaer, to "…identify opportunities for strategic benefits through better use of information technology." The task force was directed to identify opportunities in the areas of information and analytical tools for decision-makers at all levels at the Institute.

.

A data warehouse prototype was developed within the former Administrative Information Services unit. The goal of the prototype is to evaluate the technical feasibility of such a system and demonstrate that the architecture provides easy access to information. The prototype was able to demonstrate that not only was the proposed system technologically feasible, but it was of immediate interest and use to Rensselaer executives and decision-makers.

2.0 Drivers of Change

By the end of 1998, Rensselaer successfully implemented the Banner Student Information System that replaced a majority of outdated systems on campus. Today, in addition to the centralized Banner Finance and Student System, we employ several other computerized systems across the Institute to support operations such as Institute Advancement, Space Management, Axis, Contract and Grants, the Bookstore, CMMS, Meal Services, and Health Services. Moreover, many of the schools and departments on campus employ their own homegrown systems to maintain financial and student data. Although overall these operational systems are addressing the operational needs of the Institute (such as student registration and record maintenance, finance, payroll, and purchasing), the needs of Institute decision-makers go unmet.

Executives and decision-makers often need information to chronicle the past, describe current circumstances, and anticipate the future. Today, decision-makers across the Institute rely on the hard copy reports to provide information. Typically, any request for information is forwarded to the operational areas of the Institute which provide hard copy reports reflecting the data gathered in their functional areas. To analyze and transform data into useful information, decision-makers and their staff have to manually re-enter the non-integrated data into their own mini-systems. This type of operations fosters the, "You can't see the forest for the trees" syndrome, and executives are either drowning in too much data with no option to analyze it or find too little data, which means they are back to square one requesting additional information. Often the executives receive multiple, conflicting information or information that is based on the incomplete assumptions about the types of analysis required.

The needs of executives and decision-makers are often not considered in data gathering; therefore, data needed for decision-making is often not available. By the nature of the operational transaction processing systems, the types of information gathered and maintained in our centralized systems are driven by the needs of the core operational offices to support the operational needs of the Institute, such as student admission and registration, billing, payroll, and housing. While this is essential to Institute operations, we are lacking functions that will help identify types of information needed by decision-makers and facilitate the process of gathering and maintaining information.

Moreover, decision-makers across the Institute rely on their own resources to help gather, analyze, and interpret Institute data. No centralized resources are available today to help decision-makers in facilitating the flow of accurate, timely information and assist them in defining issues, selecting research designs, obtaining information, and interpreting results.

In addition to the needs of decision-makers within the Schools of Architecture, Engineering, Humanities and Social Sciences, Management, and Science, administrators in these schools are often unable to access information. In several cases the data simply does not exist. As with the needs of executives, the needs of academic administrators are often not considered in data gathering. In other cases where data does exist, it is stored in a variety of complex ERP systems and can be retrieved only in the form of a hard copy report by a professional programmer. As in the case of executives, school administrators face similar problems of first identifying the appropriate core administrative office or administrative system that might have the necessary information and then requesting the information.

Currently we don't have a process in place that will help Institute constituencies identify types of information available and how to obtain it.

As a result, the schools and departments are simultaneously investing in hardware and software, developing systems, adding staff, and maintaining these systems and staff overtime.

.

3.0 Fundamental Problems with the Reporting and the Decision-support Environment at Rensselaer

There are several fundamental reasons why our current reporting system and decision-support environment is not addressing the needs of decision-makers across the Institute. First and foremost is that our operational systems are not designed to support analytical processing and we do not employ ad-hoc tools designed to perform analytical functions such as "drilling down," "slicing and dicing," "what if?" Furthermore, the lack of an Institute-wide function that supports the needs of decision-makers and school administrators causes deficiencies in data ownership and data access policies, since the majority of these require access to information across the operational boundaries. Moreover, this information required for decision-making is non-integrated, missing, ambiguously defined, or simply inaccessible.

The deficiencies in terms of systems, tools, services, and data are summarized below.

Systems: Operational systems are not designed to support analytical processing and the needs of decision-makers.
In contrast to the current, up-to-date information used in the transaction processing system, the information used for analysis and decision-making in the Institute is fairly stable, consisting of extracts or snapshots of data taken from the transaction processing system. This information often has to be derived from multiple sources. Added value comes when information is organized so that data from one source can be compared to, combined with, or related to data from other sources. Generally, analysis and decision making requires non-volatile, historical as well as current information that is organized based on the subject analyzed. For example, the analysis of revenue will require data extracted from several sources such as Student receivables, Endowment, Institute Advancement, Research Accounting, etc.

Here is how the operational, transaction processing systems are different from the analytical and decision support systems.

Transaction Processing	*Analytical Processing*
Process oriented: Operational systems are organized by specific processes or tasks; for example, billing, registration, admissions, and financial aid. The majority of these processes rely on a separate set of tables to maintain process-specific data.	**Subject Oriented**: Analytical systems are organized around business subjects the organization needs information regarding and are populated by many operational systems. *For example, revenues will incorporate elements of Student (tuition), Alumni (gifts, endowments), Finance (research, GFA, Payroll, etc.), and Financial Aid systems.*
Organized and managed to **support transaction processing.**	Organized and managed to **support reporting and analytical needs**. Combines data from operational systems in a way that is **meaningful to the business community.**
Process driven.	**Data** driven.
Designed and optimized for efficient, **fast inserts and** updates**.**	Designed and optimized for efficient, **fast retrieval.**
Requires **certain level of computing skills**.	Appeals to **wide range of computing skills**.

.

Application specific - Designed and built separately; for example, *FAIMS, Student, FA, and Alumni Systems.*	**Integrated** – Designed or "architected" at one time and implemented iteratively over some period of time.
Primarily concerned with **current** data.	Time Variant – Generally concerned with **historical** data (enrollment over time, changes in majors); time is defined as daily, weekly, monthly, etc.
Updated constantly. **Data changes according to need**, not a fixed schedule.	Non-volatile – **Data is added regularly**, but once loaded, the data is rarely ever directly changed, thus **preserving an historic view** of information.
Updated data.	Read only data.

Tools: Executives and decision-makers across Rensselaer require better analytical tools for improved decision making. No ad hoc access to information or tools is available today to support on-line analytical processing. Currently available hardcopy reports do not meet all the specific needs of the individual departments and the general needs of the University. Typically, any request for information is forwarded to the core operational areas of the Institute, which provide hard copy reports reflecting the data relative to the departments. To analyze and transform data into useful information, decision-makers and their staff have to manually re-enter the non-integrated data into their own mini-systems.

Services: The majority of our current organizational resources is dedicated to support the operational needs of the core operational offices. We are lacking the organizational structure that is responsible for addressing the informational and the reporting needs of decision-makers and the various schools. Today each unit and division across the Institute is struggling on its own to identify what type of information is available, obtain it, conduct data analysis, and interpret the results.
There is no Institute-wide function that facilitates the flow of accurate, timely information and assists all levels of management in defining issues, selecting research designs, obtaining information, and interpreting results.

Data: For decision-making purposes, existing data is either non-integrated or missing or is ambiguously defined.

Non-integrated - Although the majority of our data is stored in the same Banner databases, our data is not integrated for decision-support. *Integration means the ability to perform an integrated analysis that incorporates data elements from different areas*. Today we employ several operational systems across the Institute. Obviously, the information stored in one system does not necessarily correspond to the information stored in another system, and our current environment does not support activities that require data from the one to be compared to, combined with, or related to from other sources. The time it takes for campus constituencies to obtain information from different sources and put it all together is interminable.

Missing - The needs of decision-makers are often not considered in data gathering; therefore, data needed for decision-making is often not available.

Ambiguously defined - Currently we lack the unified approach in data definitions and classifications. The operational needs of the various operational offices are different, and so are their needs for defining and classifying data. For example, classifications of the student populations by the Budget office is different from the classifications of the Registrar's office simply because their operational needs are different. Similarly, the organizational codes in the Banner Finance System do not correspond to the department codes in the Banner Student System.

Data policies – Currently we are lacking policies to govern the process by which information is gathered, accessed, and distributed within the Institute.

.

4.0 Defining Solution

This proposal calls for establishing a data warehouse services function at Rensselaer in support of management's decision-making process. The newly established function will be responsible for managing Institutional data to provide reliable, accurate, secure, and accessible data to meet strategic and management needs. "Institutional data" is data relevant to planning, managing, operating, and auditing administrative activities, regardless of whether the data is used or maintained by administrative or academic units.

The data warehouse architecture will be built to facilitate better access to information by a broader campus population of users not familiar with where the data is stored or how it's coded in the operational systems. The data warehouse will contain consistent, cleaned, transformed, and summarized information derived from the Banner Finance, Student, and other operational systems. The operational systems, which are both large and complex, store detailed information about nearly all student and financial transactions at Rensselaer. The data warehouse would consolidate and organize these vast amounts of transactional data into a reasonable number of fields and tables that can be queried and analyzed to support decision-making. The data warehouse will include information to help users understand the nature and use of data stored in repositories. Users will access information stored in the data warehouse through a set of easy-to-use tools that have graphical and Web-based user interfaces.

The data warehouse architecture alone will not address all existing issues. It has to be combined with a broad array of decision-support services accessible to all units at Rensselaer. The primary focus of this service will be to facilitate the collection, analysis, and interpretation of Institutional data and provide information to support planning and decision-making. The newly established function within DotCIO backed by the data warehouse-enabling technology will deliver a full, comprehensive service to Rensselaer.

5.0 Benefits

Unlike the standardized, printed reports that the Institute largely relies on to serve this function today, the data warehouse architecture can deliver information that is more flexible, timely, and responsive to emerging needs. At the same time, since the procedures used to transform the data from the operational systems to the data warehouse incorporate business rules and other quality measures, a higher degree of data accuracy and consistency can be maintained.

The major benefits of maintaining a data warehouse are summarized below:

Empowers decision-makers by enabling direct access to accurate, consistent, and non-volatile information and by offering decision-support services that will facilitate the analysis and interpretation of Institutional data.

Redirects costly personnel hours from data gathering, matching, and consolidating to data analysis. Reduces the need for Rensselaer information consumers to replicate data and maintain redundant tracking systems.

Promotes the "no walls" culture. The information sharing will lower the walls among all administrative and academic units, creating alliances and connecting people to each other and the Institute as a whole.

Improves data quality over time. Broader access to information will help better define what types of information we need on hand for effective decision-making. This will identify areas for improvement in data gathering functions and data entry.

Enhances institutional effectiveness by extending and expanding the use of management information across the Institute. The extensive use of management information will eventually promote the analytical culture that places value on information and analysis as opposed to "gut feelings."

Improves integrity and conformity of campus-wide information. As more people on campus use the data, error will be identified and corrected thereby improving data integrity over time.

.

Improves access to the historical data. Historical data will be available to the broader campus community for ad hoc access and analytical processing.

Provides direct support to Institute executives and the schools in collection, analysis, interpretation, and provision of Institute information.

6.0 Data Warehouse Project

The Data Warehouse Project is far more than a technical project. It will require an understanding of the Rensselaer processes and the environment in which the information is produced and will be used. Data warehouse staff will work closely with information consumers and producers. A clear set of policies and procedures will foster effective collaboration among different campus users groups.

To successfully handle the complexity of this project, the following guidelines should be followed:

ITERATIVE DEVELOPMENT: Within Rensselaer we have a broad variety of informational needs:

> Ability to provide decision support and executive information for many types of analysis, from the "what if" scenarios to addressing questions such as, "how should revenue be allocated among academic schools and departments?" to the need to access student lists based on certain academic performances, enrollments, or demographics.

> Ability to answer a simple question such as, "how much graduate financial aid do we pay for a particular student or certain group of students?" and from the ability to monitor students changing majors to the ability to analyze how the demographics of our students have changed over the years.

> Need to analyze the breakdown of the tuition revenues, faculty teaching and research loads, to the needs for budgeting forecasting.

These and many other informational needs cannot be addressed and solved overnight. Therefore, the data warehouse has to be developed in an evolutionary, one-step-at-a-time fashion, starting with a small, well-defined and commonly agreed to set of requirements, which will be expanded over time.

DATA STEWARDSHIP AND DATA MANAGEMENT: This project is not simply a matter of loading data from several operational systems into the data warehouse; rather it involves considerable behind-the-scenes work to reconcile data conflicts and deal with data access, ownership, stewardship, and management issues. Clear definitions on what is considered to be "Rensselaer data" and who should be allowed to access it, and what policies and procedures should be followed for accessing data need to be addressed. *The development of a clear set of policies, procedures and guidelines for Rensselaer-wide data administration activities will be essential for the success of this project.*

.

CAMPUS PARTICIPATION AND COLLABORATION: Extensive campus participation in the development of the data warehouse is perhaps one of the most critical components. The process itself will be the very foundation of its evolution. Since the data warehouse is designed to house the Institute's data, various members of the Institute should be involved in representing their needs and requirements—from the central administrative offices to administrators in the schools and departments who have a long-standing need for access to Institutional data for management decision-making in their units.

IMPLEMENT CENTRAL METADATA MANAGEMENT: Metadata, which is descriptive information about data in the data warehouse, is key to providing users with an understanding of the data contained in the warehouse. Metadata helps to identify types of information available and describes the source data and the transformation rules for migration to the warehouse. Metadata describes data in meaningful business terms that campus users can understand. To achieve an effective, central metadata management, we will need to deploy data extraction, transformation, and loading tools that support centralized metadata management. Although these tools will be used by technical staff and will be "hidden" from users, nevertheless they will enable a better-integrated approach to cleaning, transforming, and loading information from the operational systems into the data warehouse. This will help to prevent the "silo" effect of building non-integrated pockets of information and avoid conflicts in data definitions and interpretations.
Development of metadata is a critical part of evolving and improving the data warehouse over time.

PROVIDE THE ORGANIZATIONAL SUPPORT STRUCTURE: As with operational systems employed today around campus, the data warehouse is more than just hardware and software. It is also people who are highly skilled and knowledgeable in the needs of Rensselaer executives and decision-makers and the capabilities of the variety of tools used in data warehouse architecture.
Over time, they must monitor and address the informational needs of the Institute that will define types of information included in the data warehouse; serve as informational resources; promote standardization in data definitions and classifications; and coordinate with the campus community in establishing policies, standards, procedures, and roles regarding data management and data access issues. Moreover, they must be there to support and guide Institute constituencies in analyzing and interpreting data, provide training and support in using tools, install and test upgrades, and administer to the security and configuration.

Here is a quick summary of several data warehouse features, their advantages, and the associated support level they demand:

Feature	Advantage	Support Cost
Relevant, integrated, clean, summarized information.	Supports effective decision-making.	Requires ongoing monitoring and careful analysis of the informational needs. Requires standardization in data definitions and classifications. Requires coordination with campus community to establish policies, standards, procedures, and roles regarding data management and data access issues.
Web-based and desktop query reporting tools.	Allows direct access to information by the consumers.	Training, documentation, phone support, security administration. Front-end tools support for installation, configuration, and upgrades.
Data Warehouse architecture.	Allows hundreds of people access to current and historic information while at the same time ensuring responsiveness and data integrity.	Requires specialized skills and ongoing effort to monitor and control.

.

Provision of metadata.	Allows campus users to understand what information is available in the data warehouse and what it means.	Requires ongoing support in maintaining and publishing metadata to describe the source data, the transformation rules for migration to the warehouse.
Facilitating the collection, analysis, and interpretation of institutional data.	Supports strategic planning and policy making	Requires specialized skills in data analysis, survey design, and administration.

7.0 Project Plan

The iterative approach used in implementing the data warehouse architecture calls for implementation in phases. The implementation of The Data Warehouse Project will be broken down into three major phases: 1) *Infrastructure Phase* 2) *Implementation Phase* and 3) *Growth Phase*. It will take three and one-half years (3 ½) to complete the first two phases. The third phase represents future data warehouse growth and evolution.

1) Infrastructure Phase: Generally speaking, this is the time to purchase hardware and software necessary to build the data warehouse, hire and train personnel, prioritize business demands, and get the project off the ground.

2) Implementation Phase: This is when the actual implementation of the subject areas begins. Subjects within the areas of Finance, Budgeting, Human Resources, Purchasing, Research, Contracts and Grants, Admissions, Financial Aid, Students, Registrations, Alumni, and Properties will be analyzed, prioritized, and implemented.

3) Growth Phase: It is worthwhile to notice that unlike traditional systems development initiatives, change should be viewed as a sign of success, not failure. Growth phase is dedicated to continuing to address business users' demands for evolution and growth.

Below is a summary of the major activities expected during each phase:

7.1 Infrastructure Phase   Jul-01 Dec-01
The major activities during this phase break out into five groups, each described below.

7.1.1 PROJECT ORGANIZATION
Includes all the activities necessary to get the project off the ground, such as:

- Defining user groups and committees.
- Hiring and training staff.
- Developing tools selection criterion.
- Hosting data warehousing product vendor visits.
- Developing communication resources.
- Developing metadata standards.
- Creating issue resolution process.
- Establishing enhancement-tracking process.

.

7.1.2. HIGH LEVEL BUSINESS REQUIREMENTS

This stage includes a high level of business analysis with Rensselaer executives and decision-makers that will uncover the following:

- Understanding divisions' key strategic initiatives.
- Identifying divisions' key performance indicators or success metrics for each strategic initiative.
- Determining core business processes monitored by each division/unit and want to impact.
- Determining potential impact on division/unit performance metrics with improved access to information.

7.1.3. BUSINESS REQUIREMENTS PRIORITIZATION

Faced with the potential of many demands from all the administrative and academic divisions and units, this task will prioritize all business requirements.

7.1.4. DEVELOP DETAILED PROJECT SCOPE AND PLAN

Based on the prioritized list of requirements, develop project scope and plan.

7.1.5. BUILD TECHNICAL INFRASTRUCTURE

This task will include procuring and installing base system hardware and software needed to implement the data warehouse architecture. It includes:

- Developing technical architecture plan.
- Initially re-allocating of existing data base server and Oracle 8i installation.
- Selecting and installing front-end tool.
- Purchasing and installing Web and application servers.
- Selecting and installing extraction transformation and loading tool.

7.2 Implementation Phase   Jan-02 – Aug-04

The iterative approach will be used as following: For every subject area, there are six basic groups of activities described below. The actual implementation time of each subject area will vary based on the complexity of the subject areas defined during the USER REQUIREMENTS task stage. On average, for planning purposes, between four (4) to six (6) months should be allocated for every subject area.

7.2.1 USER REQUIREMENTS

Business users and their requirements impact almost every decision made throughout the implementation of a data warehouse. This includes all the activities necessary to get a full understanding of the business.

- Selecting interviewees and conducting interviews.
- Conducting data audits and analyzing results.
- Documenting findings.
- Publishing requirements deliverables.
- Prioritizing and revising scope where necessary.

.

7.2.2 MODELING
There are two major activities that are followed by the gathered business requirements and data audits—the logical design of the subject area and the detailed analysis of data sources and the transformation rules:

- Dimension modeling - Translating business requirements into star schema designs and evaluating and designing aggregate tables.
- Analyzing data sources - Identifying sources for information and the transformation rules.

7.2.3 PHYSICAL DESIGN
Similar to building a house when the blue prints are turned into the actual house, this task turns the logical modeling into a physical database.

- Physical data base design - Will include such tasks as defining standards, estimating data base size, developing index plan, developing partitioning plan, and designing physical tables and columns.
- Physical database implementation - Will include such tasks as optimizing DBMS parameters, building physical storage structure, completing table and index sizing, and creating tables and indexes.

7.2.4 DATA STAGING DESIGN & DEVELOPMENT
This stage is the core of building the data warehouse. During this task, the processes that load data from the operational systems into the data warehouse are designed and developed.

- Data staging design and development – Designing high level staging process; developing detailed staging plan by table; defining and implementing staging metadata; developing initial load; developing incremental loads; designing and implementing data cleansing and aggregation processes; developing data quality assurance process.
- Populate and validate database – Setting up production environment; loading initial data; loading historical data; performing data validation and quality assurance processes.
- Performance tuning – Setting up benchmark queries; reviewing indexes and aggregations; performing tools tuning; monitoring database.

7.2.5 APPLICATION TRACK
This is the time when the general front-end templates are designed. If necessary, standardized report frames are created. The actual time will vary based on the subject areas defined in the USER REQUIREMENTS DEFINITION phase. For planning purposes, an average of between one 1) to two (2) months should be allocated for every subject area.

- End user application specification – Categorizing and defining the end-user delivery approach (Executive interface vs. direct tool-based vs. Web-based.)
- End user application development – Designing and developing application templates and executive interfaces.
- Train core business users – Starting training core business users.
- Metadata maintenance – Populating metadata.

7.2.6 SUBJECT AREA DEPLOYMENT
This is where the subject area data warehouse is released to the business community.

- Deployment planning – Developing user education and support strategy; developing course materials; developing release plan.
- Conduct complete system testing – Running complete system testing from staging processes to end user application templates.
- Deployment – Configuring and testing desktop infrastructure; setting up security; educating users.

.

## 7.3 Growth Phase   Dec-04 - ongoing

As stated previously, maintenance and growth of the data warehouse will require a different approach than traditional operational systems. In a traditional operational systems development, the deployment simply means that the project is done and the staff is available for the next implementation. Typically there is no decline in the ongoing needs during the data warehouse maintenance and growth phase and, unlike most systems development projects, the data warehouse is never done. It will be necessary to:

- Provide ongoing support.
- Provide ongoing user education.
- Maintain technical infrastructure.
- Monitor end user query performance.
- Monitor data staging performance.
- Monitor ongoing success.
- Identify changes and enhancements.
- Identify needs for new internal or external data sources.
- Identify needs for new subject areas.

Data Warehouse Project Organization

As stated before, data warehouse implementation is more than just hardware and software. The staff needed to carry out the data warehouse implementation is described as following:

Data Warehouse Project Organization
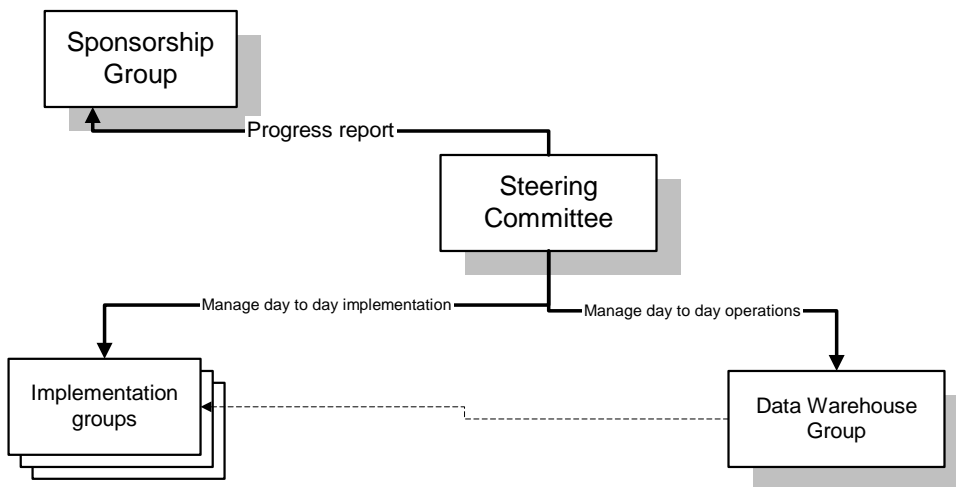
Figure 1

Sponsorship Group
The members of the Sponsorship Group are the business owners of the project. The group members are sharing the vision for the impact of improved access to information and helping to make and support key project scope decisions. The Sponsorship Group will review and approve plans for the development of data warehousing initiative. This Group will approve policies regarding access to data, data security and confidentiality.
The Sponsorship Group will:

.

1. Enhance institutional effectiveness by developing and expanding the integrated use of management information across the Institute; and by expanding the required technological capabilities to support this activity;
2. Define user requirements and important technological capabilities;
3. Make policy recommendations regarding data access, data security, and data confidentiality;
4. Ensure proper implementation of the overall strategy and institute-wide architectures in the key portfolios;
5. Review development plans;
6. Expedite decisions and issue directives in their respective organizations, as needed and as appropriate.

Members: Virginia Gregg, VP for Finance (Co-Chair); Dave Haviland, VP for Institute Advancement; Eddie Knowles, VP for Student Life; John Kolb, CIO (Co Chair); Bud Peterson, Provost; Curtis Powell, VP for Human Resources; Art Sanderson, VP for Research.

Steering Committee
The Steering Committee's global responsibility is to *ensure the proper implementation of the Data Warehouse initiative within designated time frames and budget.* The Steering Committee will ensure proper communication among and between supporting teams, and to the general public.
The Steering Committee will:
- Develop the overall implementation strategy;
- Control project scope;
- Control project budget;
- Communicate and coordinate with the following established campus groups: Banner Core Advisory Group, Web Stakeholders Group, Network and Security Group;
- Nominate and manage implementation groups (front-end tool selection, subject area specific groups, end-user training group, standards, etc.);
- Identify, resolve, or elevate policies issues;
- Request support as necessary to ensure that resources are available to meet implementation requirements;
- Work with other groups on campus for example, Purchasing to monitor relationship between vendors and Rensselaer Polytechnic Institute;
- Provide bi-monthly status reports to the Sponsorship Group;
- Prioritize change requests and present change requests to the Sponsorship Group;
- Recommend the allocation of resources;

Members: John Bradley, Director IACS (Co-Chair); Ora Fish, Data Warehousing Project Manager; Gary Gabriele, Vice Provost; Sharon Kunkel, Director of Student Records & Financial Services, Registrar; <currently vacant>, Controller (Co-Chair);

Implementation Groups
Several Implementation Groups will be formed through out project implementation. The Implementation Group will consist of the small group of campus constituencies charged with specific goal and timeframes, such as choosing the front-end query tools, identifying the information requirements, testing and approving data-marts rollouts.
Possible Members: Jackie Ellsworth, Associate Director of Financial Operations; Helen Grzymala, Associate Director of Financial Planning and Budget; Trish Lyons, Director Academic Budget & Planning; Jack Mahoney, Director Enrollment & Institutional Research; Donna Tomlinson, Manager of Fin. Operations - School of Engineering are just a few members that could potentially serve on the various implementation teams.

Data Warehouse Group
The Data Warehouse group bears the bulk of the responsibility for the design and development of the data warehouse. The following staff will be assigned or hired as necessary to carry out the data warehouse implementation:
Project Manager – Ora Fish, Project Manager. Is responsible for overall project definition, planning, and management. Manage day-to-day data warehouse implementation.
Institutional Researcher (1 Open position) – is responsible for facilitating the collection, analysis, and interpretation of institutional data.

.

Business Systems Analyst (3 positions: Kate Owens, Business Systems Analyst, 2 Open positions) – is responsible for representing business requirements through data modeling, designing and developing the staging system, and the transformational processes.
Data Warehouse Architect (1 Open Position) – is responsible for the overall back-end architecture.
Data Warehouse Analyst (2 Open Positions)– is responsible for developing and implementing project communication plans and the overall metadata accuracy and maintenance. Conducting periodic training in front-end tools and Rensselaer-specific data.

9.0 Budget

The Data Warehouse Project will require investment in hardware, software, labor, and services over a three and one-half year (3 ½) implementation period.

Hardware and Software

The data warehouse architecture consists of a back-end AIX server running Oracle database, a front-end server running Web services, and application services. There are two basic application services employed in data warehouse architecture: 1) the Extraction, Transformation, and Loading (ETL) application used by the technical staff for extracting and loading data into the data warehouse and 2) the Online Analytical Processing (OLAP) application, which provides front-end analytical tools for data analysis.

Labor and Operating Expenditures

The labor portion of the budget includes staff salaries described in the previous section.

Other Services

This category includes the following:
Educational and consulting services – This includes educational and consultant services offered by the front-end tool vendor. These services will be very important during the second year of implementation, assuming that later on during project implementation some of these services will be offered internally.

Technical Consulting and Travel and Reimbursement – It is estimated that it will take ten (10) days for technical consulting in the use of Extraction, Transformation, and Loading application.

Conference, Training, and Education Fees – This includes technical staff training and education in the area of data warehouse.

10.0 Project Risks and Risk Reduction Plan

As with any implementation project, there are several risks associated with implementing a data warehouse. Early identification and planning can substantially reduce any adverse impact on the project.

Hiring and retaining staff - Our current compensation levels cannot compete with other industries. This will greatly impact our ability to attract and retain talented people.

Most of the people will be new to the project - Consideration should be given to the initial group dynamics.

.

Personnel turnover in Finance - The Finance division's personnel turnover can highly impact this project in terms of functional resources available. One alternative could be to postpone the implementation of the financial subject areas.

Current analytical culture – Creating a data warehouse is all about providing improved access to better information to support decision-making. This can cause a cultural change as to how we place value on information and analysis.

New technology and skills set learning curve - The implementation of the data warehouse is very different from purchasing and implementing ERP systems. At its core it relies on new Web-based technology and a new set of products. The technical staff will need to have analytical skills, acquire new modeling techniques, and acquire data base performance tuning techniques. This risk can be reduced by providing the necessary training and allowing enough time for the first subject area's implementation.

Continued campus cooperation and support - Data warehouse implementation success relies heavily on the availability of the business users to specify their needs, evaluate and test front-end tools, perform data testing, and perform application testing. An effort should be made to reallocate some of the everyday activities performed by the business users to free up their time.

Addressing a data access and data ownership policy and defining the implementation procedures -Currently we don't have policies regarding to data management, data ownership, or data distribution.
Strong ties to the existing core committee can help in addressing these issues.

Office space – Options can include finding space someplace else on the campus and/or telecommuting.

11.0 Future Sustainability

The following functions have to be supported in the future:

1. Offering a broad array of decision support services to all units at Rensselaer.
2. Facilitating the collection, analysis, and interpretation of institutional data.
3. Providing information to support planning and decision-making.
4. Promoting standardization in data definitions and classifications.
5. Coordinating with the campus community to establish policies, standards, procedures, and roles regarding data management and data access issues.
6. Monitoring and addressing data integrity issues.
7. Serving as an information resource, providing an easy way for Institute constituencies to identify what information exists and how to obtain it.
8. Keeping the data warehouse current and relevant to the Institute's ever-changing environment. Re-evaluating the data warehouse content by monitoring access levels and conducting usage surveys. Maintaining on-going communication with the schools and Institute administration. Conducting ongoing needs analysis. Supporting software upgrades. Implementing new information access initiatives.
9. Providing adequate support to the end users. Delivering ongoing training and support in data warehouse tools. Maintaining Web-enabled end user documentation and metadata definitions.

.

12.0 Conclusion

Rensselaer, like any major business, collects and maintains huge volumes of pertinent data. Although most of this information is gathered to fulfill a specific unit's mission, such as payroll, registration, or financial aid, it also represents a strategic asset of the entire University when used as a decision-support and reporting tool. The degree to which Rensselaer will leverage this strategic asset depends upon our ability to deploy the information into the hands of decision makers.

The true empowerment comes when people have access to information and can perform analysis for effective decision making. Data warehouse architecture will facilitate easy access to information, but alone it will not address all existing issues. Data Warehouse architecture has to be combined with a broad array of decision-support services to all units at Rensselaer. **To paraphrase President Jackson, "How can we continue with our planning process if we don't have all the information readily available to us?"**