Institution:
The University of Missouri System consists of four distinct campuses: University of Missouri - Columbia, University of Missouri - Kansas City, University of Missouri - Rolla, and the University of Missouri - St. Louis. In addition to these campuses are the System Offices and a University Outreach and Extension Division. Total enrollment at all campuses is around 54,000 students enrolled in undergraduate, graduate and professional programs. At each campus and for each production system, data custodians are responsible for maintaining the data. The Management Information Systems (MIS) group is responsible for helping the data custodian maintain these systems at the system level.

Abstract

Few formal discussions concerning the development of data warehouses mention the organizational dynamics that can hinder or destroy such projects. The University of Missouri System began building a data warehouse for analytical reporting and ad hoc analysis more than five years ago. Though this warehouse project has been successful, the road to success would have been greatly eased had potentially negative organizational dynamics been identified early in the project. These organizational issues included the relationships between the major groups involved in the project: the data custodians at the four semi-autonomous campuses who are responsible for the data in the legacy systems, the University’s Management Information Systems group that manages the hardware and software for the campuses, and the University’s Office of Planning and Budget that gathers data and provides information to University executives and Curators. The autonomous nature of the organizational structure presented special and unique concerns about data integrity and differing policies within each legacy system and campus. Conversely, a major key to the project’s success was having developers and users work closely, cutting across organizational lines. This presentation chronicles the development of the project and discusses those organizational issues that threatened the project and those that were key to the project’s success.
The Early Road Maps: Introduction and Background

The University of Missouri includes four fairly distinct campuses. The largest campus, in terms of staff, enrollment and program offerings, is a residential campus in Columbia. Another residential campus, in Rolla, is primarily focused on engineering and mining. Two urban campuses are found in Kansas City and St. Louis. There is also a large Outreach and Extension program offered cooperatively through the System Office and the campuses. The System Office itself is also in Columbia.

Throughout this affiliation of institutions, there is a mixed bag of centralized and decentralized data and information systems housed on multiple platforms. For example, the personnel/payroll and accounting systems are mostly centrally maintained and business rules are, for the most part, centrally defined. However, the system has enough flexibility to permit some campuses to have some unique business rules. Each campus has its own student and section databases. Though the campuses receive centralized database maintenance assistance, they are responsible for their own data files. While the campuses also use the same software and follow the same basic data designs, there are no mandatory business rules or data definitions at the system level. Each campus has defined, or occasionally redefined, business rules and data definitions for their databases beyond the business rules inherent in the database structure or design.

With each major system, data custodians are responsible for data access, integrity, and providing end user support. For the student system, each campus has its own designated custodian who reports either directly or indirectly to his respective chancellor. Their primary customers are students, cashiers, academic department chairs, and campus administrators.

As part of the System Office, the Management Information Systems (MIS) provides application and database support to the data custodians. This group acts primarily as a service organization to the data custodians and to others at the system level. Also at the system level is the Office of Planning and Budget, where the system institutional research function is found. This office is responsible for all state and federal reporting, system and campus level budgeting, and responding to governing board requests for information. At each campus, there are formal or informal institutional research offices that either directly or indirectly report to their respective chancellors.

Before data warehousing was formally discussed at the University, an all-too-familiar information gathering environment existed. At the campus levels, institutional researchers used shadow systems or they would go directly to their respective data custodians for new data. When these requests came to the custodians, campus staff provided computing support, or if the request dealt with system issues, MIS may have provided the support. At the system level, shadow systems were also used. If these shadow system(s) could not answer new requests, additional data requirements would be sent to the respective custodians who would in turn ask MIS to
provide application development. These requests included a variety of data formats. Sometimes, paper reports were requested, such as the state and federal reports that provided the “official” counts. In other cases, raw data files would be requested for special ad hoc studies or to develop shadow systems for further reporting. These requests and shadow systems at the campuses and system were independent of each other and were without common definitions. Naturally, a great deal of time was spent reconciling reports to other reports. The notion of corporate data was a foreign notion in this environment.

**Early Attempts at Fixing Potholes: Data Warehousing**

The University’s data analytical warehouse has been coined the “University Integrated Data System” or UIDS. The project of warehousing data at the University had intermittent starts and stops dating back into the mid-1980's. The early initiatives had the same view of what the warehouse should look like: frozen data, derived fields, non-volatile, integrated, aggregated, time variant, timely, subject-oriented, and supporting executive decision making, common in today’s literature (e.g., Porter and Rome, 1994). These early efforts were the beginnings of defining data at the University as an institutional resource (Frost and Gohsman, 1993). We now are beginning to realize that the data are far beyond a resource, but are an institutional asset that like other assets, needs to be developed, nurtured, and continually reinvested in.

**The Early Years of UIDS**

This most recent warehouse effort began in 1989 with the identification of legacy system data that would be useful in response to the burgeoning number of ad hoc questions. The primary participants in these early meetings were the campus institutional researchers and representatives from the system institutional research office. By early 1990, the data fields had been identified and preparations were made to bring this list to the legacy system data custodians. However, by mid-1990, organizational dynamics were coming into play that would continue to affect the project.

There were many real and perceived threats that some campuses had with the development of such a system-wide data base. These concerns were raised early during initial meetings with campus representatives and revolved around the ability of the system office to understand the nuances of each campus and to treat the data differentially according to each campus. It was not yet recognized that each campus had, in time, developed legacy systems that on the surface seemed compatible, but were not. Because of these early suspicions, the development of the warehouse was beginning to be viewed as a “turf” battle rather than as a tool that could help both system and campus reporting efforts.

Simultaneously with these mid-level attempts at cooperation, the warehouse concept was endorsed at higher levels and was presented to the governing board. Thus, despite their early misgivings, the campus representatives continued to participate in the warehouse, but in a more limited manner. However, the warehouse was quickly perceived at the campuses as a system office project that would, rather than help the campuses, cause them additional work verifying
data. It was in this ongoing environment of mistrust that the project entered its second phase - a review of the data elements with the data custodians.

Computing at the University had through the years developed into a mixture of centralized hardware and software support and decentralized data definitions and business rules, even with the systems that were in appearance maintained by system level data custodians. Also, each major production system had been developed independently. These systems were integrated using customized programs that “forced” each system to talk to each other. However, there were no effective plans for integrating data at operational levels. In this environment, the role of the data custodian was paramount and his or her support was crucial to any plans for extracting production system data.

By the time the data custodians were asked to review the data elements, the project had lost most of the early campus buy-in. For the custodians, the project was an infringement upon their responsibilities for gathering and interpreting data. The project was additional work for them without immediate benefit. To the campus representatives, differences in campus business rules and definitions would need to be resolved. These campus differences reflected differing campus and school policies that had grown over the years. At the system institutional research office, these differences and their causes were not well understood and were viewed as issues that others should rectify. However, the campus custodians did not perceive these issues as problems since they correctly reflected campus policies. The custodians would provide the data and, since they were not being asked to help interpret data, they were not going to volunteer to resolve campus differences. How these campus differences were to be resolved was never formally addressed. The turf battle was being waged covertly rather than overtly and quiet acquiescence had become the order.

Caught in this void, MIS found itself between its two most significant customers - the data custodians and the system institutional research office. With neither the custodians nor the institutional research office taking leadership in resolving issues of data consistency and compatibility, MIS was faced with the task of negotiating nonstandardized data and differing business rules. Also, at this time a pilot project with IBM was initiated to carry out the data warehouse. Along with battling data definition and business rules issues, MIS now had new software and hardware with which to battle.

Road Test Ahead: The Pilot Project

A two-year agreement between IBM and the University to develop a client/server warehouse was reached during 1991 (Chatman, 1993). This agreement required MIS to implement a client/server technology that was new to the University. Also, the plan featured a new database platform with which no one at MIS had experience. During this time, three new staff members came on board at the system level to work on the warehouse. One was to work with the institutional research office and the other two with MIS. All three were unaware of the group dynamics already in place, but recognized some characteristics of the turf battles taking place. During this pilot project, relationships tended toward finger-pointing and a failure to understand the various organizational perspectives of the participants.
However, several positives came from the pilot project. Many campuses’ data definition and business rule differences were identified and resolved. Despite the organizational problems that had more than once developed into personality conflicts, valuable working relationships developed between end users, MIS, and the custodians. Getting end users and developers to discuss needs and expectations may not appear earthshattering, but for this project it was a milestone. For example, the need for data modeling was not well understood by the end users and only limited staff time was committed to this task. It was only after experiencing the performance differences between lackadaisically-tested data models and those more thoroughly tested, was this activity given the importance it needed. The experiences with this first client/server product and relational data base for MIS and system institutional research offices were indispensable to UIDS. Also, despite the costly investment in computing and staff resources, many became committed to the development of UIDS and were resolved to provide a solution despite the “gloom and doom” perspective of early participants. Additionally, the vice president to whom MIS and the institutional research office reported was committed to seeing the warehouse implemented successfully.

### Problems With the Road Crew: What Was Going Wrong

**Users wanted the data dumped and they wanted it yesterday**

The end users wanted to get in a race before they had even learned to walk. Traditionally, the institutional research office was using standard paper reports and ad hoc flat files created for singular purposes. Rather than develop experience integrating and using prototype data, it wanted to begin using the new, user-friendly warehouse to generate reports immediately. Likewise, in this rush to gain unlimited access to data, organizational issues were seen as obstacles to be dealt with later, rather than issues that required immediate and serious attention.

During the early period of the pilot project, the preferred method of warehousing was to “load the data and then shake it loose.” For both the end users and to some in MIS, the resource commitment to data modeling and testing seemed unnecessary and too time-consuming. Client/server architecture and the concept of data warehousing were in their infancy, and it took some time for those involved in the project to appreciate the need for understanding the data before dumping it into the warehouse. The traditional process of application development did not fit this new environment, yet everyone was skeptical about the usefulness of rapid prototyping, particularly while significant data issues were still being addressed. When prototyping was completed, there was little feedback from the end users or data custodians. While the custodians’ roles were paramount in extracting raw data from the legacy systems and shaping it into something that could be warehoused, their recommendations concerning warehouse design and content were deemed unnecessary. The end users, confused about their role in prototyping, requested that more data be added to the system instead of using prototypes to validate the design and learning how to use the data.
Project Management

At the onset, many wanted to view this project as a cooperative, joint adventure. As a result, the issue of project management was discussed, but never sufficiently resolved. It seemed no one was willing to take the initiative or to confront the group by demanding the leadership role openly. This group dysfunction blocked meaningful discussions of such issues. The lack of firm project leadership was a hindrance and contributed to the project’s fuzzy focus on goals and a lack of understood roles for developers and end users. This lack of a leadership function further acerbated many misunderstandings between the end users and the development team.

Organizational Dynamics

The UIDS project lacked some characteristics that have been found in other successful warehouse projects. Notably, a positive culture did not exist such as that of the Arizona State University’s warehouse project (Porter and Rome, 1995). Nor was the project directly related to a specific business goal such as with the University of Massachusetts’s warehouse project (Jackson, Hammon, and Smith, 1995). Also, the project was not constrained to a more limited “game plan” (Mundy and Thronwaite, 1995); it was to be all, immediately. The University’s organizational dynamics had created a non-lethal turf war that made delineating differences in the legacy systems’ business rules and ensuring data integrity monumental tasks. Many organizational issues and differences in perspectives were either ignored or given only cursory attention. Personality conflicts added negative dimensions to the project, making it seemingly impossible to continue and unlikely that organizational issues would ever be delineated, let alone resolved.

Repaving the Road to Data: Solutions and Work Arounds

Most of the organizational issues brought to the surface during this project have been left unresolved. Some have eased due to staff attrition and reassignments. However, there are some things done that can benefit other institutions that are attempting projects faced with similar, seemingly insurmountable organizational conflicts.

Recognition of Issues

In a culture of semi-independent campuses and opposing subcultures, many issues cannot be logically resolved. For example, there is still little effort at the system or campus levels to initiate standardization of business rules and polices. As such, data ownership issues will always be part of the project. Because of unresolved organizational issues, a number of processes changes were made to ensure continued success.

Mechanisms are now in place that call attention to unilateral changes in business rules or data definitions. The data are constantly validated and edited to insure accuracy. Rekindling campus buy-in to the project has been a slow process. The campus data custodians have expressed little interest in using the data themselves since UIDS was developed with little of their data needs in mind. However, some campus institutional research offices have begun using the
data and are quickly becoming UIDS’ best customers. There are two main reasons for their increasing use of the warehouse. Perhaps most important, the data design lends itself to being user-friendly, intuitive, and readily adaptable to many ad hoc questions. Secondly, the data reflect “official counts” and when reporting from the warehouse, there is less concern over how the “pie might have been sliced.” Increased use of the warehouse throughout the system has sparked others’ desire for access.

**Access and Training**

Access and Training

Keys to gaining acceptability of UIDS have been demonstrations and training sessions presented at the campuses. These demonstrations have been given for most all organizational levels and attendees have ranged from system vice-presidents and chancellors to information assistants. Half-day to day-long training sessions have also been presented at three of the four campuses. To bridge the gaps between the central office and the campuses, the UIDS project has treated the campuses as potential customers. Not only is a product being “sold,” but so is a service. Campus-initiated questions and requests are now treated as priorities.

At this time, no attempts have been made to deny access to the database to legitimate users at the campuses. The UIDS warehouse has been the first information system at the University that embodies the notion of corporate data.

**Functional Decentralization**

Functional Decentralization

Perhaps the most significant change that salvaged the UIDS warehouse has been the dismantling of many organizational walls between the system institutional research office and MIS. For the most part, this was accomplished more by perseverance than grand design. The main developer and an institutional researcher felt a great commitment to the project and its eventual benefits for the system and campuses. This dedication was also supported by their supervisors.

Following the completion of the initial round of data loading, including data from the legacy systems for the past five years, the principal developer was physically moved near the institutional research office. Conceptually, this is similar to Syracuse University’s “distributed positions” for information technology (Borel and Vincent, 1995). However, this distributed technical expertise is more a functional than organizational distribution. The developer’s expertise is not limited to helping system institutional research, but to all UIDS users and related requests. This type of functional decentralization has benefits that outweigh some costs and problems normally associated with decentralization (Solomon, 1994).

**Premining**

Premining

One the most important lessons learned from this project was the need to follow the “sandwich paradigm” of data warehouse building (Parsaye, 1995). The notion of premining refers to acquiring a basic understanding of the data before warehousing. In this process, data are premined and prototyped, followed by additional analysis or mining to validate the design and
data content. User feedback in this process is essential and further reinforces the need to have functional decentralization. Having developers and end users working closely in the premining, prototype, and analysis stages is crucial to the success of the warehouse project. During the early stages of UIDS, this type of process was foreign to the end users and was a difficult concept to learn.

Data Structures

The particularities of the University’s organizational dynamics called for the creation of several unique data structures. The UIDS warehouse is a mixture of extracted data, derived data, and data transformations that exist only in the warehouse. For example, each campus had a unique method of identifying first-time students based on several fields in the legacy system. UIDS has a flag (Y/N) that reflects those campus-dependent business rules identifying first-time students.

One unique feature of the UIDS warehouse is that it contains its own integrated department structure. This mythical department structure was created to integrate student, financial, and personnel data together and provides a way to tie students to money, staff and faculty. Though the UIDS department structure is time dependent, the physical table in the database has been structured to allow one to track a department’s data not only across systems, but also through time.

Continuing Operations

Data from the student, section, personnel, and financial accounting legacy systems are continually loaded. Finance and personnel data from FY1992 and student and section data from summer 1992 are loaded into the warehouse. Student data tables include data on degrees, end of semester performance, assessment tests, and demographics. Section tables contain data ranging from general section characteristics and instructor information to student level credit hour production. Finance data are divided into two main tables, revenues and expenditures. Personnel data are organized into five tables containing demographic, appointment, account distributions, titles, and departmental data. Each table has a unique level of detail. Student data have the individual student as its level of detail. Section data are aggregate at the section, student, and instructor levels. Revenue data are aggregated at the departmental, division, and campus levels while expenditure data can be looked at by subaccount, account, department, division, or campus levels.

The warehouse currently uses a client/server environment. The data are housed in an Oracle relational database on an HP 9000 server. For “middle-ware,” SQL*NET over TCP/IP is being used. All clients are Windows based and use InfoMaker to access data and to create reports. A data dictionary is being created on the web and is expected to be completed shortly. The data dictionary can be found by pointing the web browser to: http://astro.umsystem.edu/uids/uids.html.

The University is currently undergoing a process of creating data warehouses for the legacy systems. This effort seeks to get data directly into the hands of those in need. The data are loaded raw into these warehouses and the data are refreshed periodically. Since UIDS
occupies a distinctly different niche, having frozen data, the two types of warehouses should complement each other. Since both use the same user interface, InfoMaker, end user knowledge is not duplicated or wasted.

What’s Next? Or ...... “Experience is being able to recognize your mistakes when you make them again”

Gradually, the warehouse is being expanded. Grant and contract data are being prepared for loading into the warehouse. The lead developer and end users have worked closely with the data custodians in this project. The responsibility for preparing the data to be loaded has been assumed by the custodians and their buy into the project seems genuine. Student Financial Aid will be loaded next. In an attempt to counter opposing organizational dynamics previously discussed, an entirely new approach will be used during this phase. At each campus, student financial aid data are stored separately from the student data. The attempt will be to offer some functionality to the financial aid offices such as ease of data access, consistency with other data, and integration of data with the other major UIDS tables. The last area to be incorporated will be external data to ease the development peer institution comparisons that are now time consuming and tedious processes.

A Road Map to the Scenic Highway to Data: Conclusions

Just as there are normally several roads to get from one city to the next, several roads can be taken to obtain needed data. As with transportation, some highways seem fraught with potholes and unexpected turns. In developing the UIDS warehouse, several techniques and critical success factors were used to help pave the road surface.

The most critical factor was the perseverance and dedication of the end user and primary developer. Without this commitment, the project would have been abandoned at the first road block. Secondly, the recognition of competing interests and organizational dynamics is paramount if they are to be resolved or successfully dealt with. Many issues cannot be resolved unilaterally; they can however, be understood to the extent that they do not become mere personality conflicts and agents of destruction.

A clear understanding of the roles of key players including users and developers is another critical factor. Without this recognized delineation of roles, expectations and reality have a difficult time meeting themselves. Coupled with an understanding of roles, project leadership is crucial. Since the project is still without formal executive leadership, functional roles have developed where the leadership “hat” rotates between the lead developer and end users. Had this type of leadership been used earlier in the project, many confrontations could have been avoided or resolved more easily.

The need for the buy-in of the key participants in the project was missing in this project
for several years. The project required significant resources of many who saw no immediate benefit from the project. Had these key players been able to participate early in the design of the project, their active support could have easily filled many potholes.

In getting buy-ins after the fact, data validity has become paramount. Constant verification and validation processes will continue to be a part of UIDS. Treating the users of this warehouse as customers has required delivery demonstrations and training sessions that perhaps are better described as sales pitches. Constantly touching base with users to discuss problems or issues is also key to having a quality product.

Premining, warehousing, and mining have become the buzz words for future developments and additions to the warehouse. The front end time spent developing data models may seem unnecessary, but compared with the time spent later fixing problems or changing years worth of data, the time is well spent. The most important, and rewarding technique employed with UIDS was getting the principal developer and end user working together. Thus, decentralized development and support have their benefits. It can force communication across damaged or broken organizational lines and can provide an effective means for watching out for the potholes in the highway for your data. Like highways, data are resources and assets that need constant maintenance, renewal, and development.
References and Works Cited


