SAS\textsuperscript{(R)} Software and the Evolution of an Information Delivery System
Hugh G. McCabe, Syracuse, NY

ABSTRACT

This paper describes the use of the SAS System by a nationally accredited managed care organization to meet its ad hoc reporting and decision support needs. The paper will provide executives with an example of how an organization can use the SAS System to manage the problems associated with diversity in its data sources, applications, computing environments, and user requirements. Discussion will focus on the evolution of methods to deliver information in the organization during the last six years, and current work on the development of a corporate data warehouse. Emphasis will be placed on the conceptual framework for information delivery and the emergence of the data warehouse in the data base management industry.

INTRODUCTION

Health Services Medical Corporation (HSMC) is a health maintenance organization (HMO) servicing central New York and the Southern Tier of central New York. When I wrote a paper for SUGI 15, HSMC was operating four lines of business; a group model HMO, a network model HMO, and two independent practice association HMOs which were owned by the local Blue Cross and Blue Shield Plan. Our total plan enrollment was approximately 100,000. Since that time there have been some significant changes at HSMC, in the local managed care market and in the data management industry. This paper will describe HSMC's use of the SAS System to meet its information delivery needs through a period of downsizing and diversification.

THE EARLY DAYS

Prior to the introduction of the SAS System at HSMC most reporting was done by the HSMC Management Information Systems Department (MIS). Research consisted primarily of posting data from MIS reports into spreadsheets for further manipulation, analysis and presentation. This process presented a number of problems. First, manually posting summarized data was time consuming rework and was error prone. Second, additional analysis usually required data at lower levels of aggregation or detail data. For example, it was not unusual to study referrals for all medical specialties, followed by more focused analysis on one specialty, the physicians in that specialty and the procedures that those physicians performed. This additional analysis would require more requests for data and additional posting of data.

Recognizing the need to transfer and analyze data more effectively HSMC purchased the SAS System for PCs in 1988 and began to develop methods for transferring data from the online transaction processing systems (OLTP). Here, large summary files and detail level extract files from the OLTP systems were transferred to an analytic database on a PC in our research department. Transfers of data were done on an ad hoc basis.

From 1988 to 1990 this process enabled us to provide decision support of a limited scope to senior management. As demand for decision support and management reporting from the analytic database grew so too did the need for greater disk storage, memory and processing capability. The growth in demand grew not only in terms of volume but also in level of complexity.

In 1990 we purchased a SUN Microsystems Sparcstation 1 workstation to enhance our disk storage, memory and processing capability. Once the workstation was operational we installed the SAS System for UNIX and moved all of our data and SAS code from the PC to the workstation. It should be noted that all of the software developed using the SAS System for the PC was fully operational on the workstation running under UNIX.

Between 1990 and 1994 we expanded from the single workstation to a local area network with six workstations. OLTP data are downloaded from the MIS minicomputers each month. The types of work being done using the SAS System include: processing of monthly downloads (clean-up and validation), basic and complex data manipulation, ad hoc and scheduled reporting, and elaborate applications for user data queries.

Changing Enrollment

As I alluded to earlier, this investment in hardware, software and staff took place during a period of downsizing and diversification. From 1988 when HSMC first began to use the SAS System to meet its information delivery needs, to 1992, HSMC saw its enrollment swell to approximately 100,000. In January of 1993 HSMC lost its management contract for two lines of business, effectively reducing enrollment by 50,000. As of February 1995 our enrollment is back up to 85,000.

Diversification and Competition

In addition to dramatic changes in enrollment over the last six years, HSMC has seen significant changes in the Central New York managed care market. Each of these changes has added to the demand for information delivery at HSMC. Historically HSMC has been the dominant managed care plan in Central New York. Over the last six years several large national and regional competitors have entered the market and this dominance has been challenged. Along with greater competition our employer purchasers have demanded that HSMC diversify and offer a wider selection of health insurance products. This diversification of insurance products includes; traditional indemnity insurance, point of service insurance, self funded insurance and expanded networks of physicians.

As a managed care plan and manager of a physician group practice in New York State, HSMC is regulated by both the New York State Insurance Department and Department of Health. With the addition of a small Medicaid population in 1993 HSMC added the New York State Department of Social Services to the roster of state agencies requiring reporting.

In 1994 HSMC was accredited by the National Committee for Quality Assurance, the accrediting body for managed care plans. In addition to demonstrating information delivery capability for accreditation, HSMC completed Health Plan Employer Data and Information Set (HEDIS) in 1994.
At the risk of belaboring this issue I have tried to illustrate how well the SAS System has served us as a cost effective information delivery tool throughout a period of significant change and limited resources. By establishing a separate information delivery system based on the SAS System we have been able to meet a growing demand for information without compromising the goals of our MIS Department. This serves as a point of departure for discussing why we chose this approach.

Here, I will address the question of whether we could have achieved the same level of success by investing the equivalent amount of resources in our existing MIS infrastructure. This is a logical question since the MIS department does already have computers, software, data, and staff. Frequently the problems of using the existing MIS infrastructure for ad hoc reporting are explained in terms of workload and resource availability. These explanations are usually not sufficient and rarely lead to resolution of information delivery problems.

To gain a true understanding of the problems encountered in using the existing MIS infrastructure for ad hoc reporting and decision support you must look beyond the resources. Understanding the problems of using the MIS infrastructure for information delivery requires an examination of: 1) the differences between the people responsible for development and maintenance of OLTP systems and the producers and consumers of decision support information, 2) the nature of OLTP data and the data required for decision support, 3) the structure of OLTP data, 4) the tools available for information delivery in OLTP systems. I will discuss each of these areas in greater detail.

THE INTERPERSONAL DYNAMICS OF OLTP AND DECISION SUPPORT

There are a number of issues pertaining to the interpersonal dynamics of OLTP and decision support which provide insight into the vexing question of why it is difficult to use the existing MIS infrastructure to do decision support. In his 1985 SUGI paper, James Knoop of Aetna Life and Casualty addressed this dynamic by analyzing the conflicts that arise between the SAS user and the data processing professional. (Knoop, 1985) This analysis is instructive. Although Knoop focuses solely on the SAS user and his/her interactions with the data processing professional, most of his observations are equally relevant to anyone requiring decision support information from the data processing professional. I would argue that Knoop's points are generalizable and expand our understanding why it is difficult to do decision support using the existing OLTP infrastructure.

Knoop characterizes the interpersonal dynamics between SAS users and the data processing professional as that of conflict. According to Knoop this conflict occurs along three dimensions: situational conflict, cultural conflict, and structural conflict. Situational conflicts arise when, according to Knoop, "Equally legitimate organizational objectives can not be optimized simultaneously." Knoop gives the example of the SAS user being required to provide information to management which may not be available via existing MIS mechanisms. Here, the data processing professional is directed to provide the SAS user with whatever support he/she requires to complete the project. (Knoop, 1985)

In this situation or even when the information is requested directly from the data processing professional in a short time frame, the data processing professional might be forced to abandon standard procedures. These standard procedures are designed to support MIS organizational objectives like security or data integrity. Here, the organizational objective will be suboptimized in favor of meeting management's decision support need. Any attempt by the data processing professional to adhere to legitimate procedures may result in delays and will likely be viewed as a failure to respond to a decision support need.

The second type of conflict Knoop discusses is "cultural" conflict. Cultural conflicts are a result of differences in values, orientations, expectations and norms between decision support analysts and data processing professionals. Knoop outlines four different cultural contexts which characterize differences between decision support analysts and data processing professionals. Each of these cultural contexts can be a source of conflict. The first cultural context is the division of labor. The division of labor in a data processing environment is typically complex, usually reflecting a high degree of specialization. It is this specialization in data processing environments which facilitates continuity and control necessary in operational processing. The cultural context of the decision support analyst is quite different. The decision support analyst functions with almost no division of labor. The collection of information for high level decision support may require a single decision support analyst to cross all boundaries in search of data. The decision support analyst must sometimes deal with an entire cast of data processing staff. This can result in conflict from delays and misunderstandings. When decision support requires a quick response this division of labor can result in delays since it can be difficult to rapidly assemble the needed data processing staff. (Knoop, 1985)

The second cultural context which differentiates the decision support analyst and the data processing professional is their orientation. The difference in orientation between the decision support analyst and the data processing professional can result in conflict as the decision support analyst attempts to access the corporate database by working with the data processing professional. The data processing professional is process oriented while the decision support analyst is product oriented. For example, the data processing professional is concerned with systems and vigorously enforced standards and conventions while the decision support analyst is concerned with fulfilling an immediate need and attaining results quickly. (Knoop, 1985)

The third and related cultural conflict pertains to methods and procedures employed by the decision support analyst and the data processing professional. The process orientation of the data processing professional dictates use of structured, consistent, and uniform procedures. These constraints are designed primarily to promote consistency and ease in maintenance. This is contrasted by the decision support analyst's need for flexibility in their search for data to support high level decision making. The decision support analyst is not concerned with structure, consistency, or uniformity because once the project is complete they may never use the information again. (Knoop, 1985)

The fourth and final source of cultural conflict between the decision support analyst and the data processing professional is the difference in their processing cycles. In data processing environments, processing cycles are almost always regular, scheduled, and frequently permanent. The initiatives of the decision support analyst are typically ad hoc, and in spite of their sometimes extremely complex data requirements, may only be used once or twice. One can see how the difference in processing cycles could cause conflict between the decision support analyst.
and the data processing professional. (Knoop, 1985)

The last type of conflict between the decision support analyst and the data processing professional is structural conflict. Structural conflict is common in corporations where the data support analyst has developed extensive resources for conducting high level decision support analysis. These resources might include hardware, software, and data. Structural conflicts arise here when there is a lack of consensus regarding who is responsible for what in given situations. (Knoop, 1985) I have experienced this type of conflict at HSMC. My department has accumulated extensive technical resources and capabilities. We have had situations where managers have come to my department for a report needed in a very short period of time because our MIS department could not complete the reporting fast enough (perhaps for perfectly legitimate reasons). This type of situation can be the source of friction between my department and MIS.

Knoop’s discussion provides a non-reflective explanation of the conflicts which can arise as the data support analyst interacts with data processing professionals in his/her efforts to access the corporate database. These sources of conflict are the same sources of problems related to using the OLTP for decision support.

William Inmon makes some very pointed comments regarding the needs of decision support analysts. According to Inmon, traditional development methodologies employed by data processing professionals assume that system or project requirements can be identified and organized prior to the design and development process. This is usually true for operational processing where functions are repetitive and clerical. The data processing professional usually deals with that which is known, while the data support analyst usually deals with the unknown. The decision support analyst requires the flexibility to deal with “give me what I say I want, then I’ll tell you what I really want.” This is best characterized as “discovery mode.” These requirements are often at odds with the requirements of operational data processing. This orientation to identifying needs ahead of time can lead to frustration in decision support. (Inmon, 1992)

Each of these aspects of the interpersonal dynamics of the SAS user and the data processing professional contributes to our understanding of why it is difficult to rely on the existing OLTP infrastructure for decision support. A second area which takes us further in our understanding of why it is difficult to meet information delivery needs using the OLTP systems is the nature of transaction processing and transaction data.

**TRANSACTION PROCESSING AND DATA**

In their June 1992 *Medical Interface* article David Aquilina and Paul Louiselle discuss some of the limitations of using the OLTP for decision support. Aquilina and Louiselle explain that in managed care, systems were developed for routine operations, management control and decision support. Routine operations include membership processing, premium billing and claims payment. Management control centers around general ledger accounting and financial reporting. Decision support consists of data analysis for administration and executive level decision-making. Aquilina and Louiselle point out that in MIS departments of managed care organizations emphasis has been placed on operational support and management control. This situation is common in many industries. (Aquilina and Louiselle, 1992)

While managed care plans have been successful at data processing for operations and operational decision support, the higher level decision support has not been a true MIS priority. This has been changing. Recent attention on run-away costs in health care has forced governments and employer groups to demand value and quality from managed care plans. In response to these pressures, managed care plans have begun to demand high-level executive information. This type of information would include the following dimensions; price and payment, utilization efficiency, medical appropriateness, medical effectiveness, and conformance to standards. To understand the difficulties of using the OLTP system for decision support we must understand the obstacles encountered when deriving this type of high level executive information by directly accessing the operational database. (Aquilina and Louiselle, 1992)

Aquilina and Louiselle outline four obstacles to decision support when accessing operational data. The first obstacle is data element limitations. Data in the operational database are transactional. They represent administrative transactions and may not be managed in a fashion that reflects real world events, but rather operational tasks. To include data elements beyond those necessary for OLTP would adversely affect operational performance. (Aquilina and Louiselle, 1992)

Aquilina and Louiselle explain that data analysts often create data fields to make analysis more meaningful. While useful in decision support, these types of fields are of no use in operational processing and would not justify inclusion in the operational database. One example that I have encountered, has been the data field for physician visit. This is not the straight forward data element that it appears to be. Strangely enough, in spite of its analytic importance, “visit” is not an entered field in our operational database. Physicians are not reimbursed for “visits”. They are reimbursed for procedures performed and billed using standardized procedure codes. Although standardized procedure codes do exist for physician office visits, some services are provided which are coded in a manner which does not explicitly indicate a visit, but should be counted as a visit. Hence, the visit data element must be derived based on logical business rules. Even though this field is important to the data analyst for high level decision support, it does not warrant derivation and storage on the operational database. (Aquilina and Louiselle, 1992)

Visit is just one example of a “non-operational” data element. There are many others which are used regularly in decision support. There are also others which are created only once to provide decision support information and are seldom or never used again.

The second obstacle to effective decision support using operational databases is data integration limitations. For example, it is common for enrollment, claims, pharmacy and other data to be stored separately on non-integrated systems or even on separate computer platforms. This is problematic because operational processing may not require fully integrated systems, but the decision support analyst may need to integrate information from disparate systems. Integrating disparate data in an operational environment for decision support purposes is often cost prohibitive. (Aquilina and Louiselle, 1992)

This is precisely what we contend with at HSMC. The most obvious example of this problem at HSMC is the separate systems for insurance claims processing (for its enrolled membership) and the system used for managing the physician group practice. The
system for managing the group practice is a billing and appointment scheduling system which includes data for both the HSMC insured population and the fee-for-service population not insured by HSMC. This heterogeneous structure is a result of the absence of an integrated system in the software market years ago, when the enterprise began to sell prepaid insurance. Until three years ago, there was no unique identifier for our pre-paid members common to both systems. Even today it is difficult to link insurance claims information for institutional and specialty care to the information on services provided by our primary care physician (PCP) group.

The third obstacle to effective decision support using operational data is data quantity limitations. High level decision support often requires access to historical data as old as twenty-four months to ten years. Most operational systems are not capable of storing large quantities of historical data. (Aquilina and Louiselle, 1992) At HSMC, operational data includes the most recent twenty-four months of data. We are frequently required to examine data beyond the most recent twenty-four months. For example, one of our industry quality assurance indicators is the percentage of members receiving a cholesterol test in the last five years.

The final obstacle discussed by Aquilina and Louiselle is data reporting limitations. They point out that operational systems typically have a reporting module with certain standardized reports designed to support operational decision making. Some reporting modules also include facilities for custom reporting. But frequently they are not sufficient or flexible enough to support high level or strategic decision making. (Aquilina and Louiselle, 1992)

An additional system obstacle to using operational data is related to storage and processing constraints. As mentioned earlier, operational databases often only support twelve to twenty-four months of data. Even if the operational database has more than two years of data, decision support analysis may require accessing of the entire database or large portions of the database in a single operation. This is different than transaction processing where only a few records in the database are accessed for a single transaction or operation. The need to access large portions of the operational database usually places enormous and unacceptable burdens on the operational systems. (Aquilina and Louiselle, 1992)

DATA STRUCTURE AND ONLINE TRANSACTION PROCESSING

The third area which we must examine to understand the difficulties in using the OLTP system for decision support is the data structure. The most notable data structures found in OLTP systems are hierarchical data structures found in legacy systems and relational structures found in relational database management systems.

The data that are stored in legacy systems present some of the most difficult problems for use in decision support. According to Robert Giordano, most existing hierarchical data structures are not easily understood, which he attributes to the application specific design of data structures. Giordano describes several typical examples of problems interpreting existing legacy data. One of these examples is multiple redefinition of records; commonly found in flat files created by COBOL programmers. A second example of problems with accessing and interpreting existing legacy data is that fields often found in reports generated from legacy database, do not actually exist as part of the data. They are complex derivations of application subroutines. (Giordano, 1993)

Data in relational structures also present some challenges for decision support. In a recent Datamation article Ralph Kimball and Kevin Strehlo describe the characteristics of OLTP databases which limit their usefulness for decision support. Here, Kimball and Strehlo argue that the drive to reduce or eliminate data redundancy in the relational systems has resulted in OLTP databases which process transactions at great speeds but are extremely complex (sometimes containing hundreds of tables). Use of these databases requires an in depth knowledge of the data structure. Queries against these complex databases usually require equally complex structured query language programming (SQL). (Kimball and Strehlo, 1994)

REPORTING IN OLTP SYSTEMS

The final area that we must examine to understand the difficulties in using the OLTP system for decision support is the tools available for information delivery in OLTP systems. As I alluded to earlier, Aquilina and Louiselle point out that operational systems typically have a reporting module that includes certain standardized reports designed to support operational decision making. These reporting facilities do not usually provide the flexibility to do higher level decision support. (Aquilina and Louiselle, 1992)

Kimball and Strehlo discuss some of the deficiencies found in the primary reporting tool in relational database management systems, SQL. Kimball and Strehlo assert that SQL was not designed to be used as a business analysis tool and so it fails to provide the capability to answer some general business questions. Kimball and Strehlo provide several examples of this shortcoming. This would include the inability to do comparisons of data from one period to another, and do a rank ordering display of only the cases near the top or bottom of the order. (Kimball and Strehlo, 1994)

CONCLUSION

This paper provides a good example of how a medium sized corporation has been able to use the SAS System to meet its ad hoc and decision support needs. In addition to presenting a short case study in information delivery, I have discussed some of the conceptual issues surrounding the approach that we selected. These conceptual issues are also related to the emergence of the corporate data warehouse in the database management industry. The process that we have developed for information delivery is a very basic data warehouse. Over the last several years this approach to information delivery has been receiving more attention and some of the more formal database management theories and techniques have been employed in data warehouse development. It is my hope that this presentation of the conceptual issues related to information delivery will assist taking discussion beyond arguments about workloads and staffing.
REFERENCES

Aquilina, David and Louiselle, Paul "Data Analysis for Decision Support." Medical Interface (June 1992) : 42-47.


SAS is a registered trademark of the SAS Institute Inc. in the USA and other countries. *(R)* indicates USA registration.