INTRODUCTION

The Lubrizol Corporation is a specialty chemical company comprised of several business units that use chemical, mechanical, and biological technologies to develop products for diverse markets. To maintain its world leadership as a developer and supplier of specialty chemicals, Lubrizol is dedicated to a strong, ongoing research program. In 1987, consolidated research expenses increased to over $61 million dollars. Technical service costs, principally mechanical testing to qualify customers' lubricants for various standards, were an additional $31 million dollars.

BACKGROUND

In the early 1980's, Management Information Systems and Development (MIS) created databases to collect formulation and testing data for project management, project tracking, and reporting research activity. The formation of a Corporate Statistical Services Department created a need to consolidate historical data in the form of a corporate knowledge base. Senior statisticians and chemists began utilizing the data for designing experiments and building predictive models. Although SAS was available, most of these efforts began in APL or Fortran.

MISSION

In 1986, the Scientific Research MIS group was formed. Their mission was to create a hardware and software environment for the coordinated use of research data. SAS was the choice. It provided an interface to our corporate Database Management System, user friendly screening and spreadsheet capabilities, graphics, as well as portability.

CRITICAL SUCCESS FACTOR

As work began, it became evident that our foundation was weak. The databases were designed for reporting, and were corrupted with inaccurate or incomplete data. Smart systems development was never the intention of research data collection. Performing this new type of detailed data analysis was painstaking and ultimately impossible. Therefore, we took a step back and revisited our data requirements by building a data, function, and process model for Research. The data model documented all data entities used by research. It showed where the data originated and where it is used. The process model relates the data to the specific job functions that creates, uses, and stores the data. The models developed into the design of a newer, stronger foundation, geared toward information analysis, not data gathering. This data-driven approach to systems design was a drastic change in focus whose ramifications are still being felt in the restructuring of the databases, and more importantly, in the organization of the research effort as a whole.

MOVING FORWARD

The first multi-user, mainframe SAS system was called the Predict Formula Database. Its purpose was to create a clean database containing formulations run through engine test sequences. Formulation data are extracted from the reporting database, cleansed through a massive series of edit checking algorithms, and parsed into a horizontal format for statistical analysis. It utilizes arrays, pointers, and powerful proc organized into tightly coupled macros. It also utilizes advanced features of BASE, AF, FSCALC, and DB2 in a batch and interactive environment. The system helped point out the weaknesses of our current databases. This was invaluable information for the designers of our new databases. We also provided our Statistical Services Department with quality formulation data to build engine test prediction models. It's a cost-effective system, since some tests often cost more than $10,000 each to run. The Predict Formula Database System will continue to run until our new foundation, as defined by the Research models, is complete. At that time, its tremendous editing capabilities will be used to clean our existing reporting databases.
The next SAS project implemented at Lubrizol was the Blend Viscosity Predict System. Blends are formulations that must satisfy rigorous analytical tests before an engine testing program is initiated. To increase quality, throughput, and customer service, algorithms were developed to predict analytical test results. Predictions are calculated using coefficients, generated from regressed historical blend data, to predict viscosity tests with amazing success. The system includes a "blend adjustment" option to modify formulations to meet test specifications. The predictions are fast and accurate, but generating correlation coefficients was a time consuming, tedious task. The problem was resolved by the Model Developers Toolbox Utility. The toolbox documented and automated most of the tasks required to generate predictive coefficients for lubricant chemistry. SAS macros were used to generate SQL code for data extraction from DB2 into SAS. The data are manipulated, regressed, and reported using a multitude of SAS procedures. The system has provided the model developer, i.e. the chemist, with an automated analysis tool, reducing the time to develop model coefficients by as much as 80%.

**FUTURE WORK**

Future SAS work includes predictive systems for other types of chemical or elemental analysis systems to utilize, share, and capitalize on the intelligent use of research information. When the new foundation is complete, the models will validate scheduled tests before they are run. If the model predicts failure, the formulation will require further analysis. Eventually, as more rules are defined, the models will adjust formulations, based on the model's predictions. This concept already exists in the "blend adjustment" option of the Blend Viscosity Predict System. The research data will be utilized to their fullest extent. SAS has provided the tools for successful completion of the mission. From the PC to the VAX and IBM mainframes, SAS will provide the "roofing" over the database foundation. Successful chemists will leave their knowledge behind for others to learn.

The critical success factors for this large undertaking are as follows:

1 - A strong database foundation is a must.

2 - A complete "4th GL type" portable software product line providing database interfaces, full screen and spreadsheet capabilities, and mathematical/statistical tools.

3 - Thorough, detailed, all-encompassing data analysis and planning are necessary for success.

4 - Full commitment from not only users, but corporate leaders as well. This effort requires a lot of time and money before tangible benefits are realized.

5 - Patience - Is only achieved by believing the goal is reachable. Opposition, hurdles, and hard work will constantly test your patience. Of course, the larger the corporation, the more patience is required.