

Biotech Warehouse – Stretching the Limits of Columns

**Larry Bramblett, Data Warehouse Solutions, LLC
San Ramon, California**

Data Warehouse development is often focused on specific data marts that support data analysis for revenue reasons. For most businesses, the ultimate goal is to maximize revenue gain with the focus on the customer as a revenue generator. Companies that have a manufacturing component know that while the customer is important, the improvement of a process that cuts the costs of manufacturing the product also impacts the revenue bottom line. Instead of target marketing, or credit management, process improvement and results analyses are the goals of the manufacturing warehouse. This paper outlines a case study of a Bioengineering company and the journey to create a Metadata driven SAS data warehouse.

The scientists in bioengineering are moving from manual processes and systems to highly automated processes and systems. They are driven by a goal to improve both quality and quantity of products. Currently, there is a mix of both types of systems found. While the intent and design of the new, automated systems was to support the existing process, reality drives the process and the analysis of those systems to produce similar but non-conforming information.

Quality is the key concern which mandates that a "single vision of the data" supports the evaluation and analysis of both manual and automated processes within a single application.

SAS was chosen to manage the collection, conforming and transformation of the data into a single data platform. The vision for the project was to create a data model of the bio engineering process that is independent of the type of process and its location. The goal of the model was also to support US government regulations and the FDA SN88 model. While this project was highly directed around a specific model, the metadata driven process design would support other manufacturing process systems as well.

A considerable amount of time was devoted to the process data model design. Team members that

represented various process areas worked together to build the model. The result came from extensive

discussion of each process and each system. The SAS team worked with the client team to provide data as the model was being created, testing each assumption of the model. SAS/Warehouse Administrator provided the platform to extract, transform and create a view of the model that the team could evaluate. During this process it became clear that traditional column and row tables would fail to provide the analysis the end-users, the bioengineering scientists, needed.

The metadata model focused on a specific product and its data collection points. Linkage of the metadata describes the exact point in the manufacturing process that a measurement is collected. In the prototype product the metadata defined over five thousand data point of collected data.

The creation of the metadata driven process drove a view that has many repeating rows with only the lowest level metadata changing along the manufacturing process. They found data types for data points that were random along the product process. Finally, there was a need to analyze at the cell level rather than the column heading. Standard data warehouse table views did not satisfy the need to view the data as a metadata string with its associated collected value. SAS Proc Transpose provided a means to transpose the data from a traditional point of view to one that supported column headings created from the cell level data. This transpose process produced unusable tables of values and associated metadata strings that contained over 5 thousand columns! Created columns headings were less than satisfactory since only 32 bytes could support the transpose and collected data values were still in a 'text' state.

The manufacturing model contained metadata that describes each collection point data type in terms of readings, units, date, date time, and so on. Using the metadata data types, actual values could be

transformed but column headings still present a problem.

Multiple levels using SAS Proc Transpose, Arrays and Macros provided the solution. By creating tables that were half metadata string and half transposed, various levels of the manufacturing process could be created and reported on. In the end, a single row that represented the product manufacturing process was created with the correct data type for each collected data point value.

SAS Warehouse Administrator provides the process tool to document, load and transform the metadata and collected data values. Using both generated code and custom SAS procedures; SAS Warehouse Administrator provides the platform that socializes the manufacturing process into a "single vision of the data".

Larry Bramblett, Data Warehouse Architect
Data Warehouse Solutions, LLC
San Ramon, California