

## Paper 201-25

**Bringing the Data Warehouse of Standardized Medical Costs to the Web**

Priscilla Van Grevenhof, Mayo Clinic, Rochester, MN  
 Kirsten Hall Long, Mayo Clinic, Rochester, MN  
 Judith L. Wagner, Congressional Budget Office, Washington, DC  
 Ahmed S. Rahman, Mayo Clinic, Rochester, MN  
 James M. Naessens, Mayo Clinic, Rochester, MN  
 Roger W. Evans, Rochester, MN

**ABSTRACT**

A historical Cost/Utilization Data Warehouse developed in SAS® provides Health Services Evaluation researchers with valuable data for analysis. All medical and hospital services provided at the Mayo Clinic and all services provided in Olmsted County from 1987 to 1998 currently populate the warehouse. The volume of data that is now available created some interesting challenges. Distribution of the metadata was needed as well as the mapping of data elements to terms the target audience understood. The need to provide the client with an interface to access the reference tables containing coding system descriptions, coding translations, and costing procedures through the same web interface was recognized. Creating a streamline process for data retrieval permitted the inclusion of a system to log the retrievals for documentation and security. Standardized analysis, reports and relevant metadata are consistently provided in conjunction with the detailed data retrieval. Lastly we determined that it was desirable to provide a feedback mechanism to share experiences using the data. Pitfalls and successes would be shared along with publications of the usefulness of the data. The resolution of the previously mentioned issues was found in providing web interfaces on our robust in-house intranet. Using web development tools including SAS/AppDev™ Studio and SAS/IntrNet™ products made this interface possible. The development of these tools and how we accomplished the task will be discussed in this paper.

**BACKGROUND****Data Warehouse**

Developed for the researchers at the Mayo Medical Center, the Cost/Utilization Data Warehouse was created in SAS from health care billing data. The data warehouse development was in SAS for the following reasons: (1) in-house programmer expertise, (2) statistical analysis expertise, (3) knowledge of SAS tools for interfacing and accessing source systems and (4) SAS availability on four platforms at the Mayo Clinic.

On average, one year of data contains 13 million records. The warehouse contains 10 years of data from five institutions. Synchronizing the data to have common values across time was one of the most challenging aspects of the project. Almost every data element needed some modification. From dates to patient identification numbers, source systems were almost never standard over time. An example of how the data is synchronized, is found in the effort it took to code all Current Procedural Terminology (CPT) codes to a base year CPT4 code (1995). Codes may be modified, added or deleted each year. No cross-reference could be found, in-house or a vendor product, that would map codes from year to year. It was important to do this since the core of the system was a standardized cost and the cost is driven by the CPT4 code in the outpatient setting. Approximately 150 codes were researched and re-coded by a team that included physicians and the Section of Health Services Evaluation (HSE) personnel.

Data over the 10-year period of time may come in many versions. The data coming from systems that are current, archives of past versions of software and software that has been replaced all need to be obtained and analyzed for completeness. For Information about the early years of data, we could only hope to find people who had worked with the data and remembered something about the information that was needed. Eight different processes were defined to create the standardized record and a standardized cost. Over 40 tables were created to assist in the effort of standardization of data over time, systems and institutions

**Costing Methods**

Applying a standardized, inflation-adjusted estimate of the cost of each service and procedure, reflecting the national average of providing the service in constant 1995 dollars, using Medicare hospital cost reports and fee schedules was the basis for the costing methodology. The warehouse is in the process of being updated to standardize costs to 1999 dollars. The same methodology is being used but the costing tables used to drive the costs are the Medicare physician fee schedules. A more extensive description of this data warehousing effort is described in a paper given at SUGI 24.

**THE WEB AND THE WAREHOUSE**

The potential users of the data are researchers in many locations of the Mayo campus. The Mayo Intranet was used to help potential new users understand the warehouse and to keep current users informed about the latest warehouse information. The Web site devoted to the warehouse contains the following information: (1) Technical Report on the methodology used, (2) Metadata, (3) Request for Data Mart, (4) Standard summarization of the data, and (5) Standard WebEIS interface for interactive reports.

**METHODS****Technical Report**

The purpose of the technical report is to describe the costing methodology and the data preparation methodology. This was created using Adobe PageMill.® The researcher can refer to this document to assist in understanding the warehouse, and as a reference in publications. Sections of text are hyper-linked to the metadata for further explanation.

Metadata as presented on the intranet contains three types of information; technical, business, and analytical.

**Technical Metadata**

A description of each data source is contained in the technical metadata. A data flow diagram shows how costs are prepared for each of the eight sources of data. One such flow chart is shown in Figure 1. Hyper-links are provided from various hot spots of the flow chart to further describe the process represented by the

symbol. The methods and rules for data transformation are described as well.

HSE Costing Data Warehouse  
 OCH Costing Method

January 5, 1999

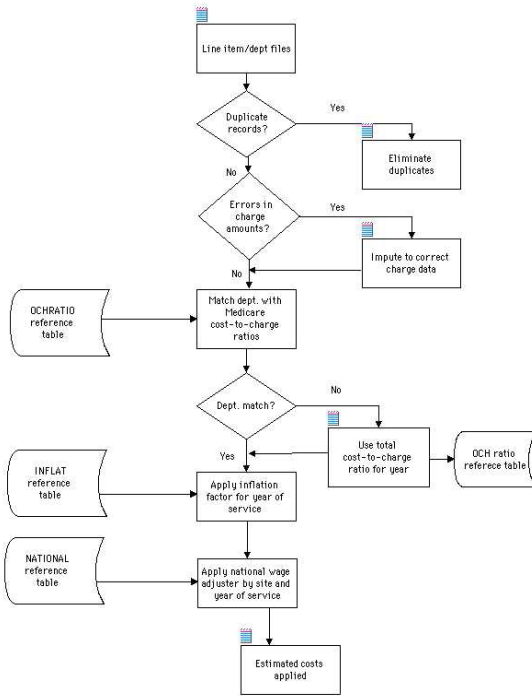


Figure 1: Flow Chart for Costing

**Business Metadata**

A table of data definitions furnishes the following information: common name, variable name, description, format, type, length and comments and limitations. Year by year summary tables inform the researcher how well the costs compare. Tables can be found on estimated costs, charges, cost to charge ratios, imputed costs and the record counts in the warehouse. Access is provided through the SAS/IntrNet to the reference tables used in the conversion of the data and the 40 descriptive tables. All of these components are an important part of business metadata.

**Analytical Metadata**

The interpolation methods and the cost algorithms are the building blocks of the analytical metadata. In this table driven system, tables are used to calculate the cost and also used in the imputation methods. Access to the 15 tables used is also a SAS/IntrNet application.

**Data Request**

The retrieval of data from the data warehouse creates a data mart. The data mart is not the usual departmental system. The system is developed for a funded research project that has Institutional Review Board approval. The research project collects additional information pertinent to that project. The additional data collected is from institutional systems, departmental systems, the medical record, surveys or questionnaires. Prior to the warehouse most of the information was collected through review of documents and data entered. This time consuming process was much more lengthy and costly. More importantly the normalized costing added an enhancement not found elsewhere. The request for creation of a data mart is also found on the web. SAS/IntrNet is used to process the request (figure 2).

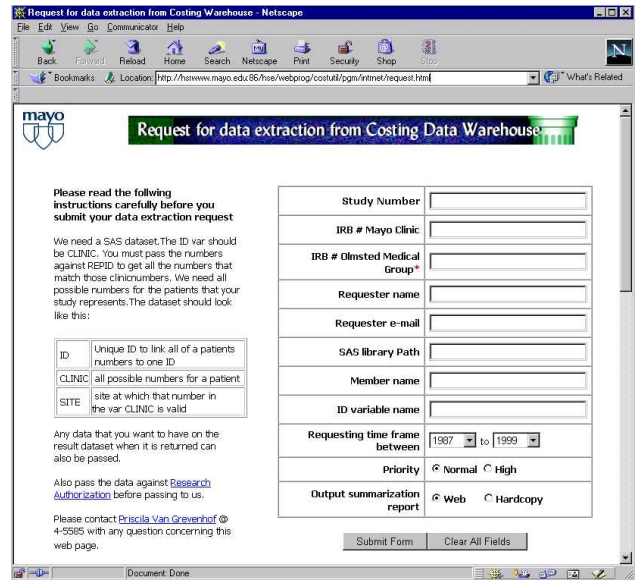


Figure 2: SAS/IntrNet Request Form

The request information is stored in a SAS dataset to log the information about the request. Upon submitting the web form, a SAS macro is invoked to produce a copy of the request that is emailed to the requester and HSE personnel. Certain security measures must be met to obtain this retrieval. Unfortunately one part of that security is still a paper system, therefore, the request must be handled in the similar manner. When it is electronically available, the processing of information can take place automatically. The data mart creation is run off-hours due to the amount of data that is processed and produced. The data mart created is a defined research project that is a large study of a group of patients usually developed for a specific department.

**Standardized Reports**

The standard reports summarize the data by logical groups across years, patient, and cost of illness. A requester has the option of hardcopy or creation of a web document. The static reports were developed using the SAS/Output Delivery System (ODS).

**WebEis Reports**

AppDev Studio is used for our latest development. WebEIS™ software permits the production of standard dynamic reports and publishes them on the Web. These are interactive reports and tables. (Figure 3)

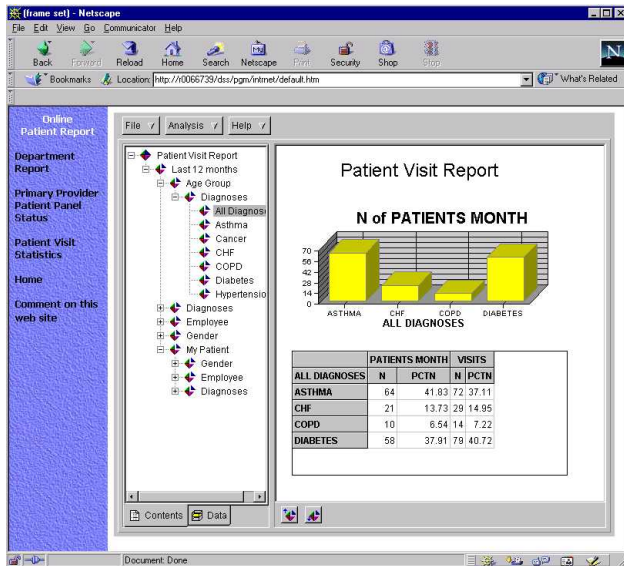


Figure 3: WebEIS

These reports are presently developed for reporting by CPT4 categories, diseases, and physicians. Another feature is the ability to drill down to the detailed data and export the data to Excel, a feature that is especially appealing to clients.

## CONCLUSION

With more requests for data marts being initiated, the demand for more standard tools will definitely increase. Utilizing SAS web tools found in AppDev Studio and the Output Delivery System applications can be developed quickly. Prototyping is made easy. Investigation of purchasing the SAS/MDDB Server is taking place. This product may add to the researcher's ability to analyze and capability to improve general reporting. Since the data mart owners are conducting research all the needs will not be served but general questions about the data will be answered with little effort.

## REFERENCES

American Medical Association: [Current Procedural Terminology \(CPT\), 1999.](#)

Ben Cochran, The Bedford Group, Raleigh, NC. *Using the SAS/MDDB and User Written Methods to Enhance SAS/EIS Objects.* SAS Users Group International Conference Proceedings. SUGI 24. p285-290, 1999.

Inmon, Bill, Imhoff, Claudia, and Gosselin, Larry. *Managing the Data Warehousing Effort: Advanced Topics.* Barnett Data Systems Course Notes.

SAS Institute Inc. (1996), The SAS Data Warehouse. A SAS Institute White Paper.

Tidename, Sue and Chu, Robert. *Building Efficient Data Warehouses: Understanding the issues of Data Summarization and Partitioning.* Proceedings of the 21<sup>st</sup> Annual SAS Users Group International Conference.

Van Grevenhof, Priscilla, Wagner, Judith L., Helgeson, Eva R., Naessens, James M., Killian, Jill M., Evans, Roger W. Using SAS® Software to Build a Data Warehouse of Standardized Medical Costs and Patient Utilization for Health Services Research. SAS Users Group International Conference Proceedings, SUGI 24. p1465-66, 1999.

## TRADEMARK

SAS is a registered trademark or trademark of SAS Institute, Inc. in the USA and other countries. ®Indicates USA registration.

Other brand and product names are registered trademarks or trademarks of their respective companies.

## CONTACT INFORMATION

Contact the Author at:

Priscilla Van Grevenhof  
 Mayo Clinic  
 200 First Street, SW  
 Rochester, MN 55905  
 Phone: (507)284-5585  
 Fax: (507)284-1731  
 e-mail: vangreve@mayo.edu