USING SAS® TO PRODUCE THE DIRECTORY OF TEXAS MANUFACTURERS

Sylvia L. Cook
Bureau of Business Research, The University of Texas at Austin

Introduction
The Bureau of Business Research at the University of Texas at Austin has published the Directory of Texas Manufacturers for the past 56 years. The 1989 edition of the Directory lists 16,329 companies that manufacture in the state. The publication of the Directory is the culmination of a year-long research effort to revise information on companies listed in the previous edition and to collect information on new companies not already listed in the Directory. The last three editions of the Directory have been produced using SAS®. The data published in the Directory are kept in SAS® datasets and SAS® is used as a programming language to maintain the data, to format the data for publication, and also to extract subsets of the data for customers. This paper describes the mechanics of the system used.

Data Organization
The data are stored in six datasets linked by common variables. The primary dataset, DTM.DATA, contains one observation and 37 variables for each company represented in the database. The variables include a unique ID number for the company, company name, plant name, street address or location, city or town in or near which the plant is located, mailing address, year plant was established, telephone number, toll-free telephone numbers, Telex and fax numbers; name, title, and telephone number of the plant executive, purchasing agent, and sales agent; and several recordkeeping variables such as the year the company was first listed in the Directory, the edition year for which an update form was last received from the company, and the date the data were last modified. The dataset also includes a code indicating the action taken when a form was last received from the company: new entry added to the database, last form received with changes, last form received without changes, unable to locate company, company no longer manufactures, out of business, or company chooses not to be listed in the Directory. Those companies whose status codes indicate that they are not to be included in the Directory are deleted from the database before the data are printed.

A second dataset, DTM.SIC, includes the ID number, a description of a product manufactured by the company, and the Standard Industrial Classification(SIC) code for the product. There is at least one and no more than five observations in DTM.SIC for each observation in DTM.DATA. Another dataset, DTM.OTHNAME, includes the company ID number, alternate or double names of the company, and a variable indicating the type of other name. Only some of the companies included in DTM.DATA are included in this dataset. A fourth dataset, DTM.HISTORY, includes the company ID number, the current name of the company, and the former name of the company if it was changed in the update process. This dataset includes only some of the companies in DTM.DATA. It is not kept after the end of the publication cycle.

Many of the companies in the database are subsidiaries of one or more other companies. The dataset DTM.HONUM includes, for each company which has a parent company, its ID number and its unique home office.
identification number. Approximately one-third of the companies in the database are represented in this dataset, some more than once, and many home offices have more than one subsidiary. The dataset DTM.HOMEOFF includes for each of the parent companies or home offices included in the database, its name and address and the home office identification number.

Updating the Database

The annual update process begins by printing forms sent to the companies in the database. Forms are sent to the companies listed in the current Directory and to the home offices. Information in the datasets is merged by company and home office ID numbers. Using the SAS® programming language, the data are written to a file and printed on letter-size paper on a laser printer.

As the forms are received and organized the update process begins. The ID numbers of companies whose update forms have been received are entered and a copy of the data for each of those companies is extracted from the database and made available to a data entry operator through PROEDIT. This copy of the data is changed to reflect the changes made by the company on the update form and is then submitted to the editor for review.

The editor reviews the data, again using PROEDIT, and after her approval, this edited copy of the data is used to modify the datasets in the database. The process of modifying the database uses the SAS® update statement to apply transactions in the modified copy of the data to the database itself to update the data.
Printing the Data

When all forms received from the companies have been processed, the data are then formatted for printing. Again using the SAS® programming language, files are created for the printing of three sections of the Directory: an alphabetic listing of companies, a listing of companies by their location, and a listing of companies by the products they manufacture. SCRIPT® files are created and then printed on mylar on an electroerosion printer. The output, which has a resolution of 600 pels per inch, is then sent to a printer for processing.

Extractions from the Database

Both the printed volumes and data from the database are sold. When the data have been finalized for publication, two datasets, RETRIEVE.DATA and RETRIEVE.SIC, are created from which data are extracted for customers. In the update process, data are extracted from the database by ID number. In the retrieval process, data are extracted from the retrieval database by SIC code, geographic area, number of employees, sales volume, or any combination of variables. These data take the form of mailing labels, printed reports, or data downloaded to diskette in fixed-field or delimited format. The SAS® programming language is used to format the data for all these media.

Advantages of Using SAS®

Our experience using SAS® to maintain the manufacturers database and to produce the Directory has been satisfactory and worth the effort of conversion. SAS® has given us flexibility in the kinds of data we collect and in the types of extractions of data we do for our customers. The flexibility in data collection has allowed us to use the database more easily in other research conducted in our office. Because we are able to produce final copy for the printer in-house, we have more control over the publication process and we are able to produce the Directory more economically. This has also shortened our publication cycle. Because we are producing the Directory of Texas Manufacturers more efficiently, we are now working on the first edition of the Directory of Texas Wholesalers.

For further information contact:
Sylvia Cook
Bureau of Business Research
University of Texas at Austin
P. O. Box 7459
Austin, TX 78713
(512) 471-1616