

Warehouse to Web with Three Powerful SAS Products

Inez Williams, Ohio Department of Health, Columbus, Ohio
Carol Martin, Ohio Department of Health, Columbus, Ohio

Abstract

Customers have urgent needs for quick and easy access to quality, consistent, and timely Ohio health data. To meet their ever-growing needs, the Ohio Department of Health (ODH) Data Warehouse was created. By integrating the SAS/Warehouse Administrator, SAS/MDDDB Server, and SAS/Intrnet, health information is available in static and dynamic formats. This paper provides an overview of the methods used to develop the Internet version of the ODH Data Warehouse using these three SAS products.

Introduction

The ODH Data Warehouse is an easy to use, point-and-click environment containing a structured collection of health data, organized into key health related topics. Figure 1 illustrates the flow of health data from the point of data extraction to the Internet. The warehouse was developed and implemented to satisfy specific information needs concerning the health status of Ohioans. The address of the warehouse is www.odh.state.oh.us/dataware.htm.

Currently, the ODH Data Warehouse consists of Ohio Vital statistics data: births, deaths (total population, infant, neonatal, and fetal) marriages and divorces. Multiple help screens, help buttons, and descriptions of the reporting variables promote a user-friendly environment. State and county level data are accessible for the years 1990 through the most current year available. The ability to drill down to city level data will be available soon. During 1999, the ODH Data Warehouse will continue to expand, with the addition of Reportable Disease and Behavioral Risk data.

To address the needs of a variety of users, the data is presented in both static ("At a Glance") and dynamic formats ("Quick Analysis"). Static reports show data for Ohio and all 88 counties, while the Quick Analysis reporting option gives users the ability to query the data. In 1999, the "Custom Reports" function will be available.

The following discussion highlights how the SAS/Warehouse Administrator, SAS/MDDDB Server, and SAS/Intrnet products contributed to the latest release of the ODH Data Warehouse: Ohio Resident Death Data.

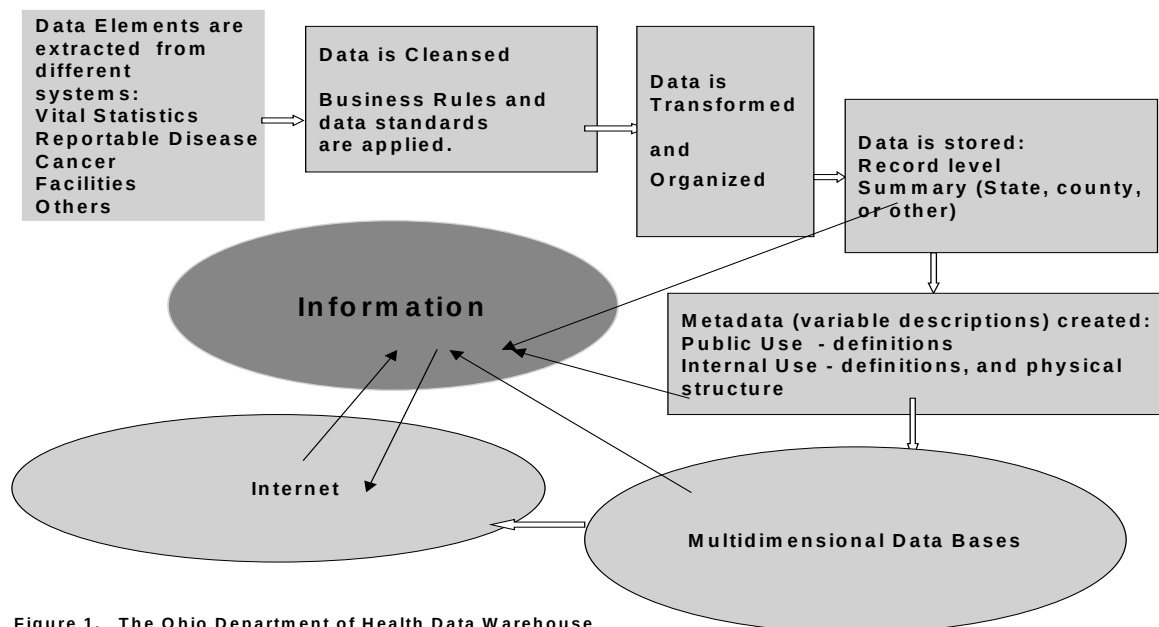


Figure 1. The Ohio Department of Health Data Warehouse
Changes Data Into Meaningful Information

Using the SAS Warehouse Administrator

As shown in Figure 1, multiple operations impact the data (extracted from various systems) as it is transformed into information. During this process, the SAS Warehouse Administrator functions as a data management tool. The software performs many activities to extract, cleanse, edit, transform and store data in the repository. Prior to storage, this information is also organized, summarized, indexed, and labeled, in preparation for use by other tools. Additionally, metadata is created which describes in detail the contents of the warehouse.

This presentation segment focuses primarily on the User Exit (user-written code) because of its role in the development of the ODH Data Warehouse. Given the structure of the vital statistics source files, the User Exit process helped to simplify data management, especially when creating the Operational Data Definitions (ODDs) for the vital statistics death data. Each ODD includes record level data for one calendar year. Among the eight vital statistics death source files used for the third iteration of the Data Warehouse, there were some differences in variable lengths, definitions, values, and some variables were not collected in given years. These issues were resolved within the User Exits.

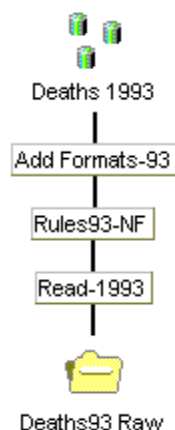


Figure 2. Process Flow to Create the 1993 Death Data ODD

Figure 2 illustrates steps to create an ODD (Deaths 1993) with three User Exits. The first Exit (Read-1993) standardizes the variable lengths and initializes uncollected variables. Rules93-NF transforms and edits the data to apply the business rules. These business rules are based on input from internal ODH staff, research, and national standards such as the National Center for Health Statistics (NCHS). Formats and labels are applied in the Add Formats-93 exit.

Within the Warehouse environment, SAS catalogs are used as a means to organize and maintain various

entries such as those containing source code (including User Exit code) and output data. Remote Library Services allows easy access to the server catalogs needed by our warehouse. One run statement generates the ODD shown in Figure 2. This process is repeated for each year of data.

Figure 3 illustrates linking of the yearly ODDs using the SAS Warehouse Administrator mapping process to form a Detail table (Resident Deaths), which is then summarized and indexed. With this step completed, the files were ready for use by the SAS/MDDB Server software.

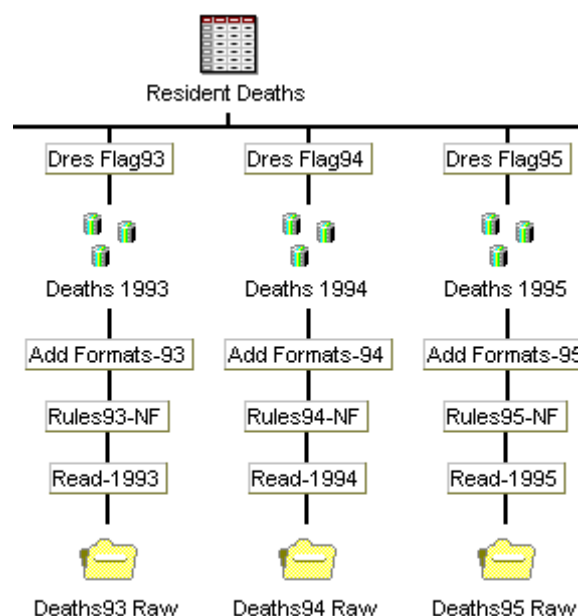


Figure 3. Linking ODDs to Create a Detail Table

USING SAS MDDB/Server

The next step in the information dissemination process was to build the multidimensional databases using SAS/MDDB Server software. The MDDB differs from the relational row and column structure by storing presummarized and cross-tabulated data in individual cells in a matrix-like structure. The SAS/MDDB Server product was selected to facilitate rapid loading and retrieval of data. To create an MDDB, PROC MDDB is used with hierarchy statements that create subtables for quick access to cells:

```

PROC MDDB data=whse.dths_sum
out=whse.dpop label='All Resident Deaths';
class cnty year racegrp sex;
var y /sum;
hierarchy cnty year racegrp sex;
hierarchy cnty year sex;
  
```

hierarchy cnty year racegrp;

The content and number of the databases is the product of two months of analysis and review. Input from various external customers, such as Minority Affairs and the local health districts we serve, was solicited. Internal customers, who were bombarded with questions from newspapers, the legislature, and the public, were also asked for feedback. Finally, production reports were reviewed to identify the most common types of questions asked.

We identified three main goals we wished to accomplish through the information dissemination process. Our first goal was to answer many of the questions our local health departments have, so that health in their communities could be assessed. These answers help the local health departments create programs to improve the health of their citizens. The second goal was to answer commonly asked questions so our researchers could be free to do more research and less ad-hoc reporting. Our third goal was to facilitate rapid and consistent information dissemination via the web, within one or two days of the official release of the data from our Vital Statistics Department. Quick dissemination of accurate data helps our ODH internal staff as well.

To meet these goals, 22 databases were created. Eight of these databases answer detailed questions about age, gender, race, and ethnicity. The remaining 14 are specific to diseases causing death in Ohio. The 22 databases, built with Ohio resident data from 1990-1997, are completely built in 3 minutes on our NT server, which houses the MDDb server software. The summary set, from which the databases are built, contains 1,391,000 observations and is commonly referenced as the base table.

Using SAS/INTRNET

After the databases were built, the next step was to create HTML pages to link the customer's requests for data with the appropriate application. We used SAS Web Publishing tools and SAS/Intrnet to create many of the HTML pages. Our SAS/Intrnet applications are written in SCL and utilize EIS methods and functionality. These HTML pages allow selection of a database (subject) and one or more Ohio counties on the left side of the page. When the "Submit" button is clicked, the first of two applications is invoked. The first application retrieves the database and the selected counties, and populates the right side of the page with available years and reporting variables from the database as shown in Figure 4. The HTML code used to invoke the application is:

```
<Input type=hidden name=_service value="default">
<Input type=hidden name=_program
      value="system.windows.nmddbd.scl">
<Input type=hidden name=pop value=sasdl.dpop>
<Input type=hidden name=_debug value=0>
```

<Input type=hidden name=county value=Ohio>

Figure 4.

The customer then chooses one or more years and a reporting variable. When the "Submit" button is clicked again, the second application intercepts the response and creates the report as shown in Figure 5. The customer can choose to print the report or save the report as an HTML file.

County	Year	Selected Causes of Death												Total
		Heart Disease		Cancer		Stroke		Pulmonary (COPD)		Diabetes		Other		
		n	%	n	%	n	%	n	%	n	%	n	%	
Ohio	1997	34,201	32.5	24,966	23.7	7,780	7.4	5,346	5.1	3,639	3.5	29,208	27.8	105,140
	1996	34,398	32.8	25,483	24.3	7,484	7.1	5,226	5.0	3,766	3.6	28,494	27.2	104,856
	1995	35,668	33.8	25,327	24.0	7,272	6.9	5,043	4.8	3,457	3.3	28,796	27.3	105,563
	1994	35,035	34.0	25,113	24.4	6,981	6.8	4,924	4.8	3,275	3.2	27,626	26.8	102,954

Figure 5.

The application and databases are now complete and we are ready for customer and researcher reviews, the final step in the development process. To enable this process to work efficiently, we have developed a

test plan that includes allowing our customers to review the application on our test server. For security purposes, pre-selected reviewers are given a URL address that is available for 10 days. At the conclusion of the 10-day period, the address again becomes secure. Customers e-mail their comments regarding content, presentation and navigation to the database administrator for review. Easily incorporated changes are made while more time consuming enhancements are placed to the "to do" list for inclusion in the next warehouse iteration.

Once customer review is completed, the entire application is copied from the test server to the production server. This migration method allows the SAS/MDDDB software to be licensed on one server where the databases can be built and rebuilt as customers are testing the application and databases. The SAS/Intrnet software resides on a different server, where external web customers cannot drill into detailed, confidential data that was originally used to build the databases.

We have exerted much effort to insure that we developed a customer friendly warehouse. Extensive help, glossaries, methods of calculation, and definitions of variables and groupings have been included in the Ohio Department of Health Data Warehouse web site. All new reports or types of data are added with clear and concise documentation that is easy to find and easy to navigate. Figure 6 illustrates this.

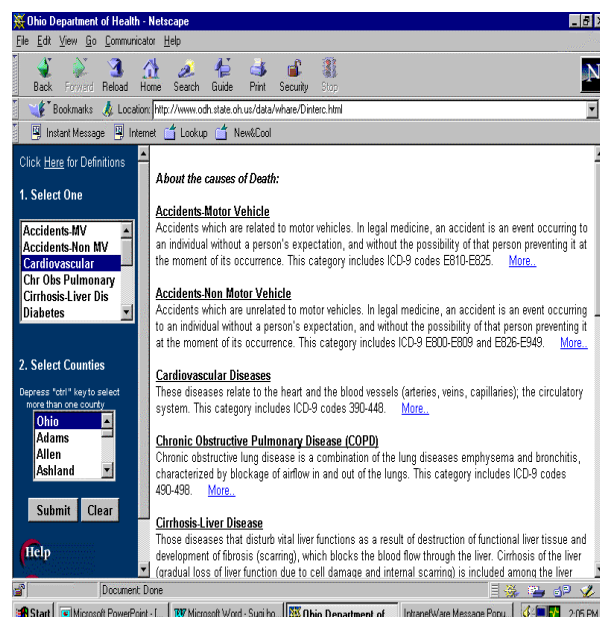


Figure 6.

Conclusion

We have been extremely pleased with the tools we used to create our warehouse. Many states and private organizations have contacted our data warehouse group with questions about the products we are using. Comments about the ease of use and the response times of our applications have been extremely favorable. In a recent survey taken at a Health Commissioner's conference, our data warehouse was determined to be the most frequently used and best liked feature of the Ohio Department of Health's web site.

During the 14 months the warehouse has been web-enabled, we have experienced no down time due to the software products we are using. We have experienced only two timeouts. The great response time and quick access to health data can be attributed to the speed and power of the multidimensional databases and the SAS/Intrnet products. The quality and flexibility of data management can be attributed to the SAS/Warehouse Administrator.

Acknowledgments

SAS, SAS/Warehouse Administrator, SAS/MDDB, and SAS/Intrnet are registered trademarks of the SAS Institute Inc. in the USA and other countries.

Contact Information:

The authors can be contacted as follows:

Carol Martin
(614) 728-4511
Camartin@gw.odh.state.oh.us

Inez Williams
(614) 728-4739
Iwilliams@gw.odh.state.oh.us

Ohio Department of Health
Department of Policy Planning
and Health Data Care Center
7th Floor
246 North High St.
Columbus, Ohio 43215

