

From Raw Data to Web-Accessible Information: SAS® Software Solutions for a State Health Data Agency

Chip Quade and Barbara A. Rudolph, Ph.D.

Wisconsin Office of Health Care Information, Madison, Wisconsin

Abstract

This paper describes a SAS-based information system that supports the data management, analysis, and information dissemination needs of a state health data agency, with special emphasis on the use of SAS software in an interactive Web-based query system. Wisconsin's Office of Health Care Information (OHCI) collects data on inpatient hospital discharges, ambulatory surgeries, health care facilities, and health care providers. OHCI cleans the raw data and produces standard reports, standard data products, and custom reports and data sets. In 1996, OHCI launched an interactive data query system on the Web. Internet customers can generate questions and quickly get answers from statewide patient-level databases containing up to 1.3 million records.

OHCI uses base SAS software, SAS/FSP®, SAS/SHARE®, SAS/SHARE*NET™, SAS/AF®, SAS/CONNECT®, and SAS/STAT® in a networked Windows NT and UNIX environment. The paper presents an overview of the architecture of OHCI's information system and a detailed discussion of the three-tiered Web query system, which will be demonstrated during the conference presentation. The three-tiered approach makes it possible for OHCI to use a single set of core programs with:

- different types of underlying data (inpatient files and ambulatory surgery files), and
- multiple user interfaces for audiences with different perspectives on health information and different levels of sophistication with respect to using the information.

Background Information About OHCI

OHCI is a state agency whose mission is not nearly as broad as its name would imply. OHCI is the steward of Wisconsin's statewide inpatient discharge data base and ambulatory surgery data base. Facilities submit about 1.1 million patient-level records to OHCI each year. OHCI also

surveys the health care provider work force and collects information from Wisconsin hospitals about their facilities, services, staffing, and operation. From this information, OHCI produces standard data products and reports. It also produces custom data sets and reports in response to end-user requests (and charges customers for the work), and lets customers query the office's data bases via the Internet (for free).

OHCI's customer base is fairly diverse and includes:

- Private-sector purchasers of health care
- Insurance companies
- Managed care organizations
- Government agencies
- Health care providers
- Health data resellers
- Policy consultants
- Teachers and researchers at colleges and universities
- Foundations, institutes, and alliances
- Law firms
- Journalists
- Consumers

Most of the above are interested in either the health of Wisconsin's population or the functioning of the health care marketplace. Information products aimed primarily at OHCI's professional customers include:

- Standard data sets
 - ♦ quarterly patient-level data sets
 - ♦ annual data sets from facility surveys
 - ♦ data sets from periodic surveys of health care providers (physicians, nurses, etc.)
- Standard reports
 - ♦ quarterly and annual *Health Care Data Reports*, with information about medical conditions and treatments for patients at Wisconsin inpatient hospitals and ambulatory surgery facilities
 - ♦ the *Guide to Wisconsin Hospitals*, which contains information about hospitals' revenues, expenses, types of services

- provided, and patient volume
- ♦ a *Small Area Analysis* report, which analyzes, by patient zip code, hospitalizations that might be prevented if patients received better ambulatory care
- ♦ a report on *Morbidity and Mortality in Wisconsin Hospitals*, which presents severity-adjusted death rates from certain conditions and information about the incidence of various other medical conditions
- Custom data sets and reports

However, consumers are an increasingly important part of OHCI's customer base, and the demands for health care information are evolving. OHCI's Web-accessible information systems grew out of these evolving demands. To meet its customers' varied and ever-expanding needs, OHCI requires an information system that efficiently manages a relatively large volume of data and makes possible many different types of analyses.

OHCI Information Systems: Customer Needs and SAS Solutions

OHCI's computing environment consists of:

- a Sun Ultra Enterprise 4000 running Solaris 2.5.2
- approximately 25 Windows NT machines
- an NT server and a Netware server
- the SAS System for Solaris, Release 6.12
- the SAS System for Windows, Release 6.12
- NCSA httpd Version 1.5.2a (Web server software) running on the Sun

The OHCI staff's main SAS-related information processing tasks are as follows:

- Review and clean incoming patient, provider, and facility data
- Create quarterly archives of patient-level data
- Create annual archives of facility-level data
- Create periodic archives of health care provider survey data
- Produce standard data products
- Create custom data sets and reports (typically, simple tabulations) in response to end-user requests
- Implement statistical analyses for both standard and custom reports

- Query SAS data from Microsoft products (Access and Excel)

In addition, internal and external customers generate queries against SAS data sets via OHCI's interactive Web system.

To support these information processing tasks, OHCI uses:

- Base SAS software and SAS/FSP for general data management
- SAS/SHARE for concurrent access to data sets
- SAS/AF for custom screens to launch in-house applications
- SAS/CONNECT for remote execution on the Sun from a PC
- SAS/STAT for regression analyses
- SAS/SHARE*NET and the SAS ODBC Driver to support queries from Microsoft products

Overview of OHCI's Web Information System

OHCI launched its Web-based information delivery system to fill a very specific gap in its spectrum of customer services. Many customers come to OHCI with specific questions that the standard reports do not address. And many of these customers lack the resources and skills to purchase OHCI's data sets and implement their own analyses. In the past, such customers had no choice but to submit a request for a custom report. But it takes staff resources to generate the custom reports, which can themselves be somewhat expensive for the customer, and OHCI's backlog of requests is sometimes quite large.

To provide an easy and less resource-intensive way for customers to get a timely response to many types of specific questions, OHCI turned to Web technology. But here OHCI faced a new set of questions, including: Just who uses the World Wide Web? What can these customers be expected to know? And what does OHCI have to do to really meet the needs of these customers?

For some organizations, a Web-based information system is nothing more than a Web-accessible data set. But providing access to *data* is not enough. A customer must be able to get meaningful *information*. And OHCI's customer

base is far from homogeneous. Some customers are familiar with health data terminology, but many are not. The approach that works for the health data specialist probably doesn't work for the consumer, and vice versa.

OHCI's design considerations included the following:

- The user interface must be organized as a function of how a user thinks, not as a function of how the data are structured. (Corollary: if different classes of users think differently, we may need more than one user interface.)
- The system must provide plenty of on-line help to assist users in finding what they want, constructing questions, and interpreting the results.
- The system must provide a response relatively quickly.
- Health care information is inherently complex, and appropriate use of the information often requires careful reading and thinking on the part of the user. And the more complex or subtle the information, the greater the risk of it being misinterpreted.
- An organization has little control over what a Web customer actually reads while visiting the site. In particular, eager users are likely to ignore caveats or cautions in their rush to click the next clickable hyperlink or button.
- Confidentiality requirements limit the amount of information OHCI can make available via any means, including the Web.

Web Query System: Implementation

The technical solution consists of a three-tiered approach, based on a system developed by the Utah Office of Health Data Analysis. The inpatient discharge query system consists of:

1. A SAS data set containing two years of inpatient discharge records.
2. Web pages that implement the user interface with standard HTML.
3. A "Middleware" component (CGI program), compiled from C source code, that:
 - interprets the user's query, as passed to it from the Web browser via the Web server software
 - generates a SAS program

- executes the SAS program
- reformats the output and returns the result to the browser

Each SAS program consists of:

1. a DATA step to select a subset of records and variables
2. a PROC SUMMARY, or a PROC SORT plus a PROC UNIVARIATE
3. a null DATA step to write the result to a text file

Configuration files govern the execution of the Middleware and the appearance of the output. Additional Web pages, support programs, and reference files implement the various parts of the on-line help system.

The SAS data set for the Web query system is a subset of the information in OHCI's archival data sets. The limited data set contains one record for each inpatient hospitalization in a two-year period - 1.3 million records in all - but the variables in each record are a subset of the variables in the archival data sets. OHCI keeps each record relatively small for three reasons. First, patient confidentiality is of supreme importance. Users must be prevented from generating queries that would let them identify individual patients, so OHCI excludes all variables that might function as direct or indirect patient identifiers. Second, a smaller data set means a quicker response time: typically 15 to 45 seconds, depending on the complexity of the query. Third, OHCI's goals are limited. The Web query system is not intended to satisfy every single customer's every need. In particular, the system does not support certain types of complex requests that would produce easily-misinterpreted results. (OHCI provides such information only through direct contact with a staff member.)

A customer comes to the Web query system with a specific question in mind, so the user interface is organized in a way that helps the customer translate the question into a query. And because the customers are not expected to be computer experts or programmers, the interface addresses the customer with language like "Do you want to zoom in on a subset of patients?" and "What do

you want to calculate?" rather than "Select variables for defining a subset of cases."

Advantages of the Three-Tiered Approach

The three-tiered approach offers some distinct advantages. First, OHCI can use two sets of Web pages to present different user interfaces to different audiences. With the health data specialist's interface, a customer frames a query in terms of diagnosis codes, procedure codes, diagnosis-related groups (DRGs), and major diagnostic categories (MDCs). With the general public's interface, a customer frames a query in terms of a "reason for hospitalization" (for example, coronary bypass surgery) independent of how that reason is represented in the underlying data. Figures one through four, at the end of this paper, show the steps a customer takes to generate a query about coronary bypass surgeries. Figure five shows the result of the query.

Second, OHCI can create query systems for additional types of data without changing the analytical Middleware. After launching its inpatient discharge query system, OHCI created a query system for ambulatory surgery data by:

- generating a Web-accessible data base of ambulatory surgery records
- building an additional set of Web pages to implement an appropriate user interface (or two, for technical and non-technical audiences)
- creating a new set of configuration files, and modifying the relevant entries to change the workings of the Middleware

In short, what looks to the customer like four different information systems is really two types of data, two user interfaces for each, and a shared program in the middle. The IS specialist's job is to make sure the HTML files, data sets, and configuration files are in sync. Figure six, at the end of this paper, shows the components of the Web-based inpatient discharge data system.

Why not SAS/IntrNet®?

The OHCI Web query system does not use the SAS/IntrNet tools. This is because OHCI obtained its license for the query system Middleware in the

summer of 1996, and implemented a working system in November 1996. At that point, the SAS/IntrNet tools were not ready for prime time. OHCI keeps an eye on the evolving SAS/IntrNet tools and may use them to build additional Web-based query systems in the future. However, OHCI's current technical solution is relatively flexible and easy to adapt to other types of data, and OHCI remains mindful of the importance of developing a complete, effective information delivery system rather than just making current data available via the Web.

Conclusion

Customer feedback about the Web query system has been very positive. The only complaint is that OHCI doesn't make enough of its information available via the Web.

OHCI's plans include expanding its Web-related information systems in three ways. First, OHCI will bring more types of data to the Web for both external and internal customers. Second, OHCI will add new query options to existing data systems. A few additional variables in the current information systems will cost OHCI relatively little and may make a big difference to some of its customers. Third, OHCI may integrate Web technology into its data submission and collection process.

To date, OHCI has saved considerable staff resources by letting customers generate questions via the Web, and more customers have been able to get more of what they want in a more timely manner. Future activities will be determined by business needs related to improving customer service and streamlining the data production process.

Acknowledgments

The following staff members of the Office of Health Care Information contributed to the preparation of this paper:

Brian Hutchinson	Judith Nugent
Lawrence Nahlik	Robert Purvis

SAS, SAS/AF, SAS/CONNECT, SAS/IntrNet, SAS/SHARE, and SAS/STAT are registered

trademarks or trademarks 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.

Web Address

<http://badger.state.wi.us/agencies/oci/ohci>
Note: Due to departmental reorganization, the Web site may have moved by the time of this paper's publication. Please contact OHCI if you have difficulty accessing the site.

Author

Chip Quade
Information Development Specialist
Wisconsin Office of Health Care Information
1 West Wilson Street, Room 372
P.O. Box 7984
Madison, WI 53707-7984
Phone: (608) 261-8568
FAX: (608) 264-9881
Email: quadecca@dhfs.state.wi.us

Figures

Figure 1. Step one of generating an inpatient query with the nontechnical user's interface

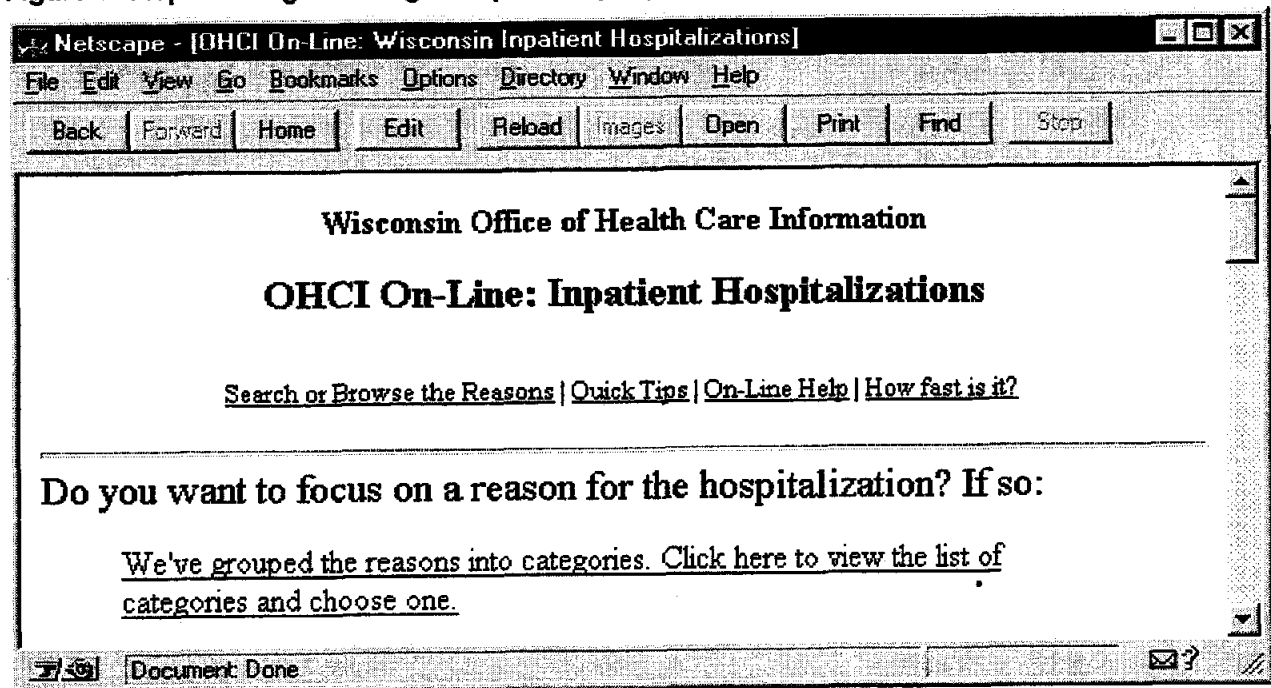


Figure 2. Step two of generating an inpatient query with the nontechnical user's interface

The screenshot shows a Netscape browser window with the title "Netscape - [OHCI On-Line: Wisconsin Inpatient Hospitalizations]". The address bar is empty. The menu bar includes File, Edit, View, Go, Bookmarks, Options, Directory, Window, and Help. The toolbar contains buttons for Back, Forward, Home, Edit, Reload, Images, Open, Print, Find, and Stop. The main content area displays the title "Reasons for Hospitalization: Heart / Cardiovascular" and a instruction: "You can choose more than one, so long as you stay within the same color or shape:". Below this is a list of 15 items, each preceded by a triangle icon and a checkbox. The third item, "Coronary Bypass with Catheterization", has its checkbox checked. The status bar at the bottom shows "Document Done" and a mail icon.

Reasons for Hospitalization: Heart / Cardiovascular

You can choose more than one, so long as you stay within the same color or shape:

- ▲ ☐ Cardiac Valve Procedure with Catheterization
- ▲ ☐ Cardiac Valve Procedure without Catheterization
- ▲ ☒ Coronary Bypass with Catheterization
- ▲ ☐ Coronary Bypass without Catheterization
- ☐ Angioplasty or Valvuplasty by Use of a Catheter
- ▲ ☐ Permanent Pacemaker with Heart Attack
- ▲ ☐ Permanent Pacemaker without Heart Attack
- ▲ ☐ Repair of Pacemaker Device
- ▲ ☐ Replacement of Pacemaker Device
- ▲ ☐ Heart Failure and Shock
- ▲ ☐ Atherosclerosis (Coronary Artery Narrowing due to Plaque Deposits)
- ▲ ☐ Hypertension (High Blood Pressure)
- ▲ ☐ Cardiac Arrhythmia (Abnormal Heartbeat)
- ▲ ☐ Angina Pectoris
- ▲ ☐ Chest Pain
- ☐ Cardiac Catheterization

Document Done

Figure 3. Step three of generating an inpatient query with the nontechnical user's interface

Netscape - [OHCI On-Line: Wisconsin Inpatient Hospitalizations]

File Edit View Go Bookmarks Options Directory Window Help

Back Forward Home Edit Reload Images Open Print Find Stop

Do you want to focus on a subset of patients? If so, specify:

Patient Characteristics:

Sex: **ALL** Male Female
 Age Group: **ALL** 0 - 28 days 29 - 365 days 1 - 4 years
 County of Residence: **ALL** Adams Ashland Barron

Year of Discharge: ☒ 1996 ☒ 1997

Expected Primary Payer: **ALL** Medicare Medical Assistance Other Government

Patient's Discharge Status: **ALL** To home or self care (routine discharge) To another short-term general hospital To skilled nursing facility (SNF)

Document Done

Figure 4. Step four of generating an inpatient query with the nontechnical user's interface

Netscape - [OHCI On-Line: Wisconsin Inpatient Hospitalizations]

File Edit View Go Bookmarks Options Directory Window Help

Back Forward Home Edit Reload Images Open Print Find Stop

Choose the question you want to answer:

- ☐ How many hospitalizations were there?
- ☐ What was the average charge?
- ☒ How long were the patients hospitalized, on average?
- ☐ What was the median charge?
- ☐ What was the median length of stay?
- ☐ What were the combined charges for all the patients?
- ☐ How many total days were the patients hospitalized?

Your answer comes in a table; how do you want to organize it?

For row headers, categorize by: For column headers, categorize by:

☐ Sex	☐ Nothing
☒ Age Group	☒ Sex
☐ County of Residence	☐ Age Group
☐ Year of Discharge	☐ County of Residence
☐ Expected Primary Payer	☐ Year of Discharge
☐ Patient's Discharge Status	☐ Expected Primary Payer
	☐ Patient's Discharge Status

Break into sub-tables after every rows and columns.

Document Done

Figure 5. The query result

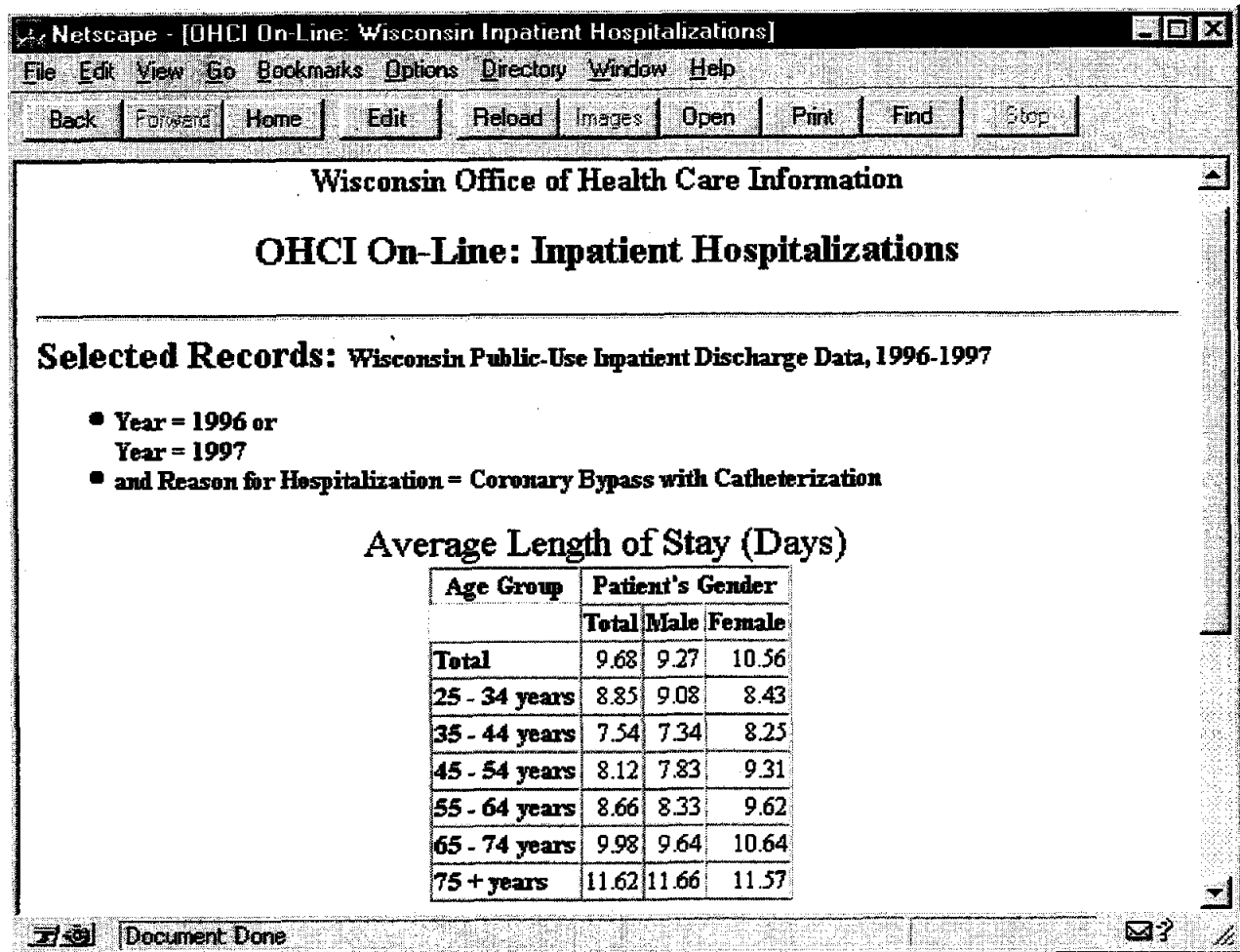


Figure 6. Schematic of OHCI's Web-based inpatient discharge query system

