

Paper 159- 27

Easy Access to Public Health Data Colorado Health Information Dataset (CoHID)

Huiyun Xiang, Colorado Dept. of Public Health & Environment, Denver, CO

ABSTRACT

The Colorado Health Information Dataset (CoHID) was developed by SAS® company, local and state public health agencies in Colorado to provide Colorado communities with easy access to local-level health data. This queryable system allows users to access community-specific information on a variety of health outcomes, conditions and health risk behaviors. Using SAS IntraNet® and other SAS web tools, CoHID allows remote users to query a variety of health-related databases such as birth and death records, population estimates, and Behavioral Risk Factor Surveillance System to help determine the health status of a neighborhood, community, county or region in Colorado. Other datasets will be added in the near future.

Like other new adventures, the development of CoHID met many challenges. This presentation will show the basic infrastructure of CoHID and technology used, and share the experiences about how to accomplish these challenging tasks with limited resources and not-so-ready technical personnel available at local and state public health agencies. The presentation is aimed at people who are interested or involved in similar projects.

INTRODUCTION

The need to have ready access to population health data at the municipal and community levels long has been recognized as crucial to fulfill the broad responsibilities for monitoring the health status of populations and for planning interventions.¹⁻⁴ But the actual availability of such data falls far short of the ideal. In the United States, there are generally well-established mechanisms for collection, analysis, reporting, and transfer of useful data at the state and national levels.³ Although more data are available now at local community levels, data needs at the local communities are usually handled as special data requests. Such needs are addressed individually by a health department data analyst who does statistical analysis or aggregates data after gathering information from the requestor. During the process, unfamiliarity with the data available and unfamiliarity with statistical terminology and data issues are often obstacles for data users to get updated and accurate data.

Several state health departments have developed interactive database query system for use by a wide range of users including general public, public health practitioners, and research communities.⁵ Ongoing evaluations and new initiatives by governmental agencies encourage further development and use of such system by more state health departments.

This article describes the background and a number of issues that have been met during the design and implementation of the Colorado Health Information Dataset (CoHID), a web-based health data warehouse.

BACKGROUND

Colorado Department of Public Health and Environment (CDPHE) provides public health and environmental protection services for the entire state. A variety of groups in this department are responsible for collecting and disseminating information such as population, birth, death, cancer, diabetes, and behavioral risk factors statistics and trends. Housed within this department is the Center for Health & Environmental Information and Statistics

(CHEIS). One of major duties of the CHEIS is to handle requests for health-related data from legislators, researchers, local public health professionals, the news media and general public. Historically, such needs were solely addressed by data analysts at CHEIS who did data analyses or obtained relevant information after talking with the requestor.

One of the greatest barriers to meet those data needs arises from the difficulty that the data requestors sometimes have in determining what their data needs are. Some of the data requests are complicated and responding fully to them are time consuming. Lack of resources and skillful data analysts often adds more difficulties to meet those data needs. So, CHEIS needs a way to reduce the cost and time associated with responding to these requests—while taking advantage of the data it has already captured.

One solution is the CoHID data warehouse, an interactive web site that provides customized public health information useful for research, legislation, and programs planning.

PROCESS FOR DEVELOPING COHID

CoHID is a joint effort by local, state public health agencies, and SAS Public Sector Pilot Program. Public health professionals, nurse, epidemiologists, statistical analysts from local and state public health agencies came together and made designing recommendations for the CoHID pilot project. Then a working team comprising of IT professionals and statisticians at the CDPHE and SAS staff, worked together to implement those ideas.

To start the project, the SAS Public Sector Pilot team met with representatives from CDPHE to determine the critical success factors that had to be met for the undertaking to be a success. They agreed that the prototype had to be provide core reporting and analysis functionality, using a thin-client architecture, with a user interface requiring minimal training, and include knowledge transfer of SAS web tool skills to be appropriate CDPHE personnel.

Within a week or so, the SAS Public Sector Pilot team delivered the web-based reporting and analysis prototype. Maintenance and other relevant knowledge were transferred to CDPHE personnel. Then, CoHID's team members at the CDPHE started to climb the learning curve and went to SAS training courses. Currently, CDPHE personnel can maintain the site and are able to add new functionalities with minimum help from SAS and its consulting company. A new data set, which provides statistics for birth defects surveillance system in Colorado, has been successfully added to the CoHID site by the CDPHE working group.

AN OVERVIEW OF COHID

CoHID is an interactive, web-based public health data warehouse that utilizes Hypertext Markup Language (HTML) forms and Common Gateway Interface (CGI) programs written using SAS web tools. CoHID was primarily designed for the non-technical users with a flexible, yet simple user interface. The users may choose from a number of data sets including death statistics, birth statistics, population data, Behavioral Risk Factors Surveillance data, and birth defects statistics. As time goes by, more data sets will be added continuously.

Using standard HTML form element such as drop-down menus, radio buttons, and text boxes, the user may choose different variables to build the query, experiment with the tool, and in doing so become more familiar with the data. The rapid return of results to user's computer screen encourages further experimentation, with a hope that the results that the user finally gets are meaningful and useful in the context of the problem at hand. Assurances are available if the user has difficulties to understand the data issues, interpret results, or to build a successful query. Explanatory documentations are provided at the CoHID site that detail data issues, statistical methods, and interpretations of results.

Before CoHID, health data have been available primarily through printed reports made available in periodical publications and specialized reports on a per-topic basis. CoHID makes data available through user-formulated queries in an interactive way that increases data accessibility considerably. It also reduces data analyst's effort to handle data requests from a variety of data users so they can put more efforts into solving technical statistical issues and adding more data sets to the CoHID site.

CONFIDENTIALITY OF DATA

Individual health data are often governed by laws and ordinances. The release of those data has a potential to reveal some information about individual data subjects. Confidentiality of individual health information is taken seriously by CoHID development group in developing the CoHID site. In addition to standard computer security measures, which are beyond the scope of this paper, several data disclosure limitation methods have been employed to prevent inappropriate attribution of information to a data subject.

The individual-level health data on the CoHID site have been de-identified, which means all names, addresses, social security numbers, and other sensitive information have been deleted prior to being placed on the web servers. When the frequencies in a table generated as the output of a query on the CoHID are relatively small, there is a potential that the information provided could allow inferences about individuals, thus threatening confidentiality of private health information. To handle this problem, statistical disclosure limitation methods have been used to hide results when low frequencies occur in the final results. In order to follow disease surveillance data release policy enforced at the CDPHE, specific race/ethnicity, gender, and age information are hidden when the total number of births in a geographical area are too small (less than 300 births) on birth defects statistics module.

MAINTENANCE OF COHID

In order to provide updated health information, or fix problems on the CoHID site, there is a need to modify existing data or SAS programs, and to add new data sets. CHEIS has a team working on CoHID projects. The team members include IT professionals, statisticians, general health professionals, and computer professionals. This team is responsible for updating statistical information, adding functionality, fixing problems found on current data modules, adding new data sets, and maintaining the web site. However, the governing board of the CoHID is responsible for making decisions about which data to be added, help get feedbacks from data users, and often provide recommendations and suggestions on relevant issues.

CHALLENGES

The invention of the electronic computer and the creation of the Internet have been said as the only two highly significant technological innovations for the health statistics over the past 50

years.⁶ The computer revolutionized statistical analysis but Internet revolutionized communication. Internet-based information dissemination technology is changing rapidly and further advances in SAS system will bring better web tools of delivering data to a wide range of public health data users.

At the same time, however, this revolution also poses challenges to a state health department. As mentioned before, implementing effective web-based queryable data system like the CoHID will require continued intensive effort and technical expertise.

The SAS web tools such as SAS IntrNet®, SAS ODS, and SAS Macros, HTML, and CGI are generally new to data analysts at state health departments. Most of them lack training in this area. To make things worse, when a variety of statistical functionalities are needed and provided on an interactive data warehouse, the data content needed for the system is complex. So, data analysts and statisticians often need to work together with IT professionals and computer professionals, or seek consultation from consulting company.

Unfortunately, IT professionals and computer professionals are often not familiar with epidemiological terminology and health statistics concepts. Unfamiliarity to each other's field between statisticians/data analysts and IT/computer professionals is often an obstacle for them to work together effectively. A strong leadership and good communications are essential for a successful project.

Availability of funding needed to develop and implement interactive data warehouse at the local and state level is another challenge. Continued funding including funding the software licensing or development costs, the necessary hardware, and the applicable information technology personnel is needed for ongoing operations and maintenance. As the data warehouse becomes more complex, the cost of development, ongoing operations, and maintenance will skyrocket. When state budget is tight for the state health department, these projects are often hit first.

CONCLUSION

SAS web tools such as SAS IntrNet and SAS ODS have created an unprecedented opportunity to provide an interactive, easy-to-use interface for access to public health data at the community level. Preliminary feedbacks indicated that interactive systems, such as the CoHID, have been received well by a variety of public health data users. The call for making local community data available and the recognized value of making data available through web will become driving forces for more development and use of such systems by state health departments.

TRADEMARKS

SAS and SAS IntrNet® 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.

REFERENCES

1. Simon, P.A., Wold, C.M., Cousineau, M.R., and Fielding, J.E. (2001), "Meeting the Data Needs of a Local Health Department: The Los Angeles County Survey," *American Journal of Public Health*, 91(12), 1950-1952.
2. The National Electronic Disease Surveillance System Working Group (2001), "National Electronic Disease Surveillance System (NEDSS): A Standards-Based

Approach To Connect Public Health and Clinical Medicine,” *Journal of Public Health Management Practices*, 7(6), 43-50.

3. Asaro, P.V., Land, G.H., and Hales, J.W. (2001), “Making Public Health Data Available to Community-Level Decision Makers – Goals, Issues, and a Case Report,” *Journal of Public Health Management Practices*, 7(6), 58-63.
4. Studnicki, J., Hevner, A.R., Berndt, D.J., and Luther, S.L. (2001), “Comparing Alternative Methods for Composing Community Peer Groups: A Data Warehouse Application,” *Journal of Public Health Management Practices*, 7(6), 87-94.
5. ORC Macro (2001), “Heuristic Evaluation of Query Interfaces for Analysis of Public Health Statistics,” *Special Reports to Epidemiology Program Office, Centers for Disease Control and Prevention*.
6. Florey, C.V., Shlomo, Y.B., and Egger, M. (2000), “Epidemiology and the World Wide Web: Is There a Net Benefit?” *Epidemiologic Reviews*, 22(1), 181-185.

ACKNOWLEDGMENTS

CoHID resides within the Colorado Department of Public Health and Environment and has been supported by Preventive Health Block Grant and Health Alert Network/Training funds. CoHID’s governing board includes representatives from both local and state health agencies. The author would like to express his appreciation to CoHID’s governing board, the development team, and SAS Public Sector Pilot team for launching the CoHID project successfully. Special thanks goes to the following people at the Colorado Department of Public Health and Environment:

Bob O’Doherty, Center Director of CHEIS

Alyson Shupe, Ph.D., Section Chief, Health Statistics Section

Rose Hammond, Statistical Analyst, Health Statistics Section

Pam Deandrea, IT Professional, System Analysis and Programming

Daniel Ondrejko, IT Professional, System Analysis and Programming

CONTACT INFORMATION

The author may be contacted at:

Huiyun Xiang, MD, MPH, Ph.D.
 Health Statistics Section
 Colorado Department of Public Health & Environment
 4300 Cherry Creek Drive S.
 Denver, Colorado 80246-1530
 Work Phone: 303-692-2169
 Fax: 303-691-7821
 Email: huiyun.xiang@state.co.us

CoHID web site can be accessed online at:

www.cdphe.state.co.us/sascohidweb/cohids.html