

# Visualizing Health: Enhancing Public Health through Weave Data Analysis and Visualization

Carol A. Gotway Crawford\*, Michael Smyser†, Georges Grinstein‡, Jim Ribble\*, Soyoun Park\*, Robb Chapman\*, Shweta Purushe‡, Patrick Ryan‡, Franck Kamayou‡, Ekaterina Galkina‡

\*Centers for Disease Control and Prevention, Division of Behavioral Surveillance, Public Health Surveillance and Informatics Program Office. †Public Health Seattle & King County and Department of Epidemiology at the University of Washington. ‡University of Massachusetts Lowell

## ABSTRACT

We present a sophisticated web-based analytical tool for use in public health and community decision support systems. This tool is based on Weave, a Web-based Analysis and Visualization Environment developed by a collaboration of computer scientists at the University of Massachusetts Lowell and members of the Open Indicators Consortium. In this presentation, we describe a new Weave component, the Analyst Workstation (AWS), being developed to enhance the data warehousing and analytical capabilities of the Weave environment. This work is being led by a collaboration of Weave developers and scientists, surveillance scientists at the U.S. Centers for Disease Control and Prevention and public health practitioners at Public Health – Seattle & King County. Our work focuses on adding strong and flexible analytic capacity to the existing Weave visualization software for purposes of community health assessment. The entire system is supported through other freely available open-source packages, but will also work with proprietary tools. Examples of Weave output and use of the AWS prototype will also be presented, as well as how the program design will be able to meet needs for automated health data assessment, and rapid response for web-visualized health data information.

**Keywords:** Health informatics, information visualization, visualization systems and tools, open source software.

**Index Terms:** H.1.2 [Information Systems]: User/Machine Systems – Human information processing; J.3.2 [Life and Medical Sciences]: Health

## 1 INTRODUCTION

Sources of public health data have rapidly increased with improved local and federal government commitments to make data available to the public [1,2]. Such data have been very beneficial to local public health organizations for local community health assessment, research into specific health trends or disease patterns and evaluation of local public health programs or interventions. The availability of so many new data sources presents many challenges in analysis and exploration, as well as presentation and dissemination. Improvements in Internet

technology now permit the public user as well as the analyst to explore the data in creative ways for understanding and assessment using dynamic displays that help a wide audience quickly grasp patterns, trends, or complex empirical information.

### 1.1 WHAT IS WEAVE?

Weave is a Web-based Analysis and Visualization Environment. It is an open-source computational and visualization package developed at the University of Massachusetts Lowell's Institute for Visualization and Perception Research (IVPR) in partnership with the Open Indicators Consortium (OIC), a collaborative of public and nonprofit organizations working to improve access to more and higher quality data [3,4]. Weave is freely available for download on Github [5] and may be used by any organization such as health departments or community health organizations that typically may not have resources to purchase or license proprietary software. These organizations include hospitals, clinics, medical offices, as well as state and local governments. Currently the OIC participates as a governing and advisory body that includes members from 18 community and governmental agencies. The Weave Users' Forum provides assistance to any user who has questions about the package. Weave allows for many types of interactive presentation formats which include standard displays of charts, graphs, mapping tools, as well as novel data mining tools such as RadViz [6] and data-driven document retrieval such as InfoMaps [7]. All visualizations and analyses can be embedded in non-web presentations. Perhaps the most important feature of Weave is its creation of "session states": an architecture that can record every event or interaction, including those with external tools having an Application Programming Interface (API). This allows users to quickly replicate all tasks leading to a particular visualization, and allows users to easily share the commands.

### 1.2 CHALLENGES: REALIZING WEAVE'S FULL POTENTIAL FOR ANALYZING AND VISUALIZING HEALTH DATA

Weave easily loads data that are ready for visualization and presentation. Although Weave can analyze data through an interface with the R-project package [8] or through exploratory visualizations, it did not, in its original incarnation,

handle complex data processing routines. Consequently, some datasets had to be pre-processed before visualization with Weave and this processing and analysis required using other analytical software. As users may also require rapid data-to-visualization output, it became important to add enhanced analytical capability to Weave that would allow the use of customized analysis routines, processing and linking of large, complex data systems, and easy storage and access of data and analytical routines.

## **2 METHODS: ADDRESSING THE CHALLENGE**

Recognizing the need for a tool like Weave for clinical decision support and community health needs assessment, CDC's Division of Behavioral Surveillance (DBS) became an active member in the collaborative Weave development process. DBS manages the Behavioral Risk Factor Surveillance System (BRFSS), the world's largest, on-going telephone health survey system [9].<sup>1</sup> Thus, an additional goal of CDC participation in the Weave development process was to allow flexible analysis and visualization of BRFSS data in a cloud-based environment. Open-source was deemed a critical component in order to leverage costs and expertise. At the same time, Public Health – Seattle & King County (PHSKC) needed to revise their analysis, visualization and dissemination system that provides an evidence base for health protection (tracking and preventing disease and other threats; regulating dangerous environmental and workplace exposures; and ensuring the safety of water, air and food), health promotion (leading efforts to promote health and prevent chronic conditions and injuries) and health provision (helping assure access to high quality health care for all populations) [10]. PHSKC reports on nearly 100 community health indicators which are presented on the departmental web pages. These indicators are analyzed for trends over time as well as for differences within demographic subgroups (e.g., age, sex, race and ethnicity) and geography [11]. In addition, PHSKC receives hundreds of requests every year for customized health assessment reports to meet the need for department funding or evaluation needs. Requests also come from local community health and other organizations, generally needed to support grant proposals or other funding requests. Much of this information is produced using legacy software created by PHSKC called VistaPHw which performs data analysis of vital record and population data, but does not produce output in formats that are ready for display or immediate use. Updates and revisions to community health indicator information are time-consuming, but need to be performed at least annually as new data are received.

Thus, in response to CDC and PHSKC needs and requirements, the IVPR team has begun development of the Analyst Workstation (AWS).

## **3 RESULTS: THE ANALYST WORKSTATION (AWS)**

With the CDC and PHSKC needs and requirements, the IVPR team began development of the Analyst Workstation (AWS). As Weave is open-source, this workstation is now being developed through active engagement by members of Open Indicators Consortium (OIC) through the Agile software development process [12] where the project is driven by member needs and participation. All worked together to develop an architecture extending Weave and a prototype to meet the collective and diverse needs of the CDC and PHSKC as well as other local health departments. The AWS is initially being developed to use data from the Behavioral Risk Factor Surveillance System since it is the largest source of state and local health data, but later releases will include enhancements to work with virtually any dataset with accompanying metadata. Analytical developments have focused on the use of the R statistical package, but additional modifications will allow the use of STATA and SAS. The system is being designed to include a comprehensive data analysis and presentation system which includes data warehouse capabilities or online data access, cloud-based use and user-defined analysis routines and visualizations.

## **4 IMPACT**

The Analyst Workstation will greatly enhance public health decision-making capability by allowing users to now fully analyze, visualize and report data in an integrated framework. It can be used to assist public health professionals in determining disease impact, recognizing disease clusters, and identifying populations and areas most affected or at risk. The result is a fully integrated data analysis and visualization system which can also be used to help prioritize and direct critical resources, and develop effective community based initiatives to improve the public's health. Perhaps most importantly, the entire Weave platform has been driven by the needs of, and in collaboration with, state and local public health practitioners. The end result of this collaboration is an open-source data analysis and visualization environment that is sophisticated, intuitive and adaptable to local public health needs.

## **5 CONCLUSION**

The collaborative efforts of Weave developers and OIC members, especially the CDC and Public Health – Seattle & King County, have been successful in the creation of designs and initial programming for adding strong and flexible analytic capacity to the existing Weave visualization software. The new Weave component, the Analyst Workstation (AWS), is being developed to analyze and dynamically display in a web-based format any public health or other data that can be loaded into the software. The entire system is integrated with other open-source packages including MySQL

(data storage and management) [13], Apache Tomcat (webserver) [14] and R (data analysis) [15], but will also work with proprietary tools such as Stata and SAS. The open-source Weave software with the addition of the AWS has potential for meeting health assessment needs of community and governmental organizations worldwide including automated health data assessment and rapid response for web-visualized health data.

## REFERENCES

- [1] U.S. Department of Health & Human Services. HealthData.gov. [Online]. Available: <http://www.healthdata.gov>.
- [2] U.S. Centers for Disease Control and Prevention. National Center for Health Statistics. Public-Use Data Files and Documentation. [Online]. Available: [http://www.cdc.gov/nchs/data\\_access/ftp\\_data.htm](http://www.cdc.gov/nchs/data_access/ftp_data.htm).
- [3] Weave. [Online]. Available: <http://info.oicweave.org/projects/weave>.
- [4] Open Indicators Consortium. [Online]. Available: <http://oicweave.org>.
- [5] GitHub. [Online]. Available: <https://github.com/about>.
- [6] Sharko J, Grinstein G, Marx KA. Vectorized Radviz and its application to multiple cluster datasets. *IEEE Trans Vis Comput Graph*. 2008 Nov-Dec;14(6):1444-51.
- [7] Kolman S, Dufilie AS, Anbalagan SK, Grinstein G, "InfoMaps : A Session Based Document Visualization and Analysis Tool," *iv*, pp.274-282, 2012 16th International Conference on Information Visualisation, 2012. [Online]. Available: <http://www.computer.org/csdl/proceedings/iv/2012/4771/00/4771a274-abs.html>.
- [8] The R Project for Statistical Computing. [Online]. Available: <http://www.r-project.org>.
- [9] U.S. Centers for Disease Control and Prevention. Behavioral Risk Factor Surveillance System. [Online]. Available: <http://www.cdc.gov/brfss>.
- [10] Public Health – Seattle & King County. Statement of Mission. [Online]. Available: <http://www.kingcounty.gov/healthservices/health/healthoffice/mission.aspx>.
- [11] Public Health – Seattle & King County. Community Health Indicators. [Online]. Available: <http://www.kingcounty.gov/healthservices/health/data/indicators.aspx>.
- [12] Beck, Kent; et al. (2001). "Manifesto for Agile Software Development". Agile Alliance (<http://agilemanifesto.org/>).
- [13] MySQL. [Online]. Available: <http://www.mysql.com>.
- [14] Apache Tomcat. [Online]. Available: <http://tomcat.apache.org>.
- [15] The R Project for Statistical Computing. [Online]. Available: <http://www.r-project.org>.

---

1 U.S. Centers for Disease Control and Prevention. Behavioral Risk Factor Surveillance System (<http://www.cdc.gov/brfss/>).