Using Secure Web Services to Visualize Poison Center Data for Nationwide Biosurveillance: A Case Study

Thomas G Savel[1], Alvin Bronstein[2], William Duck[1], M. Barry Rhodes[1], Brian Lee[3], John Stinn[3], Katherine Worthen[4]

[1] Centers for Disease Control and Prevention, Atlanta, GA
[2] Rocky Mountain Poison and Drug Center, Denver, CO
[3] Deloitte Consulting LLP, Atlanta, GA
[4] Ciber, Inc. Greenwood Village, CO

Abstract

Objectives: Real-time surveillance systems are valuable for timely response to public health emergencies. It has been challenging to leverage existing surveillance systems in state and local communities, and, using a centralized architecture, add new data sources and analytical capacity. Because this centralized model has proven to be difficult to maintain and enhance, the US Centers for Disease Control and Prevention (CDC) has been examining the ability to use a federated model based on secure web services architecture, with data stewardship remaining with the data provider.

Methods: As a case study for this approach, the American Association of Poison Control Centers and the CDC extended an existing data warehouse via a secure web service, and shared aggregate clinical effects and case counts data by geographic region and time period. To visualize these data, CDC developed a web browser-based interface, Quicksilver, which leveraged the Google Maps API and Flot, a javascript plotting library.

Results: Two iterations of the NPDS web service were completed in 12 weeks. The visualization client, Quicksilver, was developed in four months.

Discussion: This implementation of web services combined with a visualization client represents incremental positive progress in transitioning national data sources like BioSense and NPDS to a federated data exchange model.

Conclusion: Quicksilver effectively demonstrates how the use of secure web services in conjunction with a lightweight, rapidly deployed visualization client can easily integrate isolated data sources for biosurveillance.

Keywords: *public health, surveillance, architecture, web services, SOA*

Introduction

Public health surveillance is the ongoing, systematic collection, analysis, interpretation, and dissemination of data about a health-related event for use in public health action to reduce morbidity and mortality and to improve health.[1] Surveillance provides the critical information necessary to support several public health functions, including outbreak detection, estimating the impact of a disease or injury, portraying the natural history of a health condition, determining the distribution and spread of illness, facilitating public health interventions, evaluating the effectiveness of those interventions, and facilitating planning.[2] Historically, public health information systems supporting surveillance activities have been constructed to support specific program areas within health departments. These systems were often populated with data reported from healthcare providers, often via paper based forms, not surprisingly, leading to numerous "siloed" surveillance systems. Having multiple related but disparate surveillance systems, creates many inefficiencies in the areas of analysis and communications, which become most apparent during public health emergencies. With over 3,000 state and local public health agencies, tens of thousands of other public health stakeholders (including clinical providers, laboratories, first responders, and researchers) the goal of creating an efficient, secure, and flexible centralized, nationwide biosurveillance system which meets the needs and expectations of all users becomes essentially insurmountable.

Today, this lack of integration presents major vulnerabilities in the healthcare system's ability to rapidly detect and mitigate health emergencies. The goal in creating CDC's BioSense system has been to mitigate this existing vulnerability. Within the BioSense system, patient level data (specifically, chief complaint and diagnosis data) are transmitted securely, and subsequently stored and maintained in a centralized data warehouse. Various analyses are then performed, with results being shared with BioSense users as well as the original data providers. This model has proven difficult to scale, costly to maintain, and challenging for end users. It is the authors' perspective that enhancement to the current functionality in existing biosurveillance applications (e.g., Essence, RODS, or BioSense) with secure web services, distributed computing, Grid architectures, and an open federated model for data access and exchange is one approach that could potentially correct many of their current limitations.

One of the principal challenges facing existing biosurveillance implementations is their limitation in sharing data across multiple jurisdictions to support real-time surveillance activities at a national level. The ability to generate actionable health intelligence by increasing access to data and fusing multiple data streams to yield effective situational awareness is a key priority to the National Biosurveillance Strategy for Human Health.[3] Efforts to move legacy biosurveillance application models from centralized and historically siloed environments to federated and distributed models of data exchange are under way in order to effectively provide timely population-wide biosurveillance coverage. Federation of data enables local jurisdictional control over data procurement and stewardship while permitting secure access to approved external entities (e.g., state and federal agencies). Finding new methods and technologies to augment existing biosurveillance capabilities to bridge this data exchange gap has been a focus of ongoing research at CDC in collaboration with its national and international partners.

Methods

In 2008 CDC's National Center of Public Health Informatics (NCPHI) embarked on a collaborative initiative with BioSense and the American Association of Poison Control Centers (AAPCC) to visualize aggregate clinical effect data based on telephone calls to the 61 poison control centers across the nation.[4,5] AAPCC poison centers maintain a 24-hour poison help hotline staffed by pharmacists, physicians, and toxicology specialists that provide exposure management instructions and general poison information to patients. As part of their current workflow, all poison case information (clinical effect, demographics, medical outcomes, therapeutic interventions, etc.) is stored in the National Poison Data Base (NPDS) poison surveillance database and then summarized into an annual report.

Prior to this initiative, poison control data was unavailable for biosurveillance visualization and analytics purposes at a national level. Through the creation of a secure web service in conjunction with a lightweight, Java-based visualization client known as Quicksilver, CDC users were provided the tools to visualize aggregated clinical effect data both geospatially and via a time series chart. The project was able to successfully demonstrate how a federated data approach could enhance existing biosurveillance capacity by providing information access between the BioSense program and the NPDS.

Creation of the Quicksilver tool was a two-fold effort that involved close collaboration between the AAPCC's NPDS development team and the NCPHI Public Health Grid (PHGrid) team. The first effort involved creation of a web service that could provide aggregate clinical effect counts from multiple geographic regions (state, Zip-3, Zip-5) within a single query from the NPDS. Once this was completed, development then began on a Simple Object Access Protocol (SOAP) service oriented architecture (SOA) client that could accept and visualize the query results returned from the NPDS web service. The visualization client was created in Java, and is deployable to any J2EE Java Servlet container (i.e., is Java servlet 2.0 compliant). The client was also specifically tested using the Apache Tomcat application server (v5.5 and v6.0). The user interface is browser-based, leveraging HTML and JavaScript, and was tested with the Internet Explorer, Firefox, Chrome, and Safari web browsers. Quicksilver has been successfully deployed within the CDC network, with user access controlled directly by the Quicksilver business steward. A diagram outlining the relationship between the NPDS web service and visualization client can be seen in Figure 1.

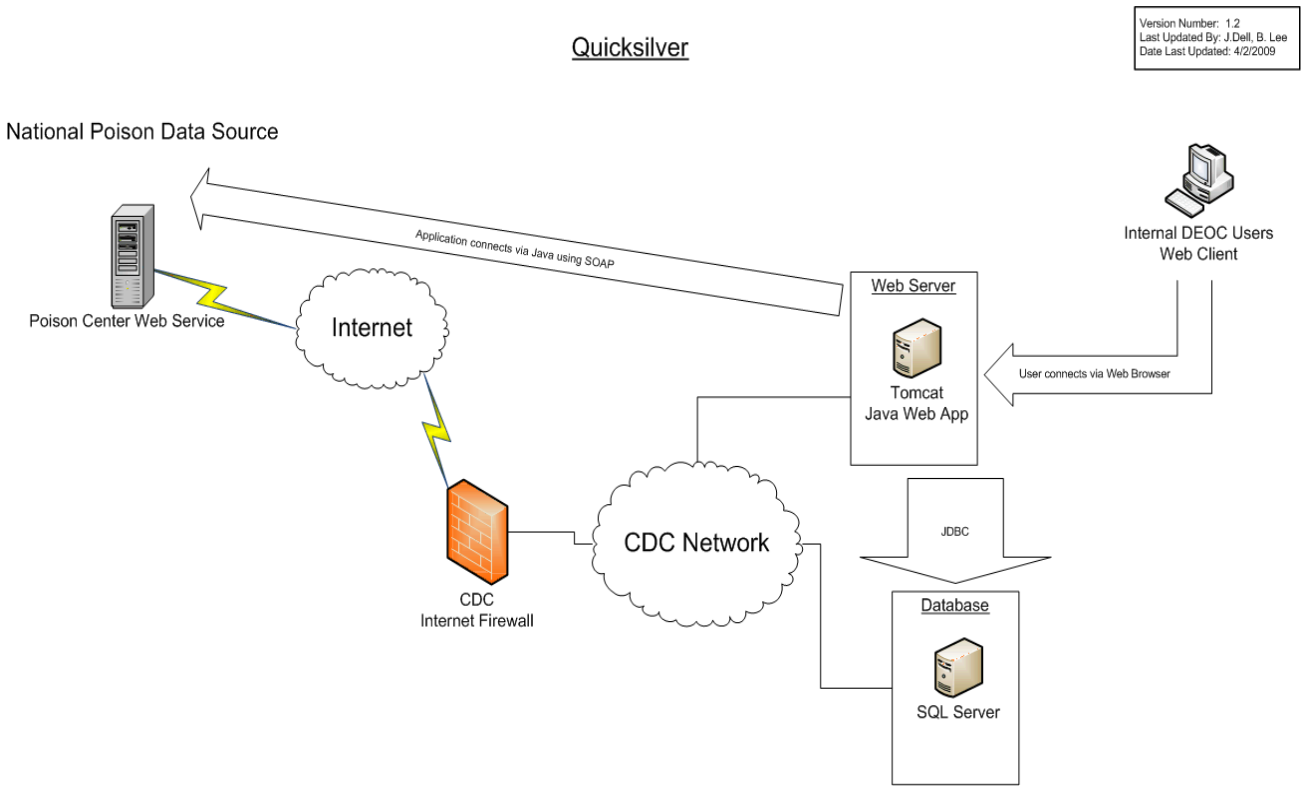

Figure 1. Quicksilver Logical Deployment Diagram

The visualization client has a small database that stores static data such as geographic information consisting specifically of spatial polygons, centroids, and list relationships for state, ZIP-3, and ZIP-5 locations. This database is quite minimal (<100MB) and is only used for assistance with drawing shapes. No poison center data is stored at the CDC. The client uses the Google Maps JavaScript API [6] to visualize the aggregate clinical effect counts within any selected geographic region. The Google Maps JavaScript API is used, rather than the Google Maps KML API to prevent any poison center data from being sent to the Google Maps internet server. To enhance the performance of the client as a user zooms in and out of different geographic entities, use of the Douglas-Peucker algorithm was applied to reduce the number of onscreen points drawn, thus reducing the rendering time of map polygons. [7] Because the visualization client was built upon Asynchronous JavaScript and XML (AJAX) APIs that support Google Maps, performance under Internet Explorer had to be enhanced to achieve timely rendering of a selected geographic region.

Visualization of aggregate clinical effects data was also achieved in a time series chart using Flot [8], an open source AJAX API used to produce graphical plots of arbitrary datasets in real-time. A modified version of the Early Aberration Reporting System (EARS) C2 algorithm was used to detect aggregate clinical effects counts $\geq$ 4 standard deviations (SD) from the mean call volume using a rolling 28-day average. [9] Data points which were 4 SD units above or below the mean were automatically highlighted (in red) to facilitate rapid examination. All statistical calculations were performed using Java's standard math libraries.

To optimize performance, the NPDS web service results from frequent user queries were cached in RAM. This transient, in-memory cache is used to reduce the number of transactions sent to the remote web service. The cache only exists temporarily and is cleared and reset each time the application server is stopped or started.

Results:

Two iterations of the NPDS web service were completed in 12 weeks. The visualization client, Quicksilver, was developed in four months, with client releases being reviewed by CDC and AAPCC representatives on a weekly basis.

When a user first enters Quicksilver, they select a geographic location (state, ZIP-3, ZIP-5), date range, and clinical effect. The NPDS web service then returns the query results within several seconds and shows a choropleth map shaded according to clinical effects thresholds established by the user (Figure 2).

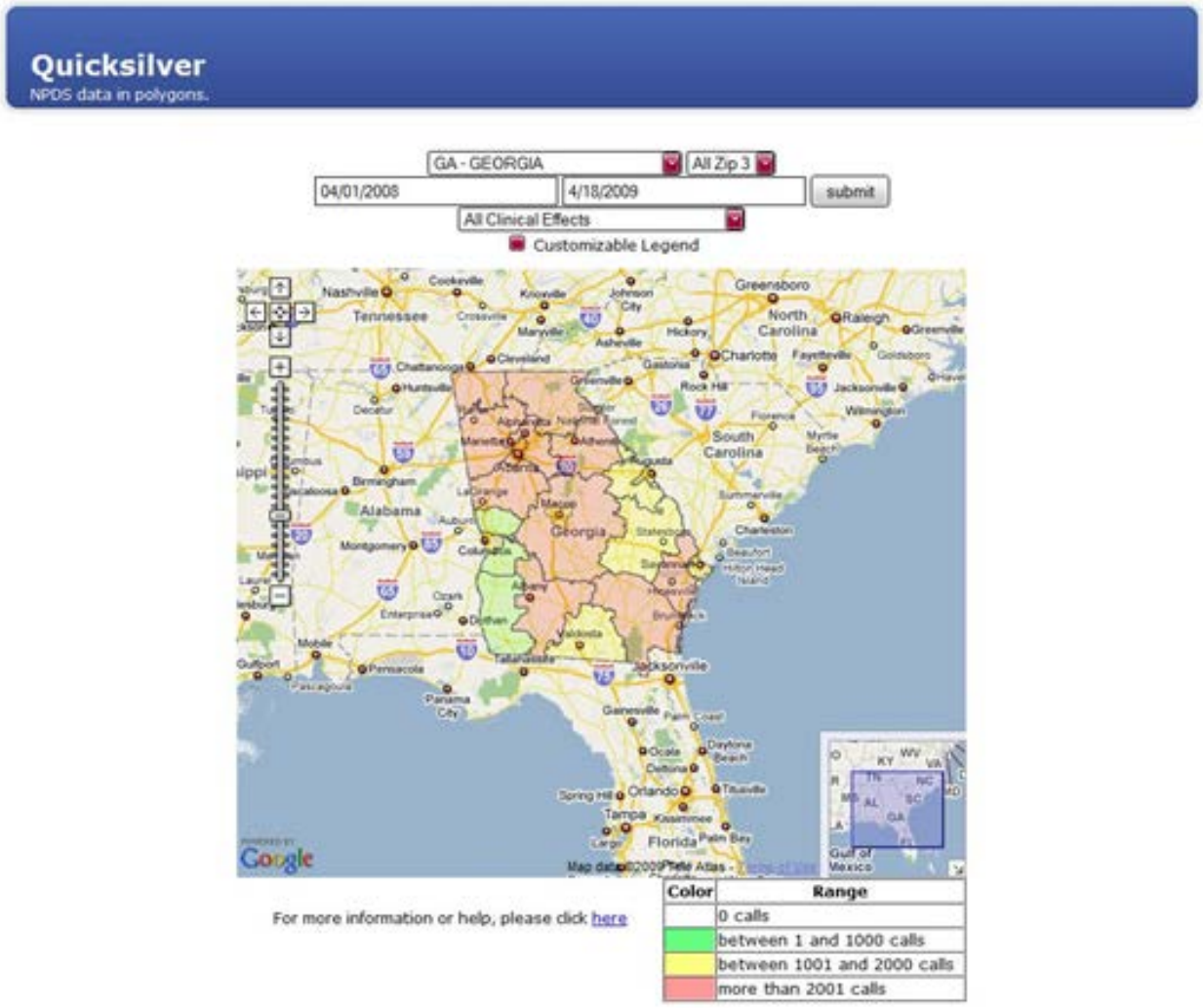

Figure 2. Choropleth Map of Georgia in Quicksilver

Each shaded area of the map represents a unique ZIP-3 or ZIP-5 region within the selected state. To show a time series of each respective ZIP-3 region within the state of Georgia, the user clicks anywhere within the ZIP-3 or ZIP-5 area and a pop-up time series graph is then displayed using the Flot open source charting API previously described (Figure 3).

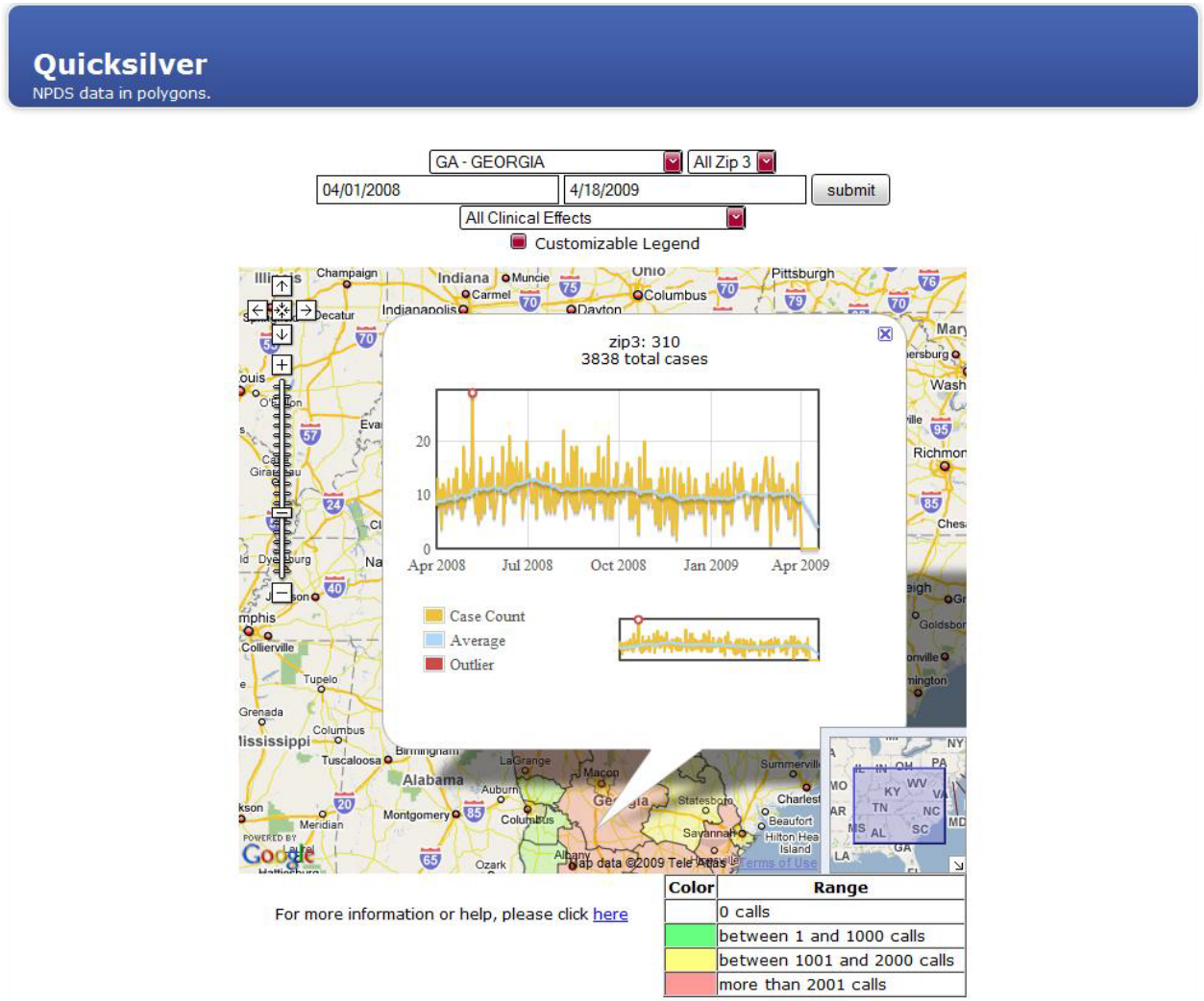

Figure 3. Time Series Graph of ZIP-310 in Georgia

Flot allows the user to also zoom into any date range within the time series to examine clinical effects outliers which fall outside of the empirically set statistical threshold (Figure 4).

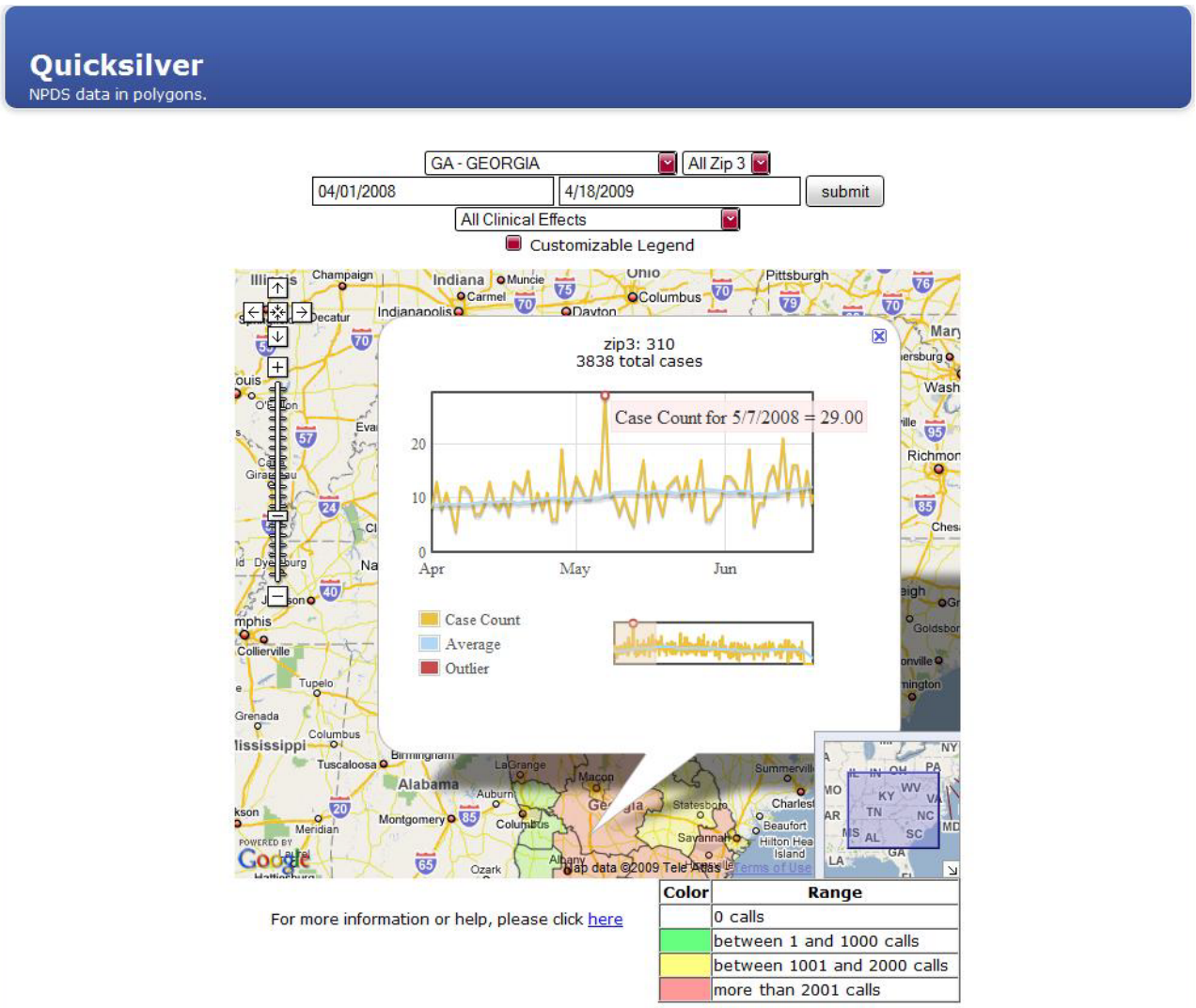

Figure 4. Magnified Time Series Chart Showing Clinical Effects Outlier

When we used Quicksilver to visualize all nausea data collected by NPDS in 2008 by state, over 13 significant (>10 SD units above 28-day moving average) increases were found.

Discussion:

The use of secure web services to enable access to previously isolated data sources for biosurveillance consumption is an emerging and appealing concept. This model aligns itself with ongoing, existing efforts within healthcare IT domain such as the Nationwide Health Information Network (NHIN). NHIN's vision of providing healthcare information exchanges among hospitals, laboratories, and independent healthcare providers is based upon federated data access and the use of secure web services in accordance with Web Services Interoperability Organization (WS-I) standards. The same protocols that were used to create the NHIN Gateway SOA infrastructure were used to create the web service described in this document. It is vital

that biosurveillance and healthcare IT infrastructures remain synchronized moving forward to ensure access and interoperability amongst web services deployed through the NHIN gateway or other related entities (e.g., PHGrid nodes).

Although this document focuses on the development of the NPDS web service, this model can be applied to other data sources which have the potential of providing valuable biosurveillance information. Other web services are currently under development, and can be accessed through the PHGrid Service Registry. [10] Although still at an early stage of development, service registries such as this, provide a method to easily survey what data sources are available for integration into applications for detection, analysis, and visualization of public health threats. These services adhere to a common set of WS-I and W3C standards (e.g. SOAP, XML, Schema, WSDL, WS-Security) enabling heterogeneous technology platforms to communicate without developing intricate adapters to bridge different implementations. Because of this adherence to standards, small teams were able to extend the functionality of a .NET based web service using Java based analysis and visualization techniques.

As the National Biosurveillance Strategy continues to move forward, all public health stakeholders should continue to support enabling technologies as they provide the means to address the priority areas of the National Biosurveillance Strategy. Ongoing informatics research should accelerate exploration of new methods to enhance how the healthcare community transforms, manipulate, and analyzes unstructured data sources. These efforts will yield actionable biosurveillance intelligence that can be used to achieve optimal situational awareness capability. Aggressive financial and intellectual investments need to be made on the behalf of the healthcare community to apply these technology tools to current biosurveillance infrastructures. Without these investments, the social and economic costs associated with recent disease outbreaks will likely only continue to escalate.

This implementation of web services combined with a visualization client represents incremental positive progress in transitioning national data sources like BioSense and NPDS to a federated data exchange model. It is the authors' hope that next steps will include extending existing web services into Grid-enabled web services to be then used over a secure and robust Grid infrastructure.

Legacy biosurveillance architectures should consider this approach as it establishes a foundation for an integrated information sharing environment with minimal financial and development resource requirements. Although data access services are extremely valuable, this approach also opens the door to the development of many other types of web services. Viable candidates for web service development and distribution over a Grid infrastructure include, but are not limited to, aberration detection, epidemiological analysis, standardized vocabularies, and visualization toolsets. Future biosurveillance applications will likely require on-demand access and query capability to multiple data and analysis streams. Implementing dynamic web services which can allow consumers to combine those streams however they see fit for their own biosurveillance purposes is the ultimate goal of the approach described in this document.

Conclusion:

Quicksilver effectively demonstrates how the use of secure web services in conjunction with a lightweight, rapidly deployed visualization client can easily integrate isolated data sources for biosurveillance. The technologies deployed to build Quicksilver provided a model to visualize NPDS data in a way that was not previously available to CDC nor the AAPCC. [11] Grid-enabled web services can allow agencies like the AAPCC to remain focused on their primary goals of data stewardship and the provisioning of clinical effects data through NPDS. Agencies like CDC which depend on this data for biosurveillance and analysis purposes can then easily and quickly obtain access to this information without the responsibility of securing third party information, and thus dramatically reducing data warehousing and processing costs. Overall, this project demonstrates the use of low-cost, open source, and federated technologies to create a valuable public health application that provides analysis and visualization of relevant data sources for biosurveillance consumption.

Acknowledgments

The authors wish to thank the many collaborators on this initiative including members of the NCPHI Grid Team. This project was supported in part by the CDC Contract # 200-2006-F016934.

References

[1] CDC. Updated guidelines for evaluating public health surveillance systems: recommendations from the guidelines working group. MMWR 2001;50(No. RR-13).
[2] Teutsch SM, Churchill RE. Principles and practice of public health surveillance. 2nd ed. Oxford, New York: Oxford University Press, 2000.
[3] Arthur R., Bailey S., Bell, B., Bernhardt J., Besser R., et al. National Biosurveillance Strategy for Human Health 2008-2013, Version 1.0 / December 2008. Available at http://sites.google.com/site/nbshh10/ Accessed November 5, 2009.
[4] National Poison Data System. Available at: http://www.aapcc.org/dnn/NPDS/NationalPoisonDataSystemInformation/tabid/311/Default.aspx. Accessed April 15, 2009.
[5] BioSense. Available at: http://www.cdc.gov/BioSense/. Accessed April 15, 2009.
[6] Google Maps API. Available at: http://code.google.com/apis/maps/. Accessed April 24, 2009
[7] Douglas-Peucker algorithm. Available at: http://en.wikipedia.org/wiki/Ramer-Douglas-Peucker_algorithm. Accessed April 15, 2009.
[8] FLOT. Available at: http://code.google.com/p/flot/. Accessed April 15, 2009.
[9] Yiliang Z., Wang W., Artrubin D., Wu Y. Initial Evaluation of the Early Aberration Reporting System --- Florida. MMWR Supplement 2005; 54(Suppl);123-130.
[10] Public Health Informatics Research Grid Wiki, Service Registry. Available at http://sites.google.com/site/phGrid/Home/service-registry. Accessed April 15, 2009.
[11] American Association of Poison Control Centers. Available at http://www.aapcc.org/DNN/. Accessed April 15, 2009.

doi: 10.5210/ojphi.v2i1.2920

Cite this item as: Savel, T., Bronstein, A., Duck, W., Rhodes, M., Lee, B., Stinn, J., & Worthen, K. 2010 Apr 9. Using Secure Web Services to Visualize Poison Center Data for Nationwide Biosurveillance: A Case Study. *Online Journal of Public Health Informatics* [Online] 2(1):e5.