Advanced Modular Design for Scalable Biosurveillance Systems

BY Reis, PhD, C Kirby, MA, E Sprecher, CA Cassa, MEng, J Brownstein, PhD, W Simons, MS
L Jordan, KD Mandl, MD MPH

Harvard-MIT-Children's Hospital Boston Informatics Program, Harvard Medical School

OBJECTIVE

This paper describes a modular approach to surveillance system design, ensuring flexibility, scalability and fault tolerance.

BACKGROUND

New data sources and analytical methods are continually being added to syndromic surveillance systems, and connections to data sources are constantly failing. In order to deal with these real-world challenges, biosurveillance systems must be scalable to incorporate new data sources and analysis methods, and robust to minimize performance loss whenever data sources inevitably fail.

METHODS

The workflow in the Automated Epidemiological Geotemporal Integrated Surveillance (AEGIS) system is divided into four modular stages: Data Sources, Predictive Modeling, Aberration Detection/Alarms, and Clients. The data sources, modeling methods, and aberration detection approaches used in each of these stages are selectable in a modular fashion. All communication between these stages occurs through a central database. The *Source Manager* manages the processing of data from the various data sources, and can scale to handle a large number of data sources. As specified by the information in its configuration table, the Source Manger reads data, converts it into the AEGIS format, pre-processes the data, including geo-coding and syndromic coding, and stores the data into the database. Timestamps are stored for each value in the database; thus, in the event of a data source failure, all other parts of the system can continue to operate, but the user is told whenever results are based on out-of-date data. The *Prediction Manager* generates expected values for the data being monitored. It can scale to handle a large number of different modeling techniques. Multiple input signals can flow into a single prediction model, enabling signal integration and multivariate modeling. Often, data from a particular source will not be available for a range of historical dates. This missing data can be imputed, e.g. as an average of the same date on the previous and following years. The *Alarms Manager* compares observed values with expected values and generates an alarm if appropriate. The configuration portion of the central database specifies which alarm methods are applied to which observed values and to which predicted values. This setup allows a combinatoric pairing of various alarm methods with various prediction models. The *Client Manager* interacts and

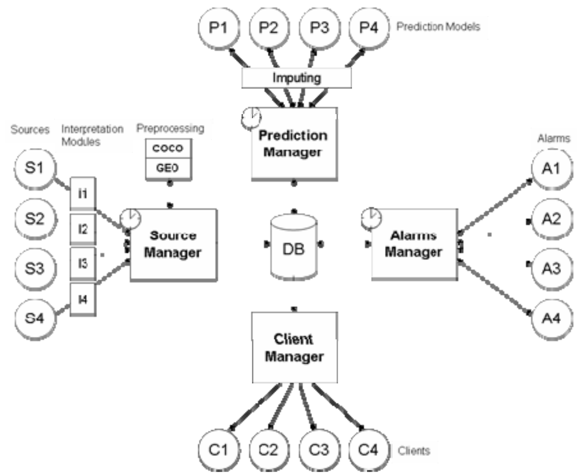

Figure 1. The modular architecture of AEGIS ensures flexibility, scalability and robustness.

communicates with users through various clients. Clients include the AEGIS Graphical Client, which displays public health and alarms data as requested by users; the AEGIS Alarms Client, which sends emails out to the Massachusetts Alert Network (MAN) with near real-time alarms; The AEGIS Administration Client enables configuration of access permissions, data source management, models and alarms setup; and the AEGIS Data Quality Client monitors the quality, accuracy and completeness of the data in the AEGIS system, reporting on the status of the various live data feeds that serve as inputs to the system.

RESULTS

At present, AEGIS monitors emergency department data from eleven hospitals and one large managed care organization in Massachusetts, and data from hospitals in New York City, the District of Columbia and Miami. Modeling approaches implemented so far include trimmed-mean seasonal models and ARIMA models [1]. Detection approaches implemented include EWMA and other temporal filters [2].

CONCLUSIONS

Modular design paradigms can help operational surveillance systems meet the critical challenges of scalability, flexibility and fault tolerance.

REFERENCES

[1] Reis B.Y., Mandl K.D. (2003) Time series modeling for syndromic surveillance. BMC Medical Informatics and Decision Making. Jan 23;3(1):2.

[2] Reis B.Y., Pagano M, Mandl K.D. (2003) Using temporal context to improve biosurveillance. Proceedings of the National Academies of Science U S A. 100(4):1961-1965.