



> Semantic Web Use Cases and Case Studies

Case Study: Semantic Web Technology for Public Health Situation Awareness

Parsa Mirhaji, School of Health Information Sciences, University of Texas,

March 2007



THE UNIVERSITY of TEXAS
SCHOOL OF HEALTH INFORMATION
SCIENCES AT HOUSTON

General Description

Public health surveillance is the ongoing collection and analysis of data to identify and respond to community health problems in a timely manner. Frequently, important public health information is distributed among many disparate databases, and includes far more than just clinical data. Infectious, gastrointestinal or respiratory illnesses may be detected by monitoring emergency room visits, water contaminations and air quality. Also, some helpful correlative data may exist online (e.g. pollens and allergens). As of now, meaningful integration of all these disparate data to inform public health practice has been the major barrier to their use in public health surveillance systems.

As the Semantic Web provides a suite of technology to enable information representation and computer reasoning, we have been able to automate many of the most challenging, resource intensive and cumbersome aspects of information integration in such a way that the system learns from the data itself as to how they should or would be integrated with other data.

Using Semantic Web technology, we have integrated unstructured text such as doctors' notes, and patient complaints into structured electronic medical records. This has enabled retrieval of all information together and through a unified query interface, regardless of how (structured, unstructured) and where (text files, database tables, spreadsheets, etc.) the data is stored.

It has not only enabled us to integrate and utilize any relevant dataset regardless of source, structure and format, but has also made it possible to contextualize their use for a variety of different tasks, by developing abstractions and models on top of integrated data (called an ontology). For example: we have created models for identifying Influenza-Like-Illnesses, Neurological, Gastrointestinal and other categories of patients (or outbreaks), and models for identifying respiratory distress patients (and outbreaks) based on multi-source data collected from 18 different air quality sampling devices and data feeds from 16 clinical information systems. This enables real-time monitoring and notification of Ozone related asthma and other respiratory distresses.

As a result, our prototype system, code named SAPPHIRE (Situation Awareness and Preparedness for Public Health Incidents using Reasoning Engines), is the only instance of an Integrated Biosurveillance System where all surveillance activities are enabled through a distributed and collaborative system. SAPPHIRE absorbs data from multiple disparate sources and puts them together as pieces of a larger puzzle that can be viewed from many different perspectives (traditional notifiable disease surveillance, syndromic surveillance, environmental protection, environmental epidemiology, biosecurity and bioterrorism preparedness, veterinary surveillance, etc.), or as a bird's eye view to the community health, for decision and policy purposes. Recently we have made significant steps in integrating biomedical information (e.g. gene expression, proteomics and genomics) in order to project and identify potential population effects of indicators found in basic science research (translational biomedicine).

If a new situation arises, SAPPHIRE can be modeled to dynamically absorb new data feeds, make new interpretations and produce new results accordingly. This dynamic instantiation and configuration of services was demonstrated during hurricane

Katrina, when SAPPHIRE was extended to accommodate data from a just-in-time PDA based questionnaire, a clinical information system from Katrina shelters and surveillance reports captured by the Houston Department of Health, and to support signal detection and reporting on diseases and syndromes defined by field epidemiologists. SAPPHIRE was the sole surveillance technology to address these needs within eight hours of the shelters opening, using Internet and Semantic Web technology.

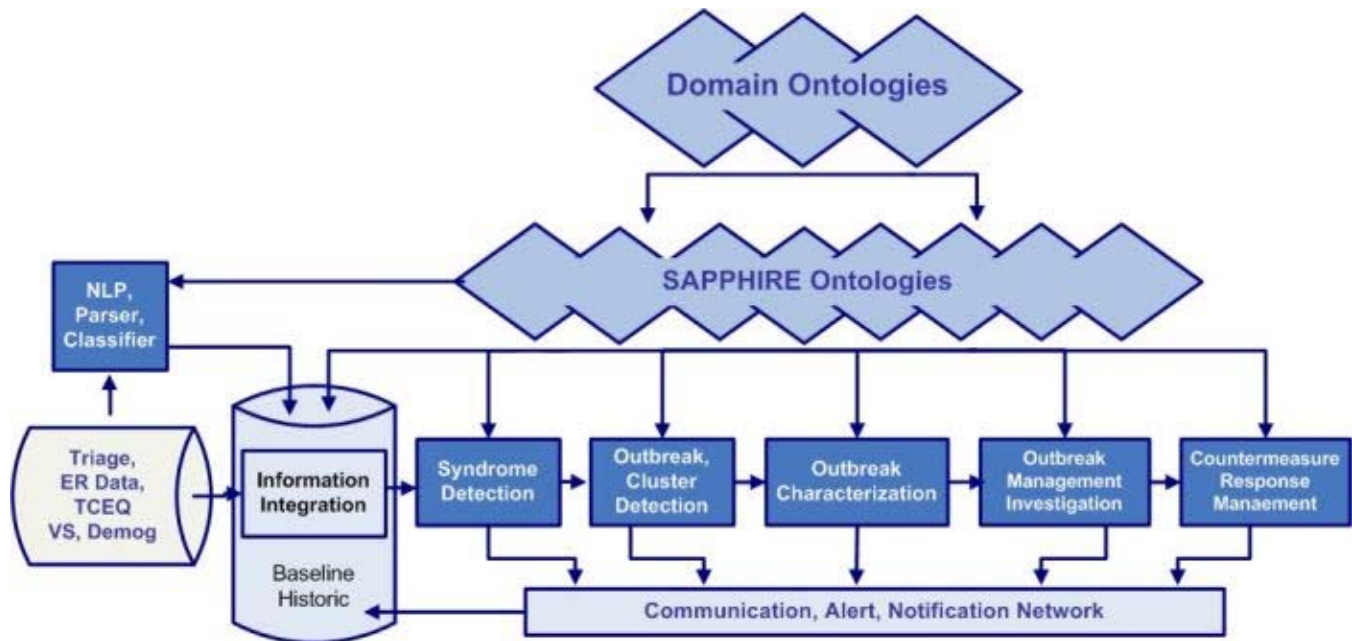


Figure 1: SAPPHIRE Ontology Driven Architecture (a somewhat larger version of the figure is also available)

Key Benefits of Using Semantic Web Technology

Use of Semantic Web technology has enabled the Center for Biosecurity and Public Health Informatics Research at the University of Texas Health Science Center at Houston to design a system architecture with the following features:

Distributed collaboration and interoperability:

disparate and heterogeneous data can be exchanged, integrated and utilized seamlessly and dynamically between remote systems.

Dynamic adaptability:

new requirements, data, functions and tasks can be introduced to the system without major rewrite or reprogramming to address novel situations or new tasks.

Multidisciplinary reuse of information:

existing data in the system can be repurposed to address unprecedented use cases.

Human computer interaction:

systems interact intelligently with human users and more effectively, intuitively and easily.