Design Science Methodology Applied to a Chemical Surveillance Tool

Zhuanyi Huang, Kyungsik Han, Lauren Charles-Smith, and Michael J. Henry
Pacific Northwest National Laboratory
zhuanyi.huang@pnnl.gov
kyungsik.han@pnnl.gov
lauren.charles@pnnl.gov
michael.j.henry@pnnl.gov

CHI'17 Extended Abstracts, May 6–11, 2017, Denver, CO, USA.
Copyright © 2017 ACM ISBN 978-1-4503-4656-6/17/05 ...$15.00.
http://dx.doi.org/10.1145/3027063.3053263

Abstract
Public health surveillance systems gain significant benefits from integrating existing early incident detection systems, supported by closed data sources, with public available data. However, identifying potential alerting incidents relies on finding accurate, reliable sources and presenting the high volume of data in a way that increases analysts work efficiency; a challenge for any system that leverages publicly available data. In this paper, we present the design concept and the applied design science research methodology of ChemVeillance, a chemical analyst surveillance system. Our work portrays a system design and approach that translates theoretical design methodology into practice creating a powerful surveillance system built for specific use cases. Researchers, designers, developers, and related professionals in the health surveillance community can build upon the principles and methodology described here to enhance and broaden current surveillance systems leading to improved situational awareness based on a robust integrated early warning system.

Author Keywords
Surveillance System; Design Theory; User Experience; Visualization

ACM Classification Keywords
H.5.m [Information interfaces and presentation (e.g., HCI)]: Miscellaneous

Introduction
Public health surveillance is the ongoing systematic collection, analysis, and interpretation of data, and is closely integrated with timely information dissemination to health professionals responsible for preventing and controlling disease and injury [2].

On April 17, 2013, a fire and explosion occurred at the West Fertilizer Company (WFC) in West, Texas. The blast caused 15 fatalities, more than 260 injured, completely destroyed the WFC facility and caused widespread damage to more than 150 offsite buildings [1]. Investigations were conducted by different agencies after this severe chemical incident took place. According to The Dallas Morning News [2], minimal information is released by government entities concerning the frequency and amount of serious chemical incidents in the United States. Big publicly available news stations are the most common outlet for event monitoring by chemical analysts. However, these top news sources provide a posteriori data and, therefore, lack the timeliness and initial incident details needed for forecasting, planning, and early warning systems.

According to Department of Homeland Security website [3], specific US government departments work to increase public and governmental awareness of potential chemical and biological threats and to strengthen the nation's response against these threats. During a chemical incident, analysts have to manually curate information compiled from multiple sources to create a report for the potential event. Sometimes this information is minimal and sometimes it is overwhelmingly large, regardless a large amount of time is needed to identify valid, important data, integrate the facts, and formulate a story. An optimized system that simplifies analysts' data collection, analyses, evaluation of the potential chemical hazard, and automates the creation of event databases for collaboration and reporting is essential to improve current situational awareness and efficiency in event reporting for active decision making and public awareness.

To address these gaps in current chemical surveillance efforts, we have designed and developed a software system called ChemVeillance. This real-time chemical surveillance system collects and integrates chemical-related official, local and national news, social media data, and reports from the National Response Center (NRC), aiming at increase the accuracy and timeliness of chemical surveillance analysts' identification and understanding of potential chemical hazardous events.

The objectives of this paper are: 1) The introduction of a novel system for analysts to easily track and analyze potential chemical-related incidents via publicly available sources (e.g., news and social media) as well as curated incident reports (e.g., NRC and FEMA) to increase situational awareness, assist in emergency response, and data driven decision making by addressing user experience to accommodate an analysts' workflow, from data collection and analytics to event reporting. 2) The application of design science research methodology to software architecture and development process within a case-specific context.

Related work
A great body of work has been done in developing surveillance systems, highlighted by Freifeild et al. [6] in an overview

[1] http://www.csb.gov/west-fertilizer-explosion-and-fire-/
[2] http://res.dallasnews.com/interactives/West/0825_datainaccuracies.html
[3] https://www.dhs.gov/

of the surveillance systems history. The very first generation of surveillance system is CDC WONDER, a Comprehensive On-Line Public Health Information System of the Centers for Disease Control and Prevention (CDC) [7]. The limitations with CDC WONDER, as described by Barthell et al. [3], are universal to most case-based surveillance systems: the data are limited to fatal injury statistics; data collection requires retrospective, labor-intensive, manual searches; the distribution of results may be delayed; and limited access to relevant local population-based data for local public health efforts.

The next generation systems to harness some of these resources is the Global Public Health Intelligence Network (GPHIN) [5]. GPHIN has shown that extensive monitoring and analysis of news media around the world can effectively aid in early detection of emerging disease threats [6]. Another successful online disease alerting service is the ProMED Mail email announcement list, with 38,000 subscribers and a panel of expert moderators [10, 11, 13].

More recently, the advent of news aggregators and visualization tools has spawned a new generation of disease-surveillance mashups (i.e., web application hybrids) that can mine, categorize, filter, and visualize online intelligence about epidemics in real time. A few examples are HealthMap [4], MedISys (Medical Information System) [1], EpiSPIDER (Semantic Processing and Integration of Distributed Electronic Resources) [8], and RODS (A Real-time Outbreak and Disease Surveillance system) [14], which all leverage informal electronic data for disease outbreak information [6].

Unlike a complete open sourcing and public facing approach, our work uniquely builds on resolving the needs of analysts by 1) collecting appropriate real-time data, e.g., the combination of chemical case data and publicly available data, and 2) presenting the information in a way that analysts can easily zoom out for a quick overview and zoom in to validate the individual data source and store curated historical data for future analysis.

Data and Methodology

In this section, we first explain the data collection approach. Then, we demonstrate our design process and design decisions by highlighting how we translated design science research methodology to our design phase.

Data Collection

Data to support the visualizations within ChemVeillance comes from three sources: highly curated News, Twitter, and the USCG National Response Center (NRC). The News feeds that support ChemVeillance are national and international news collected and filtered based on a set of curated keywords related to chemical surveillance. The Twitter feed is a random sample stream provided by Twitter. Similar to the News data, the Twitter data is filtered via the chemical curated keyword list. Finally, the NRC data includes both historic reports and daily emails. NRC incident report data is drawn from the NRC website [4], which lists all reported events since 1990., was downloaded and ingested into the system offline to support analytics. NRC email notifications contain information on all reported chemical-related incidents collected the day prior to the email.

Methodology

In the paper, "A Design Science Research Methodology for Information System Research," the authors articulate a methodology to conduct design science research in information systems [12]. The methods include six activities: 1) Identify the problem and motivation, 2) Define the objective for a solution, 3) Design and develop the solution, 4) Demonstrate the product, 5) Evaluate the results, and 6)

[4] http://nrc.uscg.mil

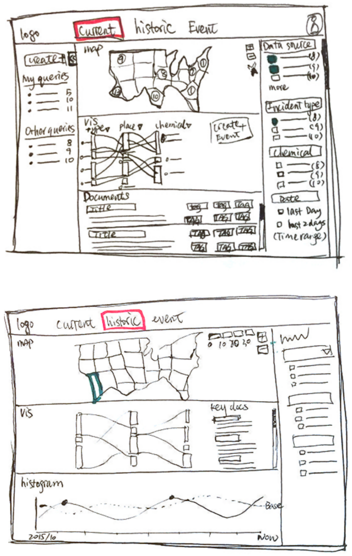

Figure 1: Drawings for ChemVeillance's early concept development.

Communicate the findings. Here, we discuss how we applied the first three activities in the design of our information system, ChemVeillance.

Activity 1. Problem identification and motivation. In this activity, we identify and frame the research questions. By justifying the motivation, we induce researchers to pursue the solution and attain consensus of the results. The main goal of ChemVeillance is to resolve the need of chemical surveillance analysts, increase situational awareness of a chemical hazard, and fill the gap in appropriate documentation for chemical incidents. To help inform our design, we aim to answer the following questions: 1) what are the responsibilities for a chemical analysts?, 2) what supports will analysts require to fulfill their responsibilities?, and 3) how can the tool benefit analysts daily work flow and have the greatest impact on current chemical surveillance technology?

Through a series of interviews with stakeholders, we arrived at the following conclusions: 1) Chemical analysts need to gather and view chemical-related evidences or information from various sources, identify chemical incidents, and create curated data-supported reports for documentation and future analyses; 2) Analysts are required to conduct those activities in one place and switch between two modes – exploratory (understand what's happening and what's new) and evaluation (dive deep to identify the situation); and 3) The system should provide a reporting trend analysis and provide a place to create and store curated data and reports.

Activity 2. Define the objectives for a solution. This activity highlights the success criteria of a solution and the ability to identify what is possible and feasible. Based on the problem identification from *Activity 1*, the objectives are for our solution are 1) increase the analysts' work efficiency, help them accomplish their daily task in one place, and 2) support analysts' difference use scenarios.

Activity 3. Design and development. Design and development is a process of iterating through multiple design constructs. For this project, we improved the design elements through several mock-ups: 1) Wire framing (Figure 1) – creating low fidelity design mock-ups by using simple symbols to present the contents, information architecture and simple work flow for concept development and discussion; 2) Higher fidelity wire frames – take the input from the first round, improve the feature with more detailed design elements; 3) High fidelity design mock-ups with color, it is polished look and feel to get aesthetic specific feedback.

Overview of ChemVeillance

ChemVeillance is a web-based application, built to help chemical surveillance analysts identify and evaluate hazardous chemical incidents from large amounts of curated data sources, e.g., news, tweets and NRC emails, in a timely fashion. ChemVeillance provides three main features and different views:

Interactive visualization: There are three different visualizations. The objectives of these visualizations are to summarize the data set from different perspectives to provide different insights and to enable data querying based on the user's need. The displays are interactive and provide cross-visualization display filtering. All the visualizations are fed by the selected harvested data sources.

Choropleth map (Figure 2, Left): The choropleth map [9] is the first view when users open the application. This visualization shows how many articles are reported in each state by the color encoded in the map. This feature helps the analyst quickly identify locations of concern and location filter. An important map feature is the ability to hover over a

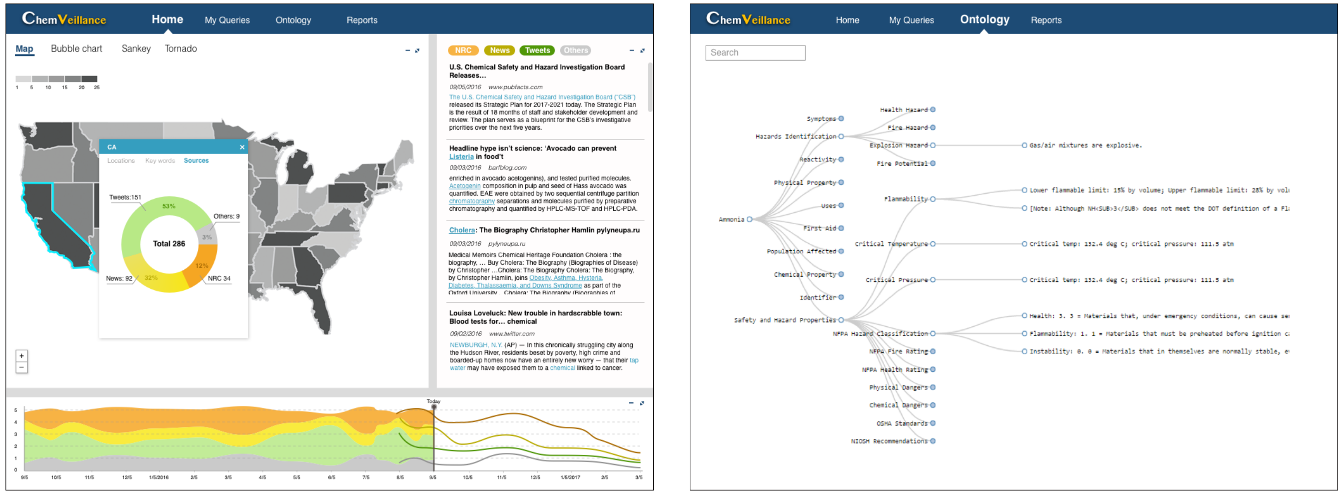

Figure 2: Overview of ChemVeillance. This program provides various types of information about the incidents (e.g., map, statistic, NRC, news, tweets, and trend; Left). Ontology of the chemical (Right).

state of interest and a small pop out window summarizing the filtered data is displayed. The pop-out window information contains the top 10 most reported articles in the state, keyword counts identified within the location-specific data, and the data sources feeding the summarized information. In addition, based on the analyst's map selection, all information in the other visualizations (bubble chart, Sankey diagram, histogram) and the document list will be updated. **Bubble chart** (Figure 3, Top):The purpose of the bubble chart is to provide a visualization that is used to quickly understand the gist of or topics contained within the filtered set of data. It reveals hidden themes and topic filter. The relative bubble size represents the keywords frequency within the data set; bubble color represents the keyword domain; proximity of bubbles represents the correlation of keywords. **Sankey diagram** (Figure 3, Bottom): The purpose of the Sankey diagram is to understand the relationship among data sources, location, and incident types and to help illuminate data noise. In this visualization, users can easily identify what data sources are feeding locations and incident types to understand the context of potential alerts. For instance, if tweets have a much higher volume than the other sources in a particular location and, therefore, are producing noise, they can easily be identified and filtered out in this view.

Saving queries – The saving queries feature refers to saving the past results and creating new queries to retrieve more data. When new data that meets the query criteria enters the system, the new data will go directly into the saved queried bucket. Analysts usually divide work by region and other elements. This feature allows analysts to have a quick

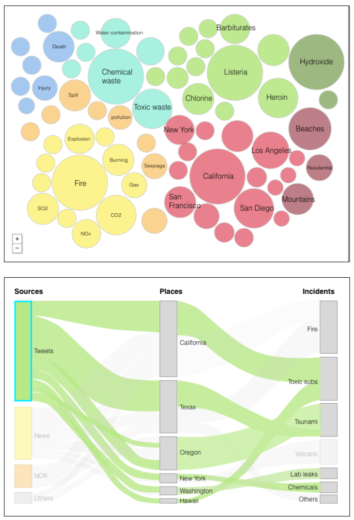

Figure 3: Bubble and Sankey visualizations used in ChemVeillance.

update and access to browse information pertaining to their specific job and interests, increasing their work efficiency.

Creating reports – The creating reports feature allows analysts to document important data sources from the article list in chronological order. This allows them to curate reports for management and/or the pubic and develop databases offline and further analysis.

Histogram – Histogram is a stacked area chart that keeps track of the past data volume and trend analysis for forecasting event severity. We encode data sources in different colors. With the histogram presentation, analysts can see the trends and detect abnormal volumes. Similar to visualizations, the histogram works as a display and filtering mechanism.

Article list – Article list (Figure 2, Left) is a list of articles collected from the data sources, which includes title, publish date, sources, and gist. In the gist, we highlight keywords to help analysts read through the content quickly. For chemical-related key words present in our ontology database, we create a hyperlink pointing the user to the keyword location in the Ontology page. The articles in this view are key for analysts to evaluate and identify chemical incidents. Populating the filterable article list with highlighted key terms pertaining to chemical surveillance and hazards provides an efficient way to skim through large amounts of articles in a short time.

Ontology map – The ontology map (Figure 2, Right) is a place where important chemical-related information for reference is stored and visualized. Analysts can get access to ontology map through the main navigation panel from the top, or by clicking a linked term in the article list. This chemical reference serves two different use scenarios: 1) An analyst is reading article that contains a word they want to learn more about; 2) An analyst is searching for a specific term.

Discussion and Conclusion

In this study, we presented the design of ChemVeillance, a chemical surveillance system for analysts, grounded by design science research methodology. Our system design and approach should inspire researchers, designers, developers, and related professionals in the biosurveillance community in how to design a surveillance system for power users. The visualizations and data abstractions provide an intuitive way for analysts to gain overall situational awareness of their domain and to investigate topical, spatial, and temporal anomalies.

We acknowledge that the paper does not contain a user evaluation on the design and use of ChemVeillance. We are engaging stakeholders and target users (analysts) to receive their feedback and guidelines. Once ChemVeillance is deployed, we will conduct a user study for evaluating the tool, which will provide us with suggestions and improvements for surveillance tool design. In addition, we are planning on performing passive observation and contextual interviews to see how well the tool is integrated in analysts' daily workflow. In the end, we will cover all six design activities for information system research [9]. We expect that an iterative design process will be applied throughout the development process of ChemVeillance.

Acknowledgements

This work was funded by the Defense Threat Reduction Agency under project number CB10190. Pacific Northwest National Laboratory is operated for the U.S. Department of Energy by Battelle under Contract.

References

[1] Wilson J. Argus. 2007. Use of Indications and Warnings for Global Tactical Detection and Tracking of Biological Events. *Georgetown Hosts 3rd Annual Conference on Infectious Disease* (2007).

[2] Thacker S. B. and Berkelman R. L. 1988. Public Health Surveillance in the United States. *Epidemiologic Reviews* (1988).

[3] Edward N. Barthell, William H. Cordell, John C. Moorhead, Jonathan Handler, Craig Feied, Mark S. Smith, Dennis G. Cochrane, Christopher W. Felton, and Michael A. Collins. 2002. The Frontlines of Medicine Project: A proposal for the standardized communication of emergency department data for public health uses including syndromic surveillance for biological and chemical terrorism. *Annals of Emergency Medicine* 39 (2002), 422–429.

[4] John S. Brownstein, Clark C. Freifeld, and Lawrence C. Madoff. 2009. Digital Disease Detection - Harnessing the Web for Public Health Surveillance. *New England Journal of Medicine* (2009).

[5] Mykhalovskiy E. and Weir L. 2006. The Global Public Health Intelligence Network and early warning outbreak detection: a Canadian contribution to global public health. *Canadian Journal of Public Health* (2006).

[6] Clark C. Freifeld, Kenneth D. Mandl, Ben Y. Reis, and John S. Brownstein. 2008. HealthMap: Global Infectious Disease Monitoring through Automated Classification and Visualization of Internet Media Reports. *Journal of the American Medical Informatics Association* 15 (2008).

[7] Andrew Friede, Joseph A. Reid, and Howard W. 1993. A discussion of rational and psychological decision-making theories and models: the search for a cultural-ethical decision-making model. *American Journal of Public Health* 83 (1993). Issue 9.

[8] Tolentino H. 2006. Scanning the Emerging Infectious Diseases Horizon-Visualizing ProMED Emails Using EpiSPIDER. *International Society for Disease Surveillance Annual Conference* (2006).

[9] Geoffrey M. Jacquez, Dunrie A. Greiling, and Andrew M. Kaufmann. 2005. Design and implementation of a Space-Time Intelligence System for disease surveillance. *Journal of Geographical Systems* 7 (2005), 7–23.

[10] Madoff LC. 2004. ProMED-mail: an early warning system for emerging diseases. *Clinical Infectious Diseases* (2004).

[11] Madoff LC. 2005. The internet and the global monitoring of emerging diseases: lessons from the first 10 years of ProMEDmail. *Archives of Medical Research* 36 (2005), 724–730.

[12] Ken Peffers, Tuure Tuunanen, Marcus A. Rothenberger, and Samir Chatterjee. 2007. A Design Science Research Methodology for Information Systems Research. *Journal of Management Information Systems* 24 (2007), 45–78.

[13] Morse SS, Rosenberg BH, and Woodall J. 1996. ProMED global monitoring of emerging diseases: design for a demonstration program. *Health Policy* (1996).

[14] Fu-Chiang Tsui, Jeremy U. Espino, Virginia M. Dato, Per H. Gesteland, Judith Hutman, and Michael M. Wagner. 2003. Technical Description of RODS: A Real-time Public Health Surveillance System. *Journal of the American Medical Informatics Association* 10 (2003), 399–408.