Epinome: A Visual-Analytics Workbench for Epidemiology Data

Yarden Livnat
University of Utah

Theresa-Marie Rhyne
Visualization Consultant

Matthew H. Samore
University of Utah

Effective detection of and response to infectious-disease outbreaks depend on the ability to capture and analyze information and on how public health officials respond to this information. Researchers have developed various surveillance systems to automate data collection, analysis, and alert generation,[1–3] yet the massive amount of collected data often leads to information overload.

To improve decision-making in outbreak detection and response, it's important to understand how outbreak investigators seek relevant information. Studying their information-search strategies can provide insight into their cognitive biases and heuristics. Identifying the presence of such biases will enable the development of tools that counterbalance them and help users develop alternative scenarios.

We implemented a large-scale high-fidelity simulation of scripted infectious-disease outbreaks to help us study public health practitioners' information-search strategies. We also developed Epinome, an integrated visual-analytics investigation system. Epinome caters to users' needs by providing a variety of investigation tools. It facilitates user studies by recording which tools they used, when, and how.[4] (See the video demonstration of Epinome at www.sci.utah.edu/gallery2/v/software/epinome.) Epinome provides a dynamic environment that seamlessly evolves and adapts to user tasks and needs. It introduces four user-interaction paradigms in public health:

- an evolving visual display,
- seamless integration between disparate views,
- loosely coordinated multiple views, and
- direct interaction with data items.

Using Epinome, users can replay simulation scenarios, investigate an unfolding outbreak using a variety of visualization tools, and steer the simulation by implementing different public health policies at predefined decision points. Epinome records user actions, such as tool selection, interactions with each tool, and policy changes, and stores them in a database for postanalysis. A psychology team can then use that information to study users' search strategies.

Simulating Infectious-Disease Outbreaks

Replicability is a key requirement for conducting psychological experiments—that is, a simulated outbreak must unfold exactly the same way on every run, provided the user made the same policy decisions. To compare different users' actions, the study must ensure that all users who followed the same path are presented with the same set of choices at the same times. Finally, for the study to have meaningful results, it must restrict the number of possible simulation outcomes.

To ensure replicability and restrict the experiments' problem size, and for training purposes, we employed a scripted, offline, large-scale simulation of *Bordetella pertussis* (the causative agent of whooping cough) activity in Utah that our collaborators at Virginia Tech's Virginia Bioinformatics Institute developed.[5] The simulation software includes a high-fidelity representation of social networks and contact patterns in computer-synthesized

Published by the IEEE Computer Society

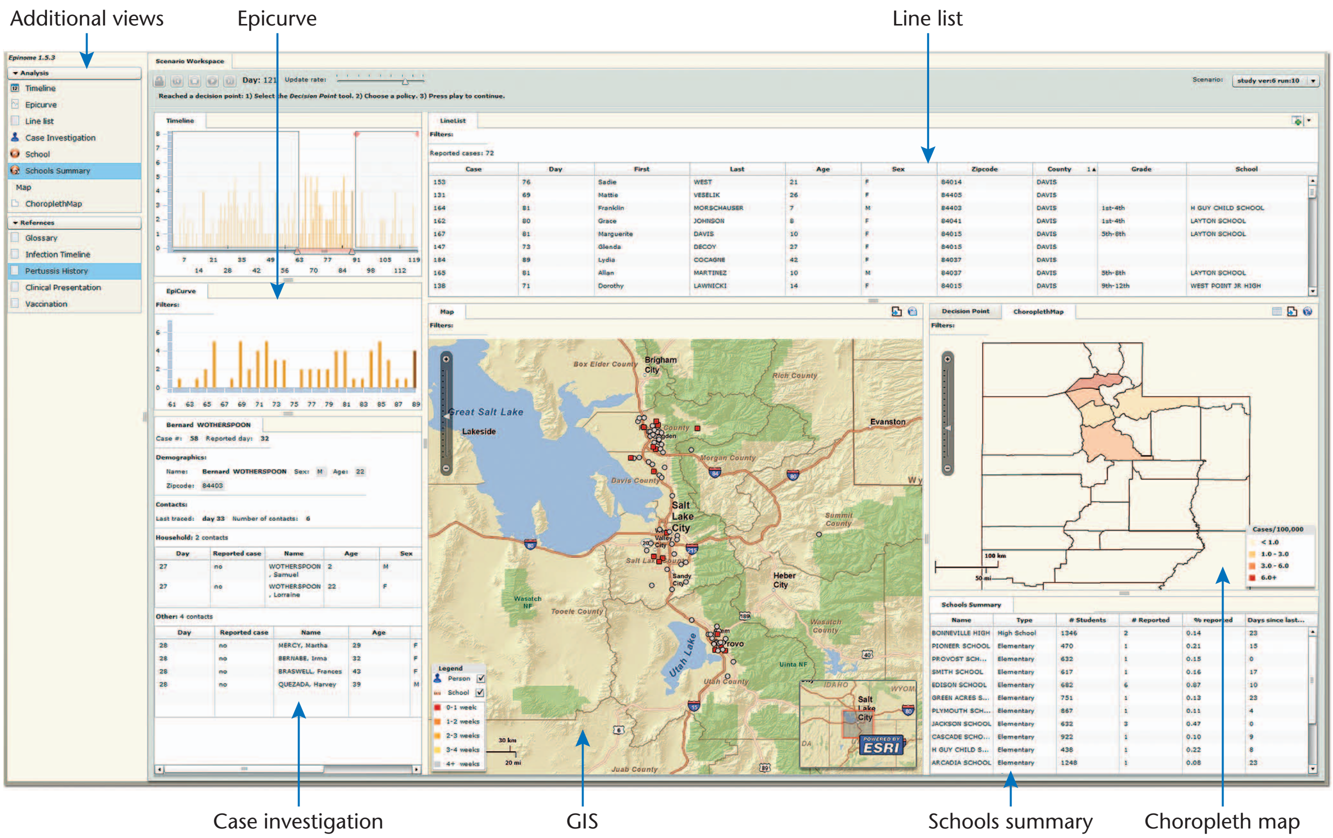

Figure 1. A snapshot of the Epinome visual-analytics system in use. Epinome includes features such as the line list, case investigations, a schools summary, and maps (such as geographic information system maps and choropleth maps).

populations, and an advanced stochastic model of pertussis transmission. The model incorporates vaccination, waning immunity, antibiotic-treatment effects, and other control measures (for example, social distancing), some of which users can variably apply.

We created realistic scripted scenarios with elements of randomness and unpredictability. The scripted simulations were based on analysis of historical numbers, published mathematical models, and input from experts in theoretical infectious disease and from in-field infectious-disease investigators. The overall simulation involved pertussis transmission across a high-fidelity simulation of 2.2 million people in Utah (using the 2000 US Census data) that allowed a more limited number of public health interventions.

We developed several outbreak scenarios with multiple decision points and a limited set of possible policy changes. Decision points included several possible policy changes and interventions—for example, starting contact tracing, changing vaccination policy for contacts, or closing schools. Contact tracing refers to identifying and diagnosing people who might have come in contact with an infected person. It's a time-consuming procedure that requires public health officials to interview an infected person and possibly additional people who might have come in contact with that person. We executed each scenario multiple times using different random seeds, through all possible paths. We saved the simulation output for each branch in each path, along with the simulation script, in a database.

When simulation playbacks reach a decision point, Epinome presents users with possible policy changes and actions based on the scenario description. Users can explore the current situation and use any of Epinome's visual-analytics capabilities before making the decision. After users choose a course of action, Epinome fetches the simulation output associated with the selected branch and resumes the playback.

Epinome

Figure 1 highlights some visual-analytics components of Epinome. The *line list* is a table view showing a collection of newly reported cases. The *case investigation* view lets users examine the detailed presentation of specific cases. Geographic information system (GIS) maps provide the geographic context of the outbreak zones under study. Choropleth maps depict a thematic display that shades areas in proportion to the number of cases reported per 100,000 people in each county. *Epicurves* depict a histogram of the number of newly reported cases

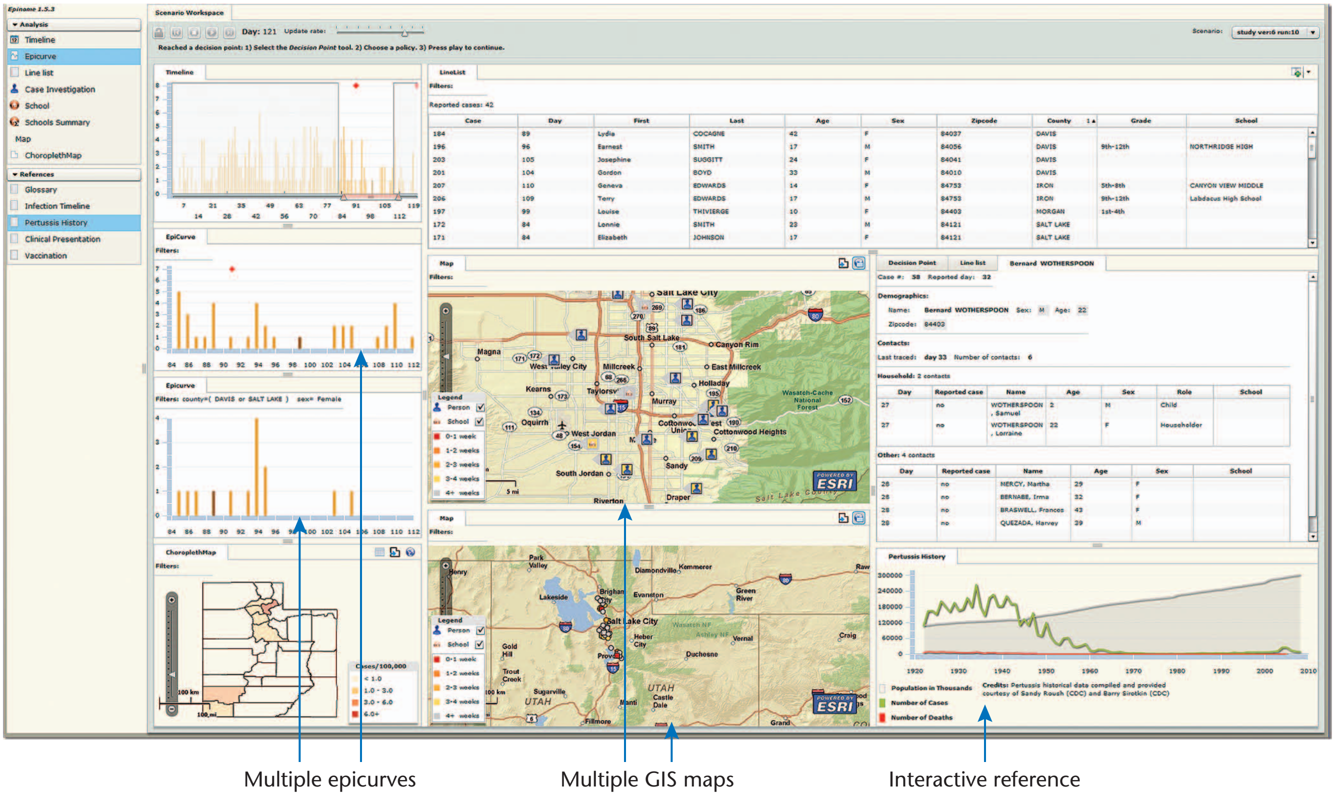

Figure 2. Epinome supports direct interaction with data and customized view creation to facilitate analytical reasoning. This figure shows the display at an earlier stage of the investigation as compared to Figure 1. Note that the user is using a different set of views and has repositioned the various views in the Epinome workspace. Introducing, removing, and rearranging views are simple drag-and-drop operations and don't require a special editing mode.

per day. These views provide the tools for visually aggregating and stratifying information associated with a disease outbreak.

The Epinome Workspace

Public health officials analyze large amounts of new information daily to support potential disease outbreak surveillance. Over the past decade, many surveillance systems have been built to automatically collect data, analyze it, and generate statistical alerts. These systems accomplish data analysis and presentation through time-series and geographical-information analysis with predefined visualizations. Typically, displays consist of a predetermined disparate collection of such graphics in a manner that might suit one particular task.

When we designed Epinome, we focused on interaction between public health practitioners and the surveillance system and on the human factors affecting information comprehension and decision making. To understand the human factors, we interviewed 50 public health personnel who work in communicable-disease control. We gathered information about their education, experience, roles, and responsibilities and the diseases they had investigated and monitored. Additionally, we inquired about their use of disease investigation protocols and guidelines to support public health decisions, responses, and actions. How the interviewees coped with the ambiguity of investigation protocols was of particular interest. We conducted contextual interviews in which we asked users to think aloud and explain their intended actions and decisions.

Loosely Coordinated Multiple Views

Using the interview sessions, we elucidated the requirements and design principles for Epinome's visual-analytics framework. The main requirement we identified was the need for an information-based[6] investigation environment that emphasizes exploration and information foraging to support both convergent and divergent thinking. In contrast, typical task-based systems emphasize predefined workflows for particular tasks.

To support dynamic searching and information foraging, Epinome empowers users to explore the data from various perspectives. It features a workspace in which users can interactively open new views that suit the current investigation, freely arrange them in the workspace, and remove views that no longer apply. Figure 2 depicts Epinome during an investigation session that illustrates the display's dynamic nature. We emphasized providing natural interaction so that users can focus on the task at hand rather than on interacting with the workspace. They can arrange views in the workspace side by side or in folders. The folder metaphor provides a simple way to temporarily move views out of the way but keep them close to other views with similar context. The folders are created and removed automatically and thus require no cognitive effort from the user.

Epinome incorporates and expands on the traditional coordinated multiple-view visualization paradigm. A coordinated multiple-view is supported through an event-driven framework that facilitates communication between disparate views. Views inform the underlying framework of important events, such as users brushing an area to highlight it. These events are routed to other views, which might act upon them or, more important, ignore them. We found that if all the views react to the highlighting of events, the resulting display flickers too much. To address this issue, we designed the system so that only views that show summaries, such as histograms, react to highlighted events.

During the contextual interviews, we noted that users usually filter data on the basis of values that are already visible on the screen, such as gender or age. We exploited this observation by providing a simple, intuitive query-by-example metaphor. Users can drag most visible data values (text or graphics) from any view and drop it into another view, into the workspace itself, or into the tabs area of a folder.

Dropping a value into another view—or even into the same one—automatically creates an appropriate filter. For example, dropping the value `Female` into another view creates the filter `gender = Female`. Similarly, dragging the graphical representation of a bar from a histogram will create a filter based on the date the bar represents. The receiving view uses the filter without having to be aware of its semantics. This approach enables views to filter data on the basis of data properties the views might not be aware of. For example, a GIS map view might only be aware of a reported case's geographic-location attributes. However, it could employ a filter that applies to the reported case's gender field.

Dropping a value into the workspace creates a filter at the workspace level—that is, a filter that applies to all the views. Finally, dropping a value into the tabs area of a folder will create a new view that's the most suitable for the data type in that folder. Epinome will then provide the value as a filter to the new view. For example, Epinome creates a new line list for a `Female` value with `gender = Female` as a filter, or a School view for a given `school id`.

This approach lets users interact directly with the data and apply filtering without needing cumbersome operations, such as navigating pull-down menus. To support this approach, Epinome views let users select almost any visible piece of information, whether it's a text field or a column in a bar chart. When an item is selected, the source view supplies the data's semantics in addition to the actual value so that a target view (or the workspace) will know how to use this value.

Epinome expands on the traditional coordinated multiple views by supporting filtering both at the view level and at the workspace level.

Epinome expands on the traditional coordinated multiple views by supporting filtering both at the view level and at the workspace level. In this way, global filtering can be used to coordinate the views—for example, focus on a geographic location or specific demographic group, while maintaining the ability to drill down in each view individually. We call this approach *loosely coordinated multiple views*, which empower users to explore different hypotheses in adjacent views yet still apply global filtering—such as specifying a time period or a geographic location—to all the views. We further expanded this approach by supporting multiple workspaces that disjoined investigations, such as using different simulation runs in parallel.

Finally, nested workspaces further support divergent thinking and facilitate exploring multiple hypotheses. We achieved this by deriving the workspace from a simple view. In this way, a child workspace behaves as any other view as far as the parent workspace is concerned. This enables users to create children workspaces, each with its own collection of views that focus on different hypotheses, yet share the same global filtering and overall focus.

Multiple instances of the same view can coexist—for example, multiple GIS maps, each zoomed on a different area, or multiple epicurves, each with its own set of filters (see Figure 2). The visualization system includes geospatial mapping using Esri (www.esri.com) mapping capabilities, choropleth mapping, movable timelines, and line-list querying. Epinome also provides other tools for

Age group	18–28	30–39	40–49	50–59	60+
	3	8	7	6	3

Job position	Director	Manager	Front line
	3	8	7

Experience	<2	2–3	4–9	10+
	2	4	8	13

Figure 3. Summary of study participants in the 2009 Epinome field study. The age groups, job positions, and experience levels show the depth and breadth of the subject population we used in our field study.

interactively aggregating and stratifying data, such as the school summaries.

The System Architecture

Epinome employs a client-server architecture. The client is written in Adobe Flex and runs in the user browser. The server is a Java servlet running on an Apache Tomcat server. As users interact with Epinome, it submits data requests to the server, which fetches the data from a database, prefilters it, and streams it back to the user client. From the user's viewpoint, there's no difference between a simulation playback and an online simulation. In fact, users needn't be aware that the simulation is a playback.

This approach has facilitated and simplified field studies for various state and local health departments throughout Utah and Colorado. This required no local installation of the client software, and user interactions were logged directly to our server in Utah.

The Field Test

We conducted a field test with 27 public health workers (14 men and 13 women that were selected from the staff directories of state and local health departments in Colorado and Utah. Three participants were between 18 and 29 years old, eight were between 30 and 39, seven were between 40 and 49, six were between 50 and 59, and three were 60 or over. Three were directors, eight were managers, and seven were frontline employees. All had outbreak investigation experience. Two had less than two years' experience, four had two to three years' experience, eight had four to nine years' experience, and 13 had 10 or more years' experience (see Figure 3).

The Scenarios

For our field test, we created 10 pertussis-outbreak scenarios. We gave each participant four random scenarios: one to train on and three to explore further. Each scenario had different initial conditions, such as the infection rate, predefined reported cases throughout Utah at various schools, and initial contact-tracing and prophylaxis policies. Epinome displayed the policies in its decision-point tool. Each scenario had some randomness regarding how the outbreak spread. All scenarios had two decision points: one at day 90 and one at day 120.

Once participants reached a decision point, the simulation playback paused, and we asked them to review the current situation and potentially change the contact-tracing and prophylaxis policies. Participants could use all of the Epinome tools for as long as needed. For our psychology study, we had to restrict the number of possible decision combinations, which we achieved by providing participants with only two decision points and three policy options at each decision point, resulting in a total of nine possible combinations.

Each option represented variations of the contact-tracing and prophylaxis policies:

1. Don't conduct contact tracing.
2. Access people who had close contact, but prophylax only high-risk contacts.
3. Access people who had close contact, and prophylax all of them.

The three options corresponded roughly to de-escalating the situation, maintaining the current policies, or escalating the preventive actions. Starting a contact tracing is time-consuming and expensive. Prophylax numerous people can put a large burden on a local health department and might have additional health consequences. (The policy restrictions and number of decision points were employed only for the field study and aren't integral to Epinome.)

In general, Epinome fetches information about the selected simulation at runtime. Such information includes a simulation run's description, the number and definitions of the decision points, and each decision point's policy options, which could be different for different decision points. Once the user selects a new policy, Epinome fetches the simulation tree's appropriate branch, and the playback continues.

Results

We aimed to examine what information participants looked for and how they used it to determine whether an outbreak had occurred. A secondary aim was for training purposes. The advantage of simulated outbreaks is that the ground truth—that is, the actual number of infected people—is precise. During a real-world outbreak, this information can only be roughly guessed.

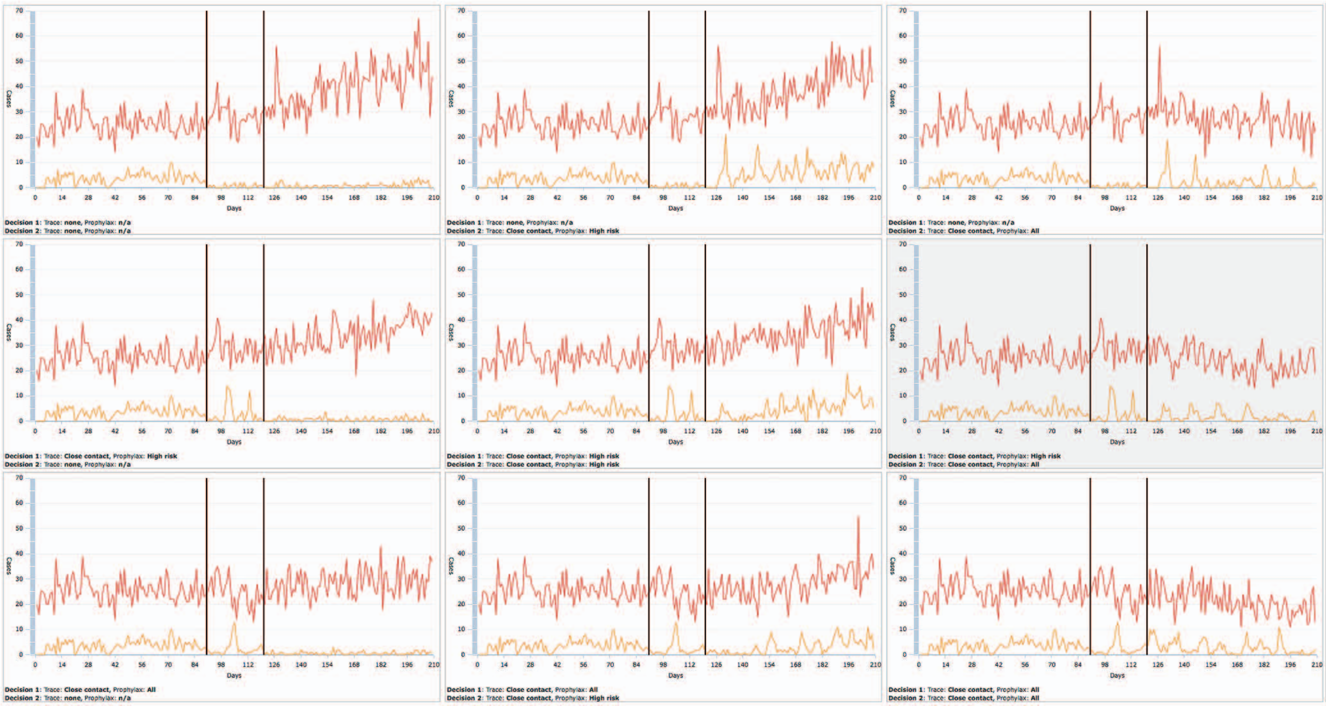

Figure 4. Displays of the ground truth matrix that public health participants could ask for at the end of their study sessions. The red lines show the actual number of people who got sick (the ground truth). The gold lines shows the number of reported cases. The two black vertical lines represent the days of the first and second decision points (90 and 120 in this scenario).

Table 1. Survey results from our field study.

	No. of responses (on a Likert scale)						
Possible benefit	**1 (disagree)**	**2**	**3**	**4**	**5**	**6**	**7 (agree)**
Provides new insights	0	0	0	1	4	6	14
Provides valuable decision support	0	0	0	0	1	10	14
Is highly usable	0	0	0	1	2	7	15
Would help do my job more efficiently	0	0	1	2	1	6	15
I would benefit	0	0	0	1	2	6	16

At the end of simulation runs, Epinome provided displays comparing the number of reported cases per day to the number of actual cases for each of the nine decision combinations (see Figure 4). For example, the last graph in the second row shows the results for choosing option 2 at decision point 1 and option 3 at decision point 2. In each graph, the orange lines show the number of newly reported cases for each day; the red lines show the number of actual cases. The two black vertical lines represent the first and second decision points, at days 90 and 120.

This simulation testing lets public health professionals gain practical insights to apply in an actual outbreak situation. For example, if the user selects option 1 at decision point 2 (see the left column in Figure 2), no contact tracing occurs. As a result, the number of reported cases decreases; however, the number of actual cases increases. This counterintuitive effect stems from the fact that stopping contact tracing reduces the reported information about the population, which means inaccurate information and not an improved population health. As a result, policy makers might determine the outbreak is subsiding, although it's actually growing out of control.

Analyzing the results with regard to user search strategies is beyond this article's scope. Table 1 shows the study results regarding Epinome's usability. Participants indicated that Epinome was a valuable decision-support tool and highly usable for visual analytics.

With Epinome, we focused on the direct and simple interaction between the user and the surveillance information as well as the human factors affecting information comprehension and decision-making. Public health professionals can explore, compare, and analyze geospatial and temporal data and can consult additional internal and external reference knowledge. This goes beyond the conventional information visualization approaches and lets users create what-if scenarios with Epinome's visual-analytics tools. We're now working with local and state public health departments in Utah and Denver using funds from the US Centers for Disease Control to deploy Epinome in their environments.

References

1. J.S. Lombardo, H. Burkom, and J. Pavlin, "Essence II and the Framework for Evaluating Syndromic Surveillance Systems," *Morbidity and Mortality Weekly Report*, vol. 53 supplement, 2004, pp. 159–165; www.cdc.gov/mmwr/preview/mmwrhtml/su5301a30.htm.
2. P.H. Gesteland et al., "Informing the Frontline about Common Respiratory Viral Epidemics," *Proc. Am. Medical Informatics Assoc. Ann. Symp.*, Am. Medical Informatics Assoc., 2007, pp. 274–278.
3. M.M. Wagner et al., "Syndrome and Outbreak Detection Using Chief-Complaint Data: Experience of the Real-Time Outbreak and Disease Surveillance Project," *Morbidity and Mortality Weekly Report*, vol. 53 supplement, 2004, pp. 28–31; www.cdc.gov/mmwr/preview/mmwrhtml/su5301a7.htm.
4. Y. Livnat et al., "Epinome—a Novel Workbench for Epidemic Investigation and Analysis of Search Strategies in Public Health Practice," *Proc. Am. Medical Informatics Assoc. Ann. Symp.*, Am. Medical Informatics Assoc., 2010, pp. 647–651; www.ncbi.nlm.nih.gov/pmc/articles/PMC3041367.
5. C. Barrett, S. Eubank, and M. Marathe, "Modeling and Simulation of Large Biological, Information, and Socio-technical Systems: An Interaction-Based Approach," *Interactive Computation: The New Paradigm*, D. Goldin et al., eds., Springer, 2005, pp. 353–394.
6. P. Pirolli, *Information Foraging Theory: Adaptive Interaction with Information*, Oxford Univ. Press, 2007.

Yarden Livnat *is a research computer scientist with the University of Utah's Scientific Computing and Imaging Institute. Contact him at yarden@sci.utah.edu.*

Theresa-Marie Rhyne *is a visualization consultant with the University of Utah's Scientific Computing and Imaging Institute. Contact her at theresamarierhyne@gmail.com.*

Matthew H. Samore *is a professor of epidemiology in the University of Utah's Department of Internal Medicine. Contact him at matthew.samore@hsc.utah.edu.*

Contact department editor Mike Potel at potel@wildcrest.com.