**SYDOVAT – Synthetic Syndromic Surveillance Data Creation Toolkit**

**Mission Need**
In order to effectively evaluate visual analytics techniques, standard test data sets must be created that require both high level and low level analysis where analysts find and communicate unexpected results while sifting through noise and other confounding factors. We have developed a novel system that allows users to generate non-aggregated synthetic data records from Emergency Departments using derived signal components from the Indiana Public Health Emergency Surveillance System. Our system synthesizes the daily, weekly and seasonal syndromic trends seen in Indiana Emergency Departments and allows users to inject outbreaks into the data, thereby creating a dataset in which analysts can be asked to solve a problem with a known solution, allowing for standard evaluations amongst various techniques. Data generated includes synthetic patient location and demographic information (age and gender) along with the Emergency Department chief complaint and chief complaint classification. Sample data sets are available for download. Please visit [http://www.purvac.org/SyntheticData](http://www.purvac.org/SyntheticData) for more details.

**Benefit:** SYDOVAT technology provides researchers with a range of synthetic data sets that are ideal for use in the evaluation of algorithms and methods dealing with multivariate, spatiotemporal data exploration, analysis and visualization.

**Collaborators:**
- Indiana State Department of Health (ISDH)
- The Regenstrief Institute
- Indiana University School of Medicine

**Funded by:**
- US Department of Homeland Security

Here a user has created a synthetic disease outbreak along the shore of Lake Michigan in mid-July. The linked epidemiological outbreak curve and age probability distribution function is shown on the right.

For more information, contact:
Dr. David S. Ebert, ebertd@ecn.purdue.edu
[http://www.purvac.org](http://www.purvac.org)