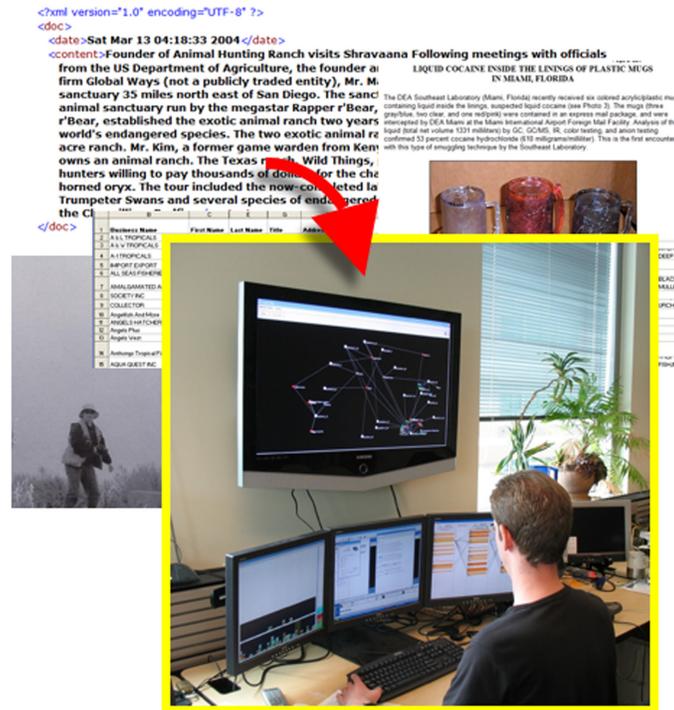


# National Visualization and Analytics Center™

## THREAT STREAM GENERATOR

### MISSION

Ever wonder if your technology will be able to detect a terrorism threat buried in a multi-source, composite data stream? The NVAC Threat Stream Generator (TSG) project team has developed a new approach to creating realistic test datasets and has created a series of datasets that are becoming increasingly popular with advanced commercial analytic tool developers, government information analysts and technology developers at universities. TSG datasets are being recognized as the test datasets of choice—with a set of freely available data—allowing users to determine how well their tools perform, because they know what the correct results should be ahead of time.



**Benefits:** TGS creates realistic datasets that are becoming the test datasets of choice by commercial tool developers, government analysts, and university technology developers.

### APPROACH

The NVAC TSG project's goals are to understand the characteristics of the data that analysts encounter and to create a tool that will build synthetic datasets mimicking real data to help analytic tool builders test and evaluate their creations. The distinct advantage of TSG datasets is that known ground truth in the form of a pre-determined threat, such as terrorist activity or a law enforcement concern, is created and translated into data cues. The cues are inserted into the dataset with a known expressivity—the number of cues and their subtlety of representation are controlled. The process is

Early Development

Lab Prototype

Open Data Distribution



September 2009

similar to writing a mystery novel. The task of finding the threat must intrigue and engage the person analyzing the

data, while inducing a suspension of disbelief, because they will know they are working with an invented scenario.



## IMPACT

The TSG team has had their data sets featured at the IEEE Symposium in Visual Analytics Sciences and Technology contest and challenge since its initiation in 2006. Since then, there have been over 600 downloads of these data sets from over 30 countries. At evaluation workshops at the symposium, tool builders and analysts expressed appreciation of this approach that allowed tools to be exercised and new insights gained in their understanding of tool performance.

## ABOUT PNNL

The Pacific Northwest National Laboratory, located in southeastern Washington State, is a U.S. Department of Energy Office of Science laboratory that solves complex problems in energy, national security and the environment, and advances scientific frontiers in the chemical, biological, materials, environmental and computational sciences. The Laboratory employs more than 4,200 staff members, has a \$918 million annual budget, and has been managed by Ohio-based Battelle since 1965.

### Contact

**Richard May**  
 National Visualization and Analytics  
 Center Director  
 Pacific Northwest National Laboratory  
 P.O. Box 999, MSIN J4-32  
 Richland, WA 99352  
 Phone: (509) 375-6976  
 richard.may@pnl.gov  
[nvac.pnl.gov](http://nvac.pnl.gov)

## Visualization and Analytics Centers



### VAC Consortium Members



For more information, contact  
**Mark A. Whiting**  
 Phone: (509) 375-2237  
[mark.a.whiting@pnl.gov](mailto:mark.a.whiting@pnl.gov)

Project Funding  
 U.S. Department of Homeland  
 Security



**Pacific Northwest**  
 NATIONAL LABORATORY

Proudly Operated by **Battelle** Since 1965