November 2007

# VAC Views

A publication of the Visualization and Analytics Centers

## Collaborating to Make a Difference

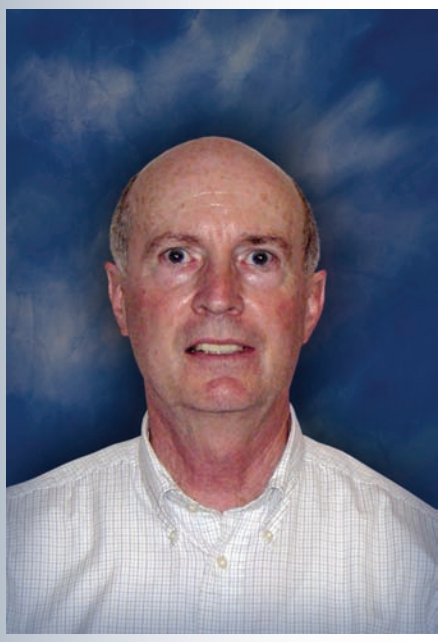
**Features:**

The **National Science Foundation** leads a collaborative program promoting data and visual analytics (page 4)

The **National Center for Foreign Animal and Zoonotic Diseases** explores visual analytics tools to mitigate the spread of disease (page 18)

Easy-to-use visual analytics tools help **law enforcement** discover crime patterns and create shared understanding (page 21)

# A Message from the Department of Homeland Security



*Dr. Hoyt has over 23 years' experience developing and managing technology projects related to homeland defense/law enforcement/counter-terrorism/counter-intelligence for the U.S. Department of Justice and DHS. He has also been a program manager at the Defense Advanced Research Projects Agency and the FBI, where he led a variety of programs to develop technical tools and methods to support both criminal investigations and foreign counter-intelligence operations. Dr. Hoyt holds a Ph.D. in computer engineering/information technology and an MS in electronics engineering specializing in communications systems.*

The mission of the U.S. Department of Homeland Security's Science and Technology Directorate is to develop technical capabilities to securely share, search and analyze homeland security information within and across jurisdictional boundaries. We seek to transition new technical capabilities in three years or less. This transition into operational use is both for DHS and for our partners in homeland security—including State, Local and Tribal government agencies as well as the private sector, which owns and operates over 80% of our nation's critical infrastructure.

For us, transition is all about how to bridge the gap between having a viable technology and moving this to a supported "product" that an end user can maintain and operate over its life cycle. The process followed varies depending on the user entity. An operational component of DHS (such as the U.S. Coast Guard, U.S. Customs and Border Protection, etc.) would traditionally migrate new technologies via a structured acquisition program (such as SBI*net* or the Integrated Deepwater System). However, transition to the State, Local, Tribal or private sector requires that a new technology become a viable commercial product that these users can buy.

Our current information-sharing environment consists of stove-piped communities that have developed their own policies, rules, standards, architectures and systems to channel information to meet mission requirements. Many gaps and seams exist in the sharing of information across these communities. Your technologies can provide key new capabilities needed to meet our most pressing issues in information fusion and visualization to support the Common Operating Picture; real-time field information access and understanding; and information sharing across organizational boundaries.

I'm excited about the research I have seen at the Consortium meetings. Some of these technologies are already in the transition process. I look forward to teaming with more of you in the future.

**Dr. John Hoyt, Director**

Knowledge Management and Information Sharing Section
Command, Control, and Interoperability Division
DHS Science and Technology Directorate

# Current Happenings

Around here, there is frost on the pumpkins and it is dark when I head to work and dark when I go home. Looks like 2007 is coming to an end. Here are some fabulous places, not to mention great conferences, to visit in 2008.

- **Visualization and Data Analysis 2008 (Part of Electronic Imaging 2008)**
  January 27–31, 2008, San Jose, California, USA
  http://electronicimaging.org

- **PacificVis 2008: IEEE Pacific Visualization Symposium**
  March 5-7, 2008, Kyoto, Japan
  http://vis.cs.ucdavis.edu/PacificVis08

- **CHI 2008**
  April 5-10, 2008, Florence, Italy
  http://chi2008.org

- **EuroVis 2008: Eurographics/IEEE Computer Society VGTC Symposium on Visualization**
  May 26-28, 2008, Eindhoven, The Netherlands
  Abstract submissions due Novermber 30, 2007
  http://www.eurovis.org

- **ICEIS 2008: International Conference on Enterprise Information Systems**
  June 12-16, 2008, Barcelona, Spain
  Paper submissions due November 30, 2007
  http://www.iceis.org

**Thank you, Jill!**

Many thanks to Jill Farris, who has kept the NVAC team on track in so many activities, including the publication of *VAC Views*. We wish Jill the best of luck in her future endeavers.

The special issue of *IEEE Computer Graphics and Applications* on visual analytics entitled "Discovering the Unexpected" came out in September 2007. This was guest edited by Kris Cook, Rae Earnshaw and John Stasko. If you don't have a copy, we strongly encourage you to get one.

For a more extensive list of events, visit our website at http://nvac.pnl.gov/events/.

**Richard May**
*VAC Views* Editor in Chief
nvac@pnl.gov

# Contents

4 Challenges and Viewpoints
- 4 National Science Foundation

6 VAC News
- 6 Spring VAC Consortium

8 Research Highlights
- 8 Geo-Visual Analytics
- 10 Global Terrorism Database
- 12 Electric Power Grids
- 14 WireVis
- 16 Assessment Wall
- 18 Foreign Animal and Zoonotic Diseases
- 20 Text/Data Integration
- 21 Law Enforcement

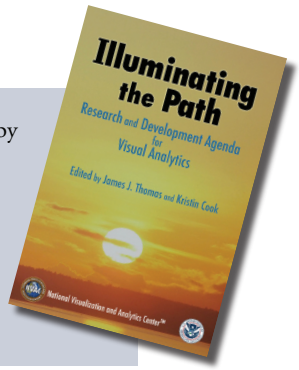
22 Partnerships

***VAC Views*** is published under the NVAC™ program with funding provided by the U.S. Department of Homeland Security.

For more information, visit the National Visualization and Analytics Center™ website at **http://nvac.pnl.gov**

This edition of *VAC Views* is also available at **http://nvac.pnl.gov/vacviews**

*Illuminating the Path: The Research and Development Agenda for Visual Analytics* is available for download at **http://nvac.pnl.gov/agenda.stm**




Pacific Northwest National Laboratory
Operated by Battelle for the U.S. Department of Energy

*P.O. Box 999, MSIN: K7-10*
*Richland, WA 99352*

The use of trade, firm, or corporation names in this publication is for the information and convenience of the reader. Such use does not constitute an official endorsement or approval by the National Visualization and Analytics Center, U.S. Department of Homeland Security, Pacific Northwest National Laboratory, or U.S. Department of Energy of any product or service to the exclusion of others that may be suitable.

# National Science Foundation Partnering to Advance Data and Visual Analytics

Larry Rosenblum, National Science Foundation

In July 2007, the Department of Homeland Security and the National Science Foundation announced that a partnership between NSF and DHS (represented by NVAC) had been created to conduct a joint research program in data and visual analytics. A formal Memo of Understanding was signed that established a 5-year plan for collaboration, dependent on available resources. The first-year funding is now in place.

This partnership between NSF and NVAC has been under development for several years. As the NSF point of contact, I have attended VAC Consortium meetings and served as an advisor to NVAC leadership. We are all excited about the opportunities this formal partnership will bring.

## The Problem

Those involved with science, engineering, commerce, health and national security all increasingly face the challenge of synthesizing information and deriving insight from massive, dynamic, ambiguous and possibly conflicting digital data. The goal of collecting and examining these data is not to merely acquire information but to derive increased understanding from it and to facilitate effective decision-making.

## Where We're Headed

To capitalize on the opportunities provided by these data sets, NSF is soliciting proposals from academia that capitalize on knowledge and expertise in the fields of mathematics, computational science and intelligent systems. The goal is to produce new data representations and transformations that will enable data stakeholders to detect the expected and discover the unexpected in massive data sets. This new program is called *Foundations of Data and Visual Analytics*, and FODAVA is the focus of the new NSFVAC. FODAVA is concerned only with a subset of the overall problem, namely the creation of the mathematical and computational sciences foundations required to transform data in ways that permit visual-based understanding.

Note that visualization is not specifically a FODAVA topic. Rather, FODAVA is concerned with the mathematical and computational models and transformations that will restructure data to provide future visualization systems with a superior ability to extract information using the new models and algorithms. FODAVA emphasizes fundamental research that will impact future-generation systems and also emphasizes novel, potentially high-impact ideas rather than advances on current algorithms and techniques.

## Two Collaborative Efforts

We are establishing two types of research efforts: FODAVA Lead and FODAVA Partnerships. The FODAVA Lead effort will be granted to a research team where all team members belong to a single academic

institution that will assume a leadership and coordination role. The FODAVA Lead will also play a key role in the development of FODAVA as a research field. In addition to forming the lead scientific research team, this institution will be responsible for assuring that results are disseminated to the FODAVA community, that effective liaison between FODAVA researchers and NVAC takes place, that testbed data sets are developed and disseminated and that the mathematics and computer science research communities become increasingly aware of the need for FODAVA-related research.

FODAVA Partnership efforts will be two-to-three-year fundamental research projects. These academic partners will actively participate with the FODAVA Lead institution in developing FODAVA as a field.

## Some Thoughts on NSFVAC from Jim Thomas

Jim Thomas, NVAC Director, says, "This partnership with NSF is the most important event since the creation of NVAC in March 2004. It brings to the front stage efforts by folks within DHS, NVAC and NSF to jointly fund the development of basic research in visual analytics supporting DHS applied mission needs.

"Bringing the concept of a program like FODAVA into reality is a long and hard row. It takes many, but without a champion, the concept will never become more than that. Our champion is Larry. His tireless commitment—through many joint meetings, many conference calls, drafting program plans, budget challenges and getting all the right players to agree—brings together the ingredients for a fine meal. Larry cooked it well. It also takes incredible patience as government processes are often slow. Also, without the strong support of Dr. Joe Kielman and other leaders within DHS, this concept would not be available for our research community.

"This program will also greatly help our education efforts. It will nearly double our university partnerships and increase the number of student interns and faculty visits. We will surely expand technology inventions and demonstrations. This partnership will help to better establish the next-generation talent base and will likely stimulate other programs supporting visual analytics for education as well. We are truly excited about working with Larry and the new researchers who will visit and collaborate with our existing research and development teams."

*This new program will create the mathematical and science foundations required to transform data to allow visual-based understanding.*

# Spring 2007 Consortium Meeting

*"The quality of speakers and presentations was excellent and the conference was great. I made a lot of contacts and see potential in the future."*

***– Max Reynolds, Indiana Intelligence Fusion Center***

Over 140 individuals attended the VAC Consortium Meeting May 14-15 in Bellevue, Washington. The meeting, titled *Visual Analytics for Regional Preparedness, Security, Response, and Health*, brought together industry, academia, other research institutions, end users and national and regional government agencies. Presenters and participants examined how visual analytic tools and techniques can be used to facilitate decision makers working on the complex problems associated with the huge quantities of multi-source, multimedia data housed at regional centers. NVAC took the opportunity to hear about the plans and progress being made both regionally and nationally toward this issue.

Consortium member Microsoft served as co-host, demonstrating the continued and growing commitment industry has towards supporting the visual analytics efforts being led by the Department of Homeland Security.

*Member demonstration session.*

## Meeting Highlights

Joe Kielman, DHS Research Futures Director, kicked off the meeting by welcoming the crowd and discussing "The Scale of Things to Come," which described the digital universe and the challenge to preserve the knowledge for rational thought.

Barbara Graff, Director of Emergency Management for the Seattle Police Department, gave the keynote address. She discussed the ongoing integration of emergency management including preparedness, mitigation, response and recovery and coordination across other emergency response agencies and community groups.

Steve Stein, Pacific Northwest National Laboratory, moderated a panel session focusing on analytical support to prepare for, respond to and recover from security and health issues. Six participants from a wide variety of fields, including the Indiana Intelligence Fusion Center and Automated Regional Justice Information System, engaged in the lively discussion.

NVAC/RVAC teams and consortium members demonstrated current technologies and products in a poster and demo session on the afternoon of day one.

The day concluded with a lovely dinner sponsored by Microsoft. John Hoyt and Joe Kielman of DHS were recognized for their continued support of the visual analytics movement.

On the second day, a research panel focused on information fusion analysis and dissemination. This panel was represented by university, national laboratories and industry. A DHS capstone panel titled "Partnering with DHS Educational Centers of Excellence" concluded the meeting.

## Member Privileges

Consortium members, including the RVACs and industrial partners, experienced the privileges of membership throughout the two-day event. Member representatives participated on discussion panels and showcased their technologies and products. Before the meeting, RVACs and representatives from DHS and the Centers of Excellence attended a technical exchange meeting. This informal gathering provided the opportunity

for individuals to discuss current projects and generate new ideas from different viewpoints.

Members were escorted on tours at both the newly opened U.S. Coast Guard Joint Harbor Operations Center and the Seattle Police Emergency Operations Center. At JHOC, members received an overview of Seattle Sector operations from Lieutenant Commander Steve Wheeler and had up-close interaction with personnel on the JHOC floor. Members saw the visual tools that JHOC staff currently use, learned about technology gaps and gained an understanding of how analytic tools can be incorporated into JHOC's work processes.

At the Seattle Police Emergency Operations Center, members learned how emergency operations are coordinated among local stakeholders and received demonstrations of the tools EOC personnel currently use. In particular, there is a need for mapping and visualization tools that can help a large team of representatives from diverse organizations stay up to date on developing situations.

*Members touring the Seattle Police Emergency Operations Center.*

The evening ended back at the magnificent Meydenbauer Center with a reception. Members used this networking time to meet key regional and federal government officials and share final thoughts.

To download presentations from the meeting, see http://nvac.pnl.gov/meeting_spring07.

For more information about the consortium, see http://nvac.pnl.gov/consortium.stm.

## Points of Contact

Jim Thomas, NVAC Director,
Pacific Northwest National Laboratory,
nvac.consortium@pnl.gov

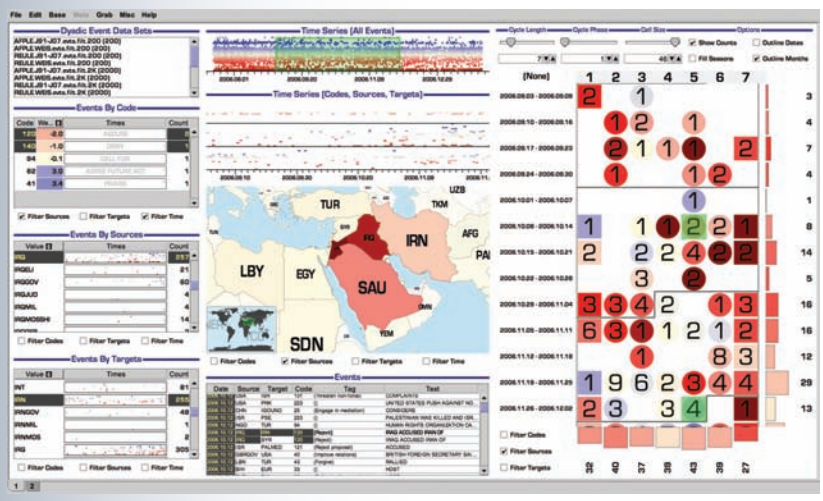
Mark Goodwin, VAC Consortium Director,
Pacific Northwest National Laboratory,
nvac.consortium@pnl.gov

*Members touring the newly opened Seattle Sector U.S. Coast Guard facility.*

# Geo-Visual Analytics to Leverage Distributed and Heterogeneous Information

Research and development underway at the North-East Visualization and Analytics Center targets advances in visual analytics that enable individuals and teams to leverage distributed and heterogeneous information, build knowledge and make decisions. Researchers at Penn State University and partner Drexel University are emphasizing methods and technologies that enable analysts and decision makers to understand and exploit the geographic ("where") and temporal ("when") elements of data—elements that are essential to anticipating, preventing and responding to major events. A key goal for the research program is to enable fast and effective national, state and local threat and vulnerability analysis, situation assessment and mitigation and disaster response and recovery.

## Visual Analytics to Turn Data into Information




*Improvise visualization of political event data generated automatically from the Kansas Event Data Systems News reports. Improvise allows intelligence analysts, policy advisors, journalists and social scientists to ask specific questions about groups of actors, kinds of events and patterns of behavior and also allows them to drill down into events.*

One goal of NEVAC research is to develop methods and tools for extracting useful information from complex, heterogeneous data of uncertain relevance. These efforts employ GeoVISTA *Studio,* the Improvise tool and the software environments (developed by NEVAC members) that support rapid development of custom geo/information visualization tools. One particularly important innovation is a *group, select and filter* analytic strategy that supports rapid visual construction and refinement of multi-part queries against large and complex data sets. This strategy enables analysts to sift and organize complex information in order to identify anomalies, find patterns and discover relationships.

Important data sources to support research, strategic assessment and decision making (whether in homeland security, business or science) exist in the form of unstructured text. While text is rich in content and typically easy for humans to interpret, machine interpretation of text remains a research challenge. To address part of this challenge, NEVAC developed FactXtractor, a tool that extracts important entities from text (e.g., locations, person and organization names, things, concepts and times). FactXtractor generates OWL files that can be ingested by a range of other tools including the Scalable Reasoning System being developed by PNNL's NVAC researchers (see p. 21) and the Context Discovery Application described below.

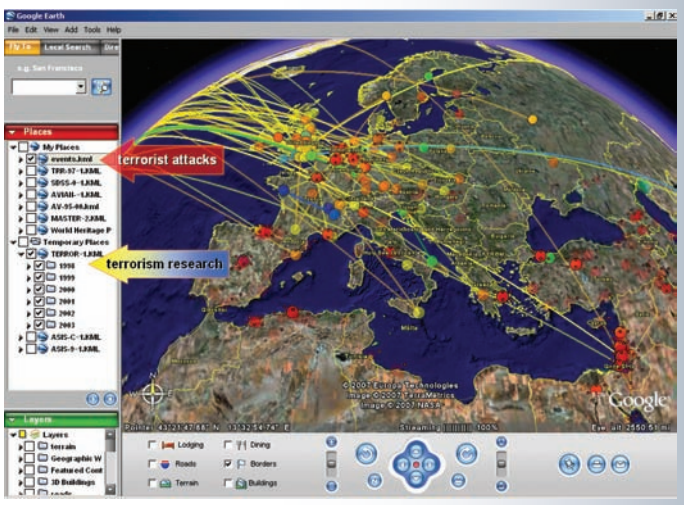
## Visual Analytics to Generate Domain-Relevant Actionable Knowledge

Disparate information fragments become actionable knowledge as connections across them are identified and new knowledge is grounded in existing knowledge. NEVAC researchers are developing, implementing and testing a range of visual analytics methods and tools to support this process.

One of these tools, the Context Discovery Application, builds on theoretical work in Local Model Semantics (a semantic model for reasoning with contexts) to support situation assessment and monitoring for disaster response events. More specifically, CDA integrates tools for knowledge-enabled information foraging from open source documents and for

space-time information contextualization of retrieved information. Specific tools connected in CDA include: geographically constrained retrieval of open source documents (e.g., news stories) based on a user-specified situation context; computational processing of retrieved documents using FactXtractors and tools for geocoding extracted place references; formal ontology integration to find potentially relevant non-spatial dimensions within data retrieved; and web-based mapping and analysis tools that support individual and group work with the derived information.

Closely related research focuses on tools for tracing the conceptual and geospatial diffusion of knowledge. In this work, methods to retrieve and extract entities, concepts and relations from text documents are being integrated with web map and web feature services to geographically connect and contextualize evolving information about important events and about related knowledge reported in scientific, news or informal (e.g., blogs) sources.




*Geographic distributions of terrorist incidents and scientific research networks for research about terrorism* (HCI International 2007). *The figure shows multiple themes simultaneously on the same map, including collaboration networks on three different subjects and terrorist incidents. Israel is marked by a dense cluster of incidents. Note the higher concentration of collaboration links in Europe. A few long lines across the globe indicate joint publications between Israeli researchers on terrorism and remote collaborators.*

NEVAC researchers are also extending ConceptVISTA (an ontology creation, management and visualization tool developed by team members and discussed in the February 2007 *VAC Views*) and connecting it with other tools (e.g., FactXtractor and Improvise) to support development and management of knowledge derived from heterogeneous sources and interpreted from multiple perspectives. NEVAC researchers have implemented methods to enable dynamic, coordinated linking of concept maps, text browsing and web search to help users quickly browse, locate and explore relevant information.

## Visual Analytics to Leverage Individual and Distributed Team Expertise

To reach their full potential, visual analytics technologies must support group as well as individual work. To address this need, NEVAC researchers are taking a comprehensive, user-centered design approach to developing collaborative visual analytics technologies. The goal is to both understand and enhance performance of teams addressing information analysis, sense making, policy formulation and decision-making tasks. The approach taken integrates field observation of practitioners and knowledge-elicitation from practitioners as well as simulation-based experiments. The latter use NeoCITIES, a small-world simulation environment, to study the impact of new geospatial analysis tools for synchronous and asynchronous different-place work on threat assessment and emergency response.

For more information about the research described here, see http://www.geovista.psu.edu/NEVAC.

*A web-map service interface developed to support distributed, asynchronous teams participating in emergency response activities. It has been incorporated with a NeoCITIES experiment to evaluate the impact of three geo-collaborative tools on situation awareness, distributed cognition and team performance.*

## Team Members

Penn State University: Mark Gahegan, Frank Hardisty, Junyan Lou, Alan M. MacEachren, Michael Mitra, Ian Turton, Chris Weaver

Drexel University: Chaomei Chen

## Point of Contact

Alan M. MacEachren, Director, GeoVISTA Center,
maceachren@psu.edu

# The Global Terrorism Database

DHS has as a primary mission to prevent terrorist attacks within the U.S. and reduce the vulnerability of the U.S. to terrorism. With support from DHS, a team of researchers at the National Consortium for the Study of Terrorism and Responses to Terrorism—or START—has developed a new tool designed to help analysts, practitioners and policy-makers achieve DHS's mission.

The Global Terrorism Database provides detailed information on terrorist events that have occurred all around the world since 1970, including all domestic cases (e.g., an American attacking a target within the U.S.) as well as international events (where the perpetrator attacks a target in a foreign country, as with the 9/11 attacks).

Compiled by a team of social scientists, the data has been assembled to allow for systematic analyses of trends in terrorist activity over time, across places and by different terrorist organizations. While the world may face unique terrorist threats today, analysts can learn a great deal about today's threats by exploring the behaviors of terrorist groups in the distant and recent past, and the impacts that these actors, and the events they perpetrated, have had. GTD is an essential tool to allow for such exploration.

*The Global Terrorism Database interface.*

## Approach

GTD began in 2001 when researchers at the University of Maryland obtained a large collection of data originally collected by the Pinkerton Global Intelligence Services. From 1970 to 1997, Pinkerton trained researchers to identify and record terrorism incidents from wire services, government reports and major international newspapers. With funding from the National Institute of Justice, the Maryland team finished computerizing the original Pinkerton data in December 2005, making corrections and adding information wherever possible.

With the 2005 launch of START by DHS, additional funds were made available to conduct validity and reliability checks of the existing data. In April 2006, the Center for Terrorism and Intelligence Studies, working directly with START, received additional funding from the Science and Technology Directorate of the DHS to extend GTD beyond 1997. By May 2007, data collection and coding was completed through 2004. During the next year, the new data will be systematically integrated with the original data to form a single source of information on terrorist attacks, from 1970 to 2007.

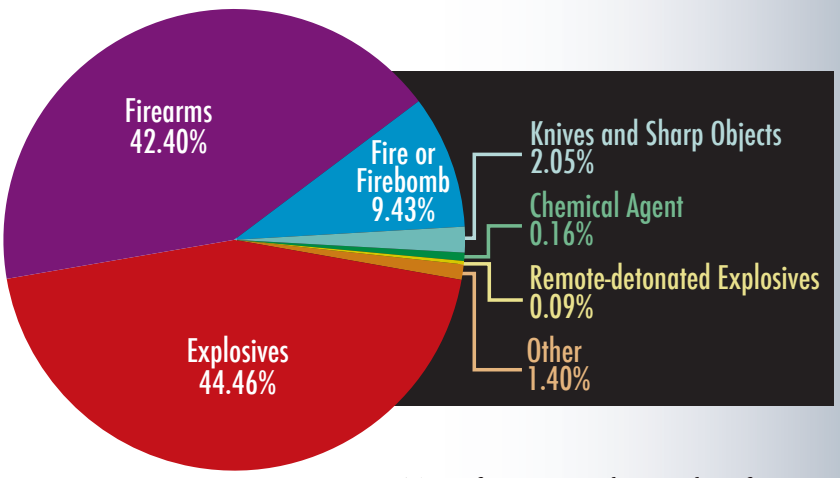
While data collection continues, START has already made access to data on events through 2004 available through a new online interface for GTD, available at http://www.start.umd.edu/data/gtd. This interface, available to the public, allows users to search the data for specific events, for all events in a selected country, for incidents linked to a specific terrorist group and more.

## Impacts and Outcomes

GTD contains information on over 80,000 international and domestic terrorist attacks between 1970 and 2004. GTD identifies more than 30,000 bombings, 13,400 assassinations and 3200 kidnappings. It also provides details on more than 1200 terrorist events within the U.S. since 1970.

For each attack, GTD includes information on:

- date of the event
- type of attack
- target and location, geo-coded to the city level
- weapons used
- number of casualties
- group claiming responsibility
- more than 50 other variables.




*Types of weapons used in incidents from 1970 – 1997, as seen in the Global Terrorism Database.*

The GTD team and START researchers have already begun conducting systematic analyses of the GTD data and have identified some myth-busting findings: Average age of a terrorist group is one year or less; 74 percent of all groups known to launch attacks between 1970 and 1997 lasted for 1 year or less; majority of attacks resulted in zero fatalities; 1.2 percent of cases resulted in the death of 25 or more people; and terrorists have been far more likely to use traditional weapons than exotic, advanced weaponry (see figure on types of weapons used in incidents from 1970 to 1997). Terror incidents began rising some in 1998, and that level remained relatively constant through 2004. In this way, GTD has already begun to provide empirical grounding for many of the debates surrounding discussions of terrorism and counter-terrorism.

For more information about START and GTD, see http://www.start.umd.edu/.

## Team Members

University of Maryland/START: Laura Dugan, Gary LaFree, Kathleen Smarick
Pinkerton Global Intelligence Service: Hugh Barber
National Consortium for the Study of Terrorism and Responses to Terrorism: Gary Ackerman
Center for Terrorism and Intelligence Studies: Charles Blair, Elle DiLornezo

## Point of Contact

Gary LaFree, Director, START, glafree@smart.edu

# Protecting the U.S. Electric Power Grids

Our society is vitally dependent on a network infrastructure of natural, man-made, and human resources—from food to water supplies, from electric power to other fuel sources, and from communication and transportation to medical and emergency services. While these resources are seamlessly integrated into the fabric of our society, electric power has the highest "network reachability"—and all other network resources depend on it. Losing electric power inevitably impairs the ability of the other resources to perform, which could cripple society if a widespread outage persisted for prolonged periods.

Soon after the August 2003 blackout in the Northeastern U.S. and Great Lakes areas, a scientific panel (Gilbert 2003) testified before the U.S. Congress and painted a grim scenario of the consequences of an extended blackout of our electric power systems:

"With the power out even a day or two, both food and water supplies would soon fail. Transportation systems would be at a standstill. Waste water could not be pumped away and so would become a health problem. In time natural gas pressure would decline…. Communications would be spotty or non-existent.... Business and production would [stop]. All in all our cities would not be very nice places to be.... It would not be a very safe place to be…. Martial law would likely follow…."

The vulnerability of our nation's electric power systems has also become the focus of research for a group of scientists and engineers at the Pacific Northwest National Laboratory since 2006. This article highlights some of the work.

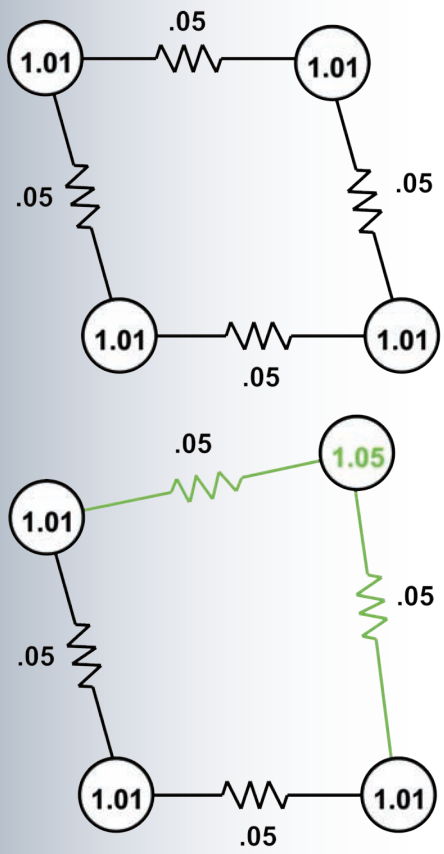
## Power Grids Visualization

The application of visualization has long been used by the electric power industry to facilitate power-grid operations. However, the main functionalities of existing visualization technology are mostly about depicting collected or processed information using color or icons on top of a geographic layout such as a U.S. map. This presentation-oriented approach, which can best be described as visualizing the *geography* of the grids, fails to take advantage of the analytical strengths naturally gained by the visualization itself.

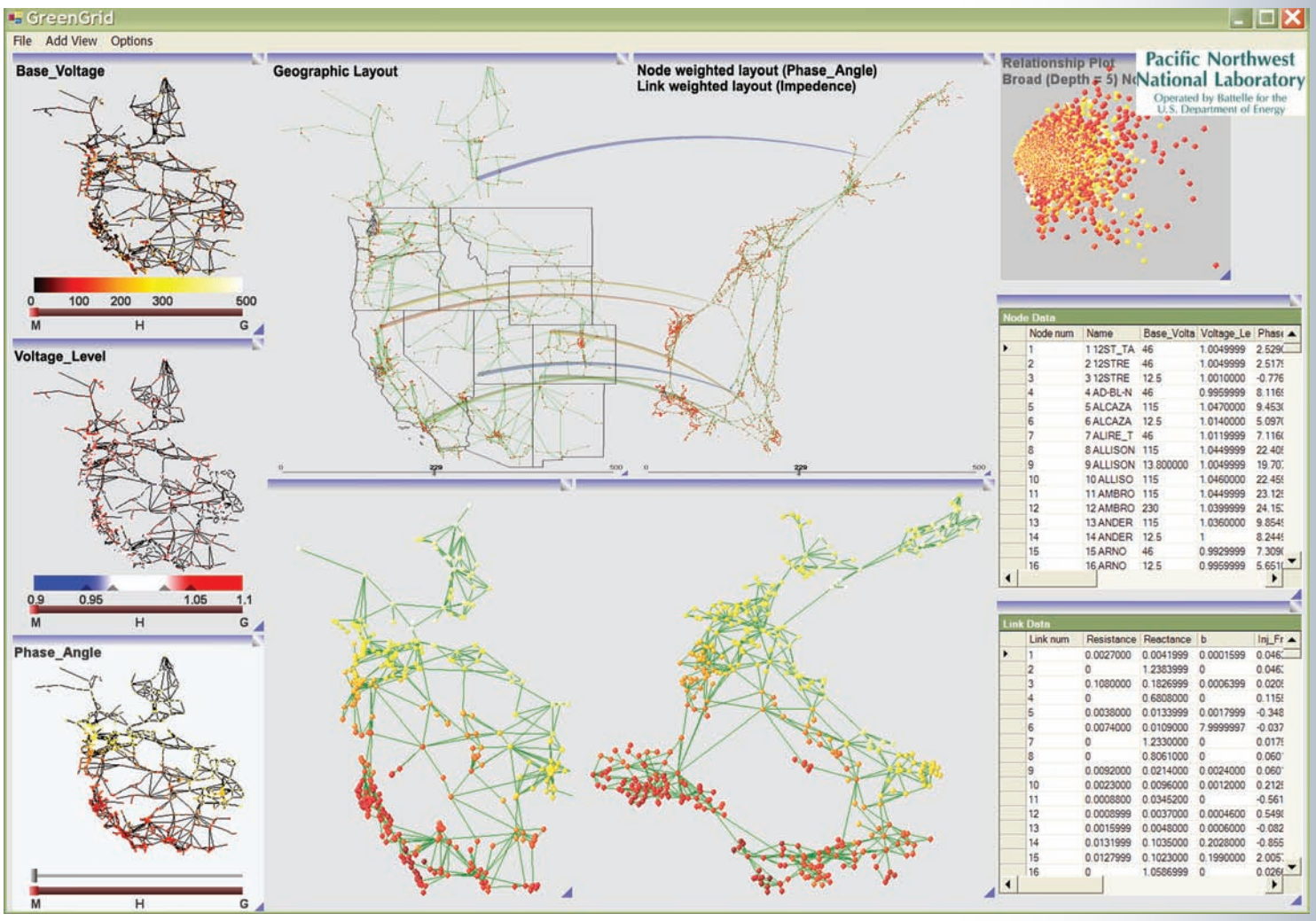
## Visualizing the Physics of the Power Grids

Instead of plotting the power data alongside a geographic layout, the PNNL team uses a weighted graph to simulate the attractive and repulsive forces of the grid operations. For example, all four nodes of the circuit graph in the figure to the left are weighted at the same per-unit value of voltage and all the lines are weighted at the same per-unit values of impedance. At the bottom of the same figure, the upper right node weighting has increased from 1.01 to 1.05 per unit. The result is a change in the way that the network is visualized. This visualization design reflects the true *physics* or *electronics* of the power grid operation. While the graph layout is distorted, the shape of the overall graph still preserves the relative locations of the graph nodes. Based on this visualization approach, the team has developed an interactive tool, known as GreenGrid, to support the planning and monitoring of the North American power grids.

One application of GreenGrid is to identify the vulnerabilities of the power grids at first sign of trouble by detecting the onset of a network separation. Failing to address the separation problem quickly could result in a full-blown blackout in the power grids. Data collected moments before the 1996 blackout of the western grids is used to demonstrate the team's design approach. In the figure on page 13, the left side shows the power grid system, 230kV and higher, geographically represented. The right side shows the impedance representation of the same physical system. In the impedance representation, the nodes are weighted using the voltage phase angle, and the links are weighted using the transmission line impedance values. The impedance representation clearly shows that the system nodes are clustered into




*Top: Same per unit values of impedance and per unit value of voltage generate a symmetric circuit graph. Bottom: Same per unit values of impedance but different per unit value of voltage (upper right one is higher) generate a distorted graph.*




*The left side shows the geographic layout of Western Electricity Coordinating Council, 230kV or higher. The right side shows a GreenGrid layout with additional weights applied to both nodes (using voltage phase angle) and links (using impedance).*

various groups. The weak coupling between strongly coupled areas indicates the lines where an unstable system is most likely to separate. This is seen in the figure where the points of separations of the August 10, 1996 blackout are highlighted. The points of separations are indicated by the brushing links between the two visualizations.

Green Grid is a component of *Have Green*, a suite of visual analytics technologies described in the July 2006 edition of *VAC Views*. GreenGrid is currently installed and undergoing evaluation in the Electricity Infrastructure Operations Center at the Pacific Northwest National Laboratory.

## References

Gilbert, PH (2003) "Implications of Power Blackouts for the Nation's Cybersecurity and Critical Infrastructure Protection: The Electrical Grid, Critical Interdependencies, Vulnerabilities and Readiness," Testimony of the first session of 108th U.S. Congress, Sept 4, 2003.

## Team Members

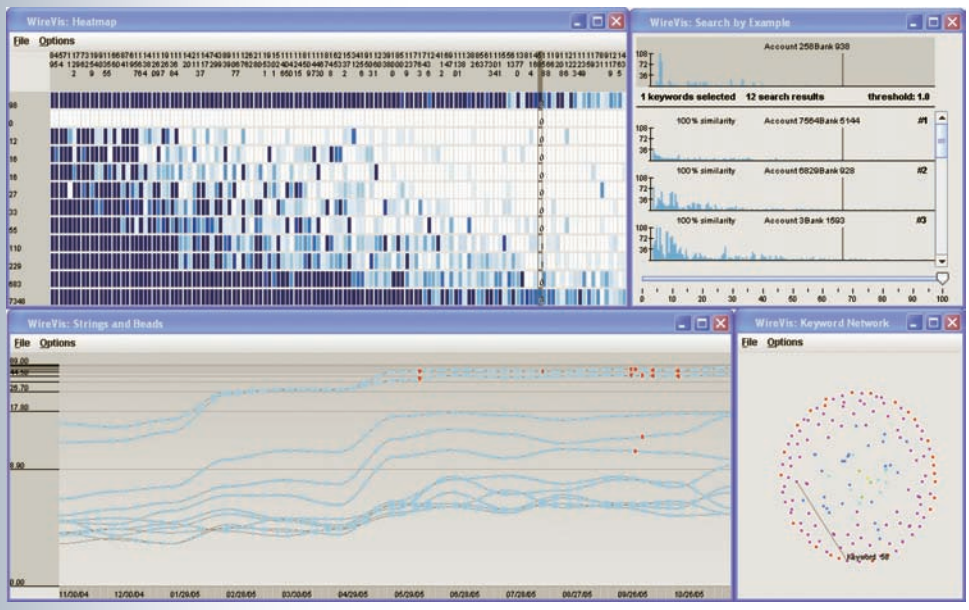
Pacific Northwest National Laboratory: Harlan Foote, Ross Guttromson, Patrick Mackey, Kevin Schneider, Jim Thomas, Pak Chung Wong

## Point of Contact

Pak Chung Wong, Chief Scientist, PNNL, pak.wong@pnl.gov

# Discovering Suspicious Financial Activity

Banks have the difficult problem of keeping watch over massive flows of transactions. When they see potential fraudulent activity, such as money laundering, they must investigate further, spending considerable time to winnow out the few cases that necessitate action. Federal law, in fact, requires that they do this. Banks face large fines or even closure if they do not carry out this responsibility effectively. Bill Fox, Senior Vice President and Senior Compliance Executive at Bank of America, likens this process to "looking for a needle in a very large stack of needles," indicating that transactions for fraudulent purposes look very much like regular transactions. This is a visual analytics problem requiring visual exploration, discovery and analysis. To attack it, members of the SouthEast RVAC developed a novel system called WireVis.




*A view of WireViz showing the heatmap (top left), search by example (top right), keyword graph (lower right) and strings and beads (lower left).*

## WireVis Tools and Capabilities

Fraudulent transactions involve deception and hidden or missing information about the activity and people involved. Perpetrators often change their patterns or have different patterns. Bank investigators' tasks are similar to other investigative analysts' tasks. This is especially true for investigation of wire transfers, where the people or companies involved may not be the bank's customers, the transaction path may be complicated and the bank may not have full information about this path.

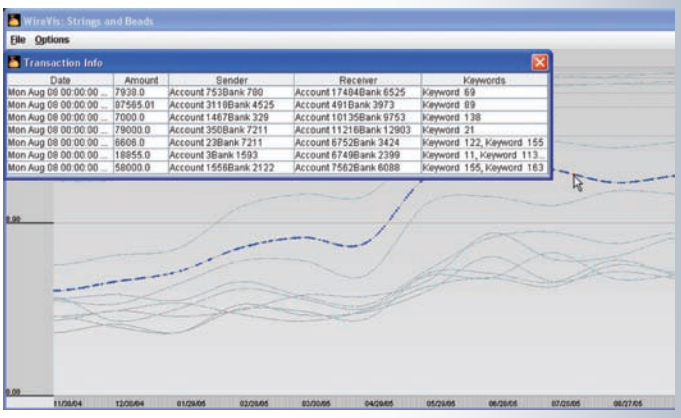
Analysts must monitor millions of wire transactions over a given period of time. Looking at financial patterns over time is not sufficient to identify suspicious activity. Analysts therefore have developed a proprietary set of keywords based on information in the transactions. They look for relations among keywords and also how keywords relate to transaction patterns. Experienced analysts have developed a "tradecraft" where they can quickly identify patterns that require further investigation. Current approaches have significant challenges when dealing with the amounts of data that must be explored.

SRVAC researchers decided to capture this approach in WireVis, a highly interactive system that would give the analysts overviews and detailed capabilities they never had before. WireVis has three main views: keywords versus clusters of transactions (heatmap), patterns of financial activity for the clusters over time (strings and beads) and a network map of relations among keywords. These views must be tightly coupled and balanced, so that users can launch an exploration from any of them and see the results of interactions immediately highlighted in the other views.

The heatmap displays keywords along the horizontal axis versus clusters along the vertical axis. A cell in the heatmap with darker, more saturated color indicates a high concentration of that keyword in the cluster. With a few clicks, users can drill down at any point and get to clusters that resolve individual accounts, even for large numbers of accounts. When users drill down, the strings and beads display is updated with the selected clusters; conversely, if users select a time period in strings and beads, the clustering and keyword distributions are updated in the heatmap view. When users slide

the cursor along the horizontal axis, the dates where that keyword appears in that cluster are shown as bright red beads on the strings. Similarly, when users slide the cursor along the vertical axis, keyword distributions for selected keywords (either selected individually or by rubber-banding) at that cluster are shown as bright red beads. Moving the cursor along either axis produces a sparkling animation where patterns can quickly be picked out. A year's worth of data at any level of clustering can easily be seen in this view.

WireVis also has a powerful and useful search-by-example tool. Instead of figuring out and then typing in a complicated query, users simply need to identify a pattern of interest in either keyword distribution or transaction activity (or both), and the system will find patterns that are similar to or not similar to the example pattern. WireVis lets users see quantitative distributions of keywords in a cell, see detailed information on individual accounts or groups of accounts, select different types of transaction patterns for the strings and beads view, and perform several other interactions.




*Selecting a specific bead brings up the related wire transactions in a separate window. The user can interactively drill down into individual transactions.*

## WireVis in Action

Researchers worked with Jim Price, head of the Bank of America wire transfer investigation group, to evaluate the WireVis's capabilities on a sanitized bank transaction data set. Using the heat map and beads and string overview, Price quickly separated normal transaction behavior (the great bulk) from behavior requiring more scrutiny. Price and the researchers focused on small sets of keywords, looking for peculiar temporal patterns in the heatmap view. One keyword showed transaction activity only in the second half of the year. Zooming in, Price noted that few transactions involved this word. Switching to the beads and string view, he found one transaction at $3M near year-end that stood out from the others, which were in the range of tens to hundreds of thousands of dollars. Double-clicking the bead for this transaction gave more details. Search-by-example showed that such a transaction is usually done by a large bank. Although not individually compelling, all these pieces of evidence taken together made a strong case for further investigation.

Bank of America officials are excited by the possibilities and want to see WireVis put into action. Bill Fox wants to see a transformation of bank investigators so that they are more like other investigative analysts and sees WireVis as a tool that can be used for both exploratory analysis and training. SRVAC researchers are completing work on a scalable version of WireVis that will efficiently handle the large amounts of real transactions in a year or more of bank data. Then they will put it to work. It is also clear that WireVis tools can be used on other financial analyses.

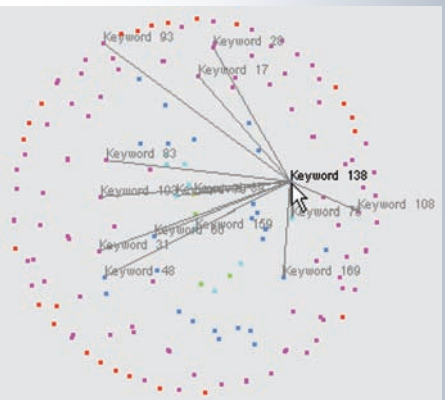
For more information about the research described here, see http://www.srvac.uncc.edu.

## Team Members

University of North Carolina at Charlotte: Remco Chang, Robert Kosara, William Ribarsky, Jing Yang

## Point of Contact

William Ribarsky, Bank of America Endowed Chair in Information Technology, UNC Charlotte, ribarsky@uncc.edu




*The most frequent keywords in the keyword network view appear in the middle, while the less frequent ones appear on the outskirts. When a user highlights a specific keyword, lines are drawn from that keyword to all related keywords.*

# Lightweight Visual Analytics: The Assessment Wall

Visual analytics research often produces powerful but complex interfaces to support information discovery and analysis. How can researchers simplify this discourse with information to make adoption easier and shorten the time to useful insights and productive analysis? The Assessment Wall is one such effort, designed to help users more easily monitor and explore continually updating text sources, such as news or reports, through a high-resolution touch-screen interface. The Assessment Wall team's goals are to help users find the latest on topics of interest along with emergent and unexpected themes. Additional interactivity enables finding relationships and patterns, all in context of the broader streaming data. Usability doesn't have to come at the expense of power and insight.

## Approach

*The Assessment Wall application on 45" LCD and touch overlay. Triage runs persistent queries in the upper left. ThemeView portrays emergent topics on the right. The TimeSlicer (bottom left) shows temporal patterns.*

The Assessment Wall consists of linked interactive visualizations for streaming text documents on a high-resolution touch-screen interface. The commercially available hardware components were chosen specifically to provide an attractive, engaging experience that leads naturally into intuitive touch interactions. Underlying the Assessment Wall is a mature visualization tool called IN-SPIRE™.

Whereas many visualization applications are expert tools, the team's research effort was devoted to adapting a subset of useful visual representations to empower natural interactions with immediately useful insights. Available tools and functions were reduced to a critical few, and most existing interactions were modified for easier use. The resulting interface provides easy-to-learn, high-value, engaging interactions for analyzing information.

The Assessment Wall occupies a singular position among large displays for work and ambient systems. In an ambient capacity it conveys important themes that change with the data, persistent query matches and patterns in time, any of which can be perceived at a glance. Unlike many ambient systems, it also supports interaction that helps to discover relationships between concepts and temporal patterns without training or special equipment. With this balance of ambient and interactive, novice users can maintain awareness of the data, dig deeper into topics of interest for simple analysis and use the system as a focal point for discussion and impromptu collaboration on work-related data.

As users' analytical interests change or when installed for new users, the interface is easily customizable to show different queries and geographical regions, broader or narrower time frames and even emphasis on topics in the emergent ThemeView visualization.

## Impacts and Outcomes

Periodic usability studies guided the initial design choices. The first installation at a large government client provided the opportunity for getting feedback and observing the system in use. Overall results were quite positive, and the team identified several key usability improvements and touch-optimizations that have since been integrated.

Several ongoing and proposed projects will use the Assessment Wall in operational environments to support DHS missions. The team will be customizing the visualizations to new data feeds and assessing user tasks and needs to provide the most value while preserving the ease of use and cross-functioning of tools. Operational use promises many

future opportunities to support effective display and analysis while maintaining a lightweight and usable interface.

DHS is also acquiring an Assessment Wall to place outside the undersecretary's office for use by DHS staff and external visitors. In addition to interactively exploring dynamic data, users will also be able to navigate a wealth of information on other DHS Science & Technology initiatives and their real-world impacts, Centers of Excellence, and RVACs.

*The Assessment Wall in use with live data at a meeting.*

## Research Directions

Moving forward, the Assessment Wall team is especially interested in empowering users with personalized options, supporting collaborative use and making the application more easily adaptable to new environments and tasks.

All users bring their own analytic context to their interpretation of the Assessment Wall. What is important to one user may be different from the next person. When individual users find a pattern or document of interest, the team would like to help them transition to their primary analysis environment.

Analytic teams have become a priority for DHS and the broader community. The Assessment Wall team wants to investigate the ways in which the Assessment Wall could assist in sharing analytic artifacts and supporting shared insight.

Finally, the team must be responsive to the diverse needs of each environment that integrates an Assessment Wall into its analytic practice to ensure its utility. While some types of customization are simple, the addition of whole new tools or integration with existing applications may lead to a new, more agile software architecture to support the broad range of DHS mission areas.

For more information about the research described here, see http://infoviz.pnl.gov/technologies.html.

## Team Members

Pacific Northwest National Laboratory: Barb Cheney, Nick Cramer, Richard May, Dave McColgin, Grant Nakamura, Ian Roberts, Randy Scarberry

## Point of Contact

Nick Cramer, Senior Research Scientist, PNNL, nick.cramer@pnl.gov

# Foreign Animal and Zoonotic Diseases:
## Boldly Visualizing Where We've Never Analyzed Before

*The FAZD Center works to understand, mitigate and prevent the spread of animal and animal-to-human diseases and is exploring the use of visual analytics tools to aid its efforts.*

How do you prevent a disease? How do you detect it? Does it spread? How do you contain it? How do you mitigate the impacts of a pandemic? How are animal diseases different from or the same as human diseases? What are the economic impacts of an animal disease outbreak? What about zoonotic diseases that can affect both animals and humans? What if an animal disease mutates into a human disease? What if a disease is purposely engineered or introduced as a terrorist weapon?

The National Center for Foreign Animal and Zoonotic Diseases is tasked to help answer these questions as a national university Center of Excellence for the Department of Homeland Security. The FAZD Center has begun working with NVAC and RVACs to explore using emerging visual analytics tools to better discover underlying relationships and present the essential concepts and findings regarding the spread and possible mitigation of animal and zoonotic diseases. Jim Thomas, NVAC Director, says, "This is an ideal match of needs, talents and technology. We are really looking forward to this growing partnership."

### The FAZD Center

The FAZD Center's purpose is to "defend the agriculture and food system against terrorist attacks, major disasters and other emergencies." The research and education mission of the Center is uniquely challenging in that it not only addresses the potential movement of pathogens and diseases in the complex infrastructure of American agriculture and food systems but also simultaneously addresses the potential impacts of animal diseases on public health. The Center's products protect America from biological attacks that use diseases that can be transmitted directly to humans or to humans through contact with animal populations. The Center is unique in its focus on the interface between animal and human diseases, recognizing that it will be impossible to prevent or successfully respond to catastrophic and pandemic diseases without a comprehensive understanding of the movement of diseases within animal populations and the environment and their interactions with human populations. The Center conducts research on high-priority diseases including avian influenza, rift valley fever and foot-and-mouth disease, with research aimed at producing improved diagnostics and vaccines.

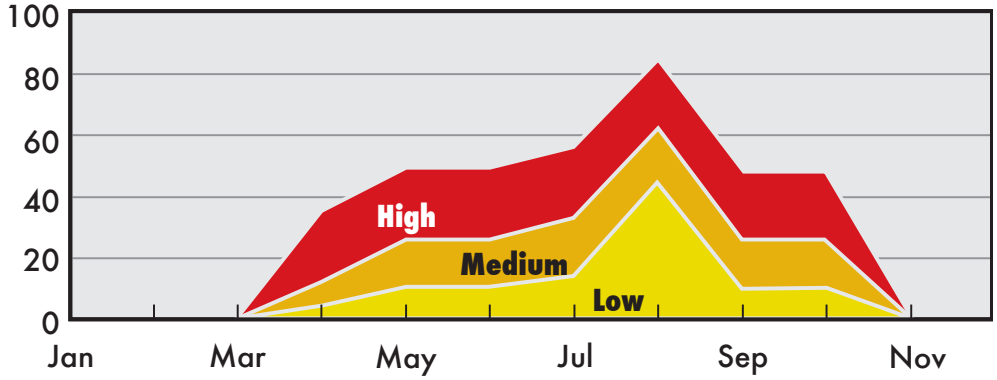
*FAZD researchers are currently using visualizations depicting temperature, moisture, presence of disease vector (e.g., mosquitoes) and presence of host (e.g., cattle) to help predict and analyze disease outbreak.* Graphic provided by SPADAC Inc. (http://www.spadac.com, info@spadac.com) copyright © 2007 all rights reserved.

The Center is developing and expanding on a number of disease models. These models build from an understanding of the basic biology of viral and bacterial pathogens, to the epidemiology of host-pathogen-environment interactions, to estimating potential consequences on animal and public health as well as potential economic consequences. Disease events occur in time and space. They start in a place or places and then spread and subside over time. Intervention can prevent, reduce or mitigate the disease and its consequences.

The FAZD Center relies heavily on Geographic Information Systems that map the spread of a disease or related factors in space; these maps can be shown in sequence or be animated to show how the presence of disease moves through space and time. While GIS is valuable and necessary to map out "where" and "when" diseases occur, the Center also needs new ways to make information more visually useful to carry FAZD content.

## The Need for VAC Tools

The FAZD Center is challenged to interpret large data sets and present emerging research findings so that security planners and crisis event managers can better prevent, contain and facilitate recovery from calamitous disease outbreaks. Disease event managers are now facing a virtual flood of data streams of varying degrees of usefulness. New tools are critically necessary for managing large data sets, and systems being researched by the VACs have new visual analytic tools needed by the FAZD Center. These tools make it possible to think in multiple dimensions and at a level of abstraction not previously possible for planners and event managers. However, to apply these tools within the context of responding to real-world events, there needs to be a subject matter expert conceptual framework imposed on available data sets and information streams, identifying data relations into a decision-making "space."




*Vulnerability windows are another technique to show risk over time. Relative risk levels—in this case low, medium and high—are used to show probability of disease outbreak.*

## The Next Steps

The next set of challenges is to begin to conceptualize how the vast information flows of the information space being developed by the VACs can be shaped and formed to address FAZD's interests and needs. This requires abstracting from FAZD space-time information the truly critical aspects that are under consideration. If researchers were approaching the boundaries of the "what is known" map for FAZD, then they will have entered the part of the map where the lines disappear and there is only the notation, "there be dragons here." Approaches for slaying those dragons will be presented in a future article.

For more information about the research described here, see http://fazd.tamu.edu.

## Team Members

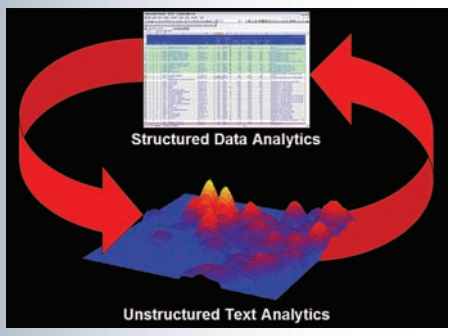
National Center for Foreign Animal and Zoonotic Diseases: Neville Clarke, Terry Nipp

## Point of Contact

Terry Nipp, University Outreach, FAZD Center, tlnipp@ilioco.com

# Text/Data Analytics Integration

The past few years have yielded tremendous advancements in visual analysis techniques. They have helped provide unprecedented insights into large document collections and enabled us to correlate vast quantities of structured data. Technologies that analyze document collections extract information from unstructured text by evaluating word content, key phrases or statistical properties of the collections. Structured data analysis techniques take advantage of the implied relationships within the data structures and other known dimensions of the information such as the numeric or categorical properties of the data. Often a focus on either documents or structured data is appropriate; however, many analysis activities span both types of information. The Text/Data Analytics Integration effort explores approaches that will leverage analysis techniques used for unstructured text analysis with those used in structured data analysis.




*The Text/Data Analytics Integration effort seeks to join two robust but contrasting analysis technologies to address combined structured and unstructured data analysis and visualization.*

## Making Allies of Text and Data Analysis Techniques

Under sponsorship of the DHS Science and Technology Support to Operational Components program, PNNL is working with private industry, NVAC researchers and the DHS/Immigrations and Customs Enforcement Office of Intelligence National Initiatives Unit to deploy an evaluation pilot. "We're surrounded by so much information ... we need to be able to consolidate it and make it ... understandable to average users," says Joe Garofalo, Acting Chief of the National Initiatives Unit. This pilot will help determine how two robust but contrasting analysis technologies can work together to address combined structured and unstructured data analysis and visualization. IN-SPIRE™, which is used for unstructured text analysis, and Tableau Software™, which supports visual analysis of structured data, will be deployed at ICE and applied to data sets consisting of both structured data and unstructured text. The project will investigate techniques that combine or integrate the analysis outcomes from these tools and evaluate their effectiveness. Many technical approaches are under consideration, but a key premise of this research is to apply the technologies in an environment as close to real operations as possible. As such, formal evaluation criteria will be developed in conjunction with the evaluation pilot.

## Capturing What Is Learned

The evaluation criteria will examine technical approaches from both an end-user, value-added perspective and an operational support standpoint. Utility of the technology is the critical factor in adoption by end users. The technology must be easy to use and add value with a minimum time or intellectual investment. "Our job is to take information and turn it into visual intelligence ... so that large volumes of data are simpler to understand," Garofalo says. A second key to infusion of the technology is integration with the organization's data sources and ensuring it can be supported by information technology staff. The evaluations will incorporate quantitative and subjective measures to determine how well technologies can be integrated into existing analytic flow, and if improved processes can be introduced into the ICE operational environment as a result of increased functional capabilities. The results of the assessment will be used to consider both technical and cultural process improvements to maximize benefit.

## Team Members

Pacific Northwest National Laboratory: Gus Calapristi, Jean Scholtz, Jim Thomas
DHS/ICE: Joe Garofalo

## Point of Contact

Gus Calapristi, Project Manager, PNNL, gus.calapristi@pnl.gov

# Visual Analytics for Law Enforcement

Law enforcement personnel must make rapid decisions in the field. They need to piece together chains of evidence, determine how crimes and suspects are associated and keep on top of developing information. Simple, easy-to-use visual analysis can help officers and crime analysts discover patterns of related crimes and create shared understanding.

The NVAC Scalable Reasoning System project is beginning a pilot deployment of its web-based and mobile visualization tools with the Automated Regional Justice Information System, a San Diego-area law enforcement information network. ARJIS connects local, state and federal personnel to criminal activity databases throughout California and the west. The deployment transitioning SRS from research to application is sponsored by the DHS Science and Technology Support to Operational Components program.

*SRS links mobile and office personnel with real-time information analysis and notification tools. ARJIS investigators can use web-based and PDA interfaces to find patterns in data and disseminate their findings to colleagues.*

## Web-Based Visualization Speeds Crime Analysis

Currently, ARJIS users query for data by supplying a name or other identifier. But when their searches return many results, it can be difficult to determine if a suspect has outstanding warrants or if there are attributes—like a common vehicle description—that connect multiple crimes. SRS addresses this problem. With SRS clustering tools deployed through the web-based interface to ARJIS, search results from many different data sources can be integrated and presented in visual groups that reveal their hidden structure.

In addition to data clustering, other SRS tools can be easily integrated into existing data warehouses. These include techniques for detecting temporal trends in text reports and street-level maps (using the NEVAC FactXtractor—see page 8) that show where crimes in a given cluster occurred. Designed to be easy to use, SRS is a good fit for ARJIS.

## Maintaining Situational Awareness in Distributed Teams

SRS also provides support for collaborative investigation. When users find something of interest, they can immediately share it with colleagues. This real-time sharing is particularly important to ARJIS, which has deployed 500 handheld computers to officers. Using these handhelds, officers can retrieve information such as warrants and booking photos in the field. But with mobile visualization through SRS, it becomes easier to quickly navigate to the right information and to make connections that would otherwise be missed.

SRS includes tools for capturing multimedia information in the field, automatically tagging it with GPS coordinates when available and disseminating it to team members. An officer might query for a suspect description, annotate it with a photo snapped on his handheld and with one click distribute the result to other members of a task force. SRS users can also set up customized "watch lists" that automatically push relevant information to them, regardless of whether they are using mobile or desktop interfaces.

The ARJIS pilot will serve as a model for how simple but powerful visual analysis can help investigation teams and mobile personnel work together.

*Through simple visual interfaces, NVAC's Scalable Reasoning System gives ARJIS investigators a readily understood picture of temporal, geographic and topical structure in their data. The same functionality is available on web browsers and PDAs (insets), allowing information collected in the field to be integrated with other data streams and quickly disseminated.*

## Team Members

Pacific Northwest National Laboratory: Bob Baddeley, Joe Bruce, Richard May, Bill Pike, Rick Riensche, Katarina Younkin

## Point of Contact

Bill Pike, Senior Research Scientist, PNNL, bill.pike@pnl.gov

## A Partnership with Academia, Industry and Government

## Consortium Members – Provider Level

BOEING
FUTURE POINT Systems, Inc.
Objectivity
GREENLINE SYSTEMS
Microsoft
GCS research
i2 A ChoicePoint® Company
TABLEAU SOFTWARE
Raytheon

## Program Representatives

**Joseph K. Kielman**
Department of Homeland Security

**John Hoyt**
Department of Homeland Security

**James J. Thomas**
National Visualization and Analytics Center

# Resources

**The following publications are related to visual analytics. This list is by no means inclusive, nor is it meant as an endorsement of those mentioned here versus those that are not.**


Committee on Basic and Applied Research Priorities in Geospatial Science for the National Geospatial-Intelligence Agency, Mapping Science Committee, National Research Council. 2006. ***Priorities for GEOINT Research at the National Geospatial-Intelligence Agency.*** The National Academies Press, Washington, D.C.
This report defines 12 problems in geospatial science that NGA must resolve to meet future needs.

Heuer R. 1999. ***Psychology of Intelligence Analysis.*** U.S. Government Printing Office, Washington, D.C. Available online at http://www.cia.gov/csi/books/19104/
Discusses how subconscious cognitive processes can limit our reasoning abilities and how we can try to understand and negate these effects.

Johnston R. 2005. ***Analytic Culture in the U.S. Intelligence Community.*** Central Intelligence Agency, Washington, D.C.
An inside look, from an anthropological perspective, at the analytical culture in the intelligence community.

Jones M. 1995. ***The Thinker's Toolkit: Fourteen Skills for Making Smarter Decisions in Business and in Life.*** Crown Business, New York.
Provides practical methods for simplifying problems and making faster, better decisions.

Lowenthal M. 2005. ***Intelligence: From Secrets to Policy.*** 3rd edition. CQ Press, Washington, D.C.
Describes how the intelligence community's history, structure, processes, and functions affect policy. This 3rd edition analyzes pre- and post-9/11 issues and events.

National Commission on Terrorist Attacks. 2004. ***The 9/11 Commission Report.*** WW Norton & Company, New York.
The report is a comprehensive examination of the events leading up to the 9/11 terrorist attacks. The commission traces the roots of al-Qaeda's strategies, describes how the 19 hijackers entered the United States and boarded airplanes, and details the missed opportunities of law enforcement officials.

Schwartz P. 1991. ***The Art of the Long View.*** Currency Doubleday, New York.
Outlines a scenario approach for planning. Schwartz describes new techniques, originally developed within Royal/Dutch Shell, based on many of his firsthand scenario exercises with the world's leading institutions and companies.

Thomas JJ, KA Cook (Eds). 2005. ***Illuminating the Path: The Research and Development Agenda for Visual Analytics.*** IEEE Computer Society Press, Los Alamitos, CA.
Describes the field of visual analytics and outlines the visual analytics science and technology research agenda.

2003
2004
2005