VACCINE ANNUAL REPORT – PERIOD 1

July 1, 2009 – March 31, 2010

Cooperative Agreement No. 2009-ST-061-CI0001

VACCINE
500 Central Drive, Room 127
West Lafayette, IN 47907

(765) 496-3747
vaccine@purdue.edu
www.VisualAnalytics-CCI.org

Table of Contents

I. Executive Summary

A. Overview

The Visual Analytics for Command, Control, and Interoperability Environments Center (VACCINE) was established on July 1, 2009 to co-lead, with Rutgers University, the Department of Homeland Security's (DHS) Center of Excellence in Command, Control and Interoperability. VACCINE's mission is dedicated to creating methods and tools to analyze and manage vast amounts of information for all mission areas of homeland security. VACCINE accomplishes its mission through an integrated program of research, education and outreach, spanning the disciplines of visualization and computer graphics, engineering, computer science, geographic information systems, cognitive psychology, information technology, and emergency management and public safety. VACCINE is an international center with overall management and the lead research component handled by Purdue University and the education lead directed by the Georgia Institute of Technology. The original VACCINE team consisted of 16 schools. Since July 1, 2009, the Justice Institute of British Columbia, Navajo Technical College, and Swansea University have been incorporated into the program. The VACCINE team is currently comprised of the following 19 universities with the associated Principal Investigator at each school:

University	***PI***
Purdue University	Dr. David Ebert
Georgia Institute of Technology	Dr. John Stasko
Pennsylvania State University	Dr. Alan MacEachren
Stanford University	Dr. Pat Hanrahan
University of North Carolina at Charlotte	Dr. William Ribarsky
University of Washington	Dr. Mark Haselkorn
Florida International University	Dr. Shu-Ching Chen
Indiana University	Dr. Maureen Biggers
Jackson State University	Dr. Loretta Moore
North Carolina A & T State University	Dr. Gerry Dozier
University of Texas at Austin	Dr. Kelly Gaither
Virginia Tech	Dr. Chris Barrett
University of Houston, Downtown	Dr. Richard Alo
Simon Fraser University, CA	Dr. Brian Fisher
University of British Columbia, CA	Dr. John Dill
University of Stuttgart, Germany	Dr. Tom Ertl
Justice Institute of British Columbia, CA	Mr. Jack McGee
Navajo Technical College	Prof. Mark Trebian
Swansea University, Wales	Dr. Min Chen

Figure 1

VACCINE is strategically positioned to support the Department of Homeland Security in confronting the challenges of safeguarding our nation in preventing, responding to, and recovering from events in applications including criminal investigation, health surveillance, fraud detection, emergency response, natural disasters, and border and infrastructure security. VACCINE focuses on the research, development, and deployment of interactive visual analytic environments for communicating and disseminating information and deriving insight from the massive homeland security data deluge. VACCINE is developing tools to help homeland security personnel, responders, and decision makers make sense of the sea of text, sensor, audio, and video data by developing powerful analytical tools and interactive visual decision making environments that enable quick, effective decisions as well as effective action and response based on available resources. VACCINE integrates data and analysis into interactive visual displays to enable users to make discoveries, decisions, and plan action. Applications include public health, animal health, public safety and emergency response. Turning massive data into actionable knowledge through visual analytic techniques is vital to the mission of the Command, Control, and Interoperability (CCI) Division of The Department of Homeland Security, as well as all of the mission areas of DHS.

VACCINE is addressing the following four of the five homeland security missions outlined in the Quadrennial Homeland Security Review:

Mission 1: Preventing Terrorism and Enhancing Security
Mission 2: Securing and Managing Our Borders
Mission 3: Safeguarding and Securing Cyberspace Mission
Mission 4: Ensuring Resiliency to Disasters

B. Global Accomplishments

Since the VACCINE Center officially began on July 1, 2009, these accomplishments reflect a nine month period of performance. Nevertheless, the VACCINE team worked diligently to create a robust portfolio of research and academic projects as well as make significant progress in our outreach activities. Highlights of our team's efforts include the data in Figure 3 as well as these activities:

- VACCINE reviewed and advised the Meadowlands Stadium Authority (NFL Giants & Jets) on security, networking, communications interoperability, and video surveillance plans for the new Meadowlands stadium.
- VACCINE formed a multi-agency public safety consortium in Indiana to develop tools for crime analysis, law enforcement planning, and decision making. We have embedded a graduate student within the law enforcement community to work directly with these agencies to understand their requirements.
- VACCINE has been working to create a unique corporate affiliate program in the form of a non-profit Limited Liability Corporation. Membership will be by invitation

only. Corporations will be complementary rather than competitive. Several companies are already in line to become founding members.

- We have formed a partnership with the Harrisburg, PA, Police Department to develop an extensible web-based map application that supports exploration of and sensemaking about criminal activity in space and time.
- VACCINE developed an interactive pandemic preparedness, surveillance, and training tool for the Indiana State Department of Health and adapted this tool for use by the State of Washington, New Jersey, and New York. This has been used for all health districts in Indiana and at least two county exercises in the state of Washington.
- VACCINE and CCICADA have been collaboratively working on four research projects and four education projects. This equates to 36% of the total VACCINE portfolio.
- We are developing law enforcement tools and evacuation planning tools for the Charlotte Mecklenburg Police Departments.
- We are developing an interactive intelligence visual analytics tool for use by law enforcement and intelligence analysts in the Atlanta area.
- The VACCINE team developed an interactive visual analytics dashboard with the FAZD center that was used in a DNDO nuclear detection full-scale exercise in Seattle in September 2009, working with USCG, Seattle Sector.
- We developed visual analytics course modules that were used in middle school and high schools in Indiana in 2009 and 2010.
- We are working with the START Center to develop visual analytics tools for the Global Terrorism Database and Minorities at Risk databases for the START Center.
- We are developing visual analytics flooding and preparedness tools for the DHS Natural Disasters, Coastal Infrastructure and Emergency Management Center of Excellence at the University of North Carolina.
- VACCINE has been researching novel visual analytics tools for Rift Valley Fever for the FAZD Center, incorporating epidemiological modeling as well as economic modeling and analysis.
- We have developed a network of Hispanic Serving Institutions and Tribal Colleges and Universities for engaging in visual analytics and network security research and education. We are in the process of incorporating the Navajo Technical College into VACCINE.
- We have developed cybersecurity visualization and analysis tools for US CERT.

- VACCINE is developing visual analytics tools for U.S. Customs and Borders Security to analyze their gigabytes of data.
- We have brought the Justice Institute of British Columbia on as a VACCINE School. Planning is underway to have a Visual Analytics for Public Safety workshop in September, 2010, in Vancouver, CA.
- VACCINE is finalizing plans for a joint United States-Germany research project to apply visual analytics to disaster prevention and crisis response, with a focus on critical infrastructures in logistics, transportation, food safety, digital networks and power grids at the national levels. This project is in collaboration with several international partners, NVAC, and NCFPD.
- We have expanded our network of universities to include three new schools: Justice Institute of British Columbia (Vancouver, BC, Canada), Navajo Technical College (Navajo Nation, Arizona), and Swansea University (Swansea, Wales, United Kingdom). A map of our current members and our partners can be seen below in Figure 2.

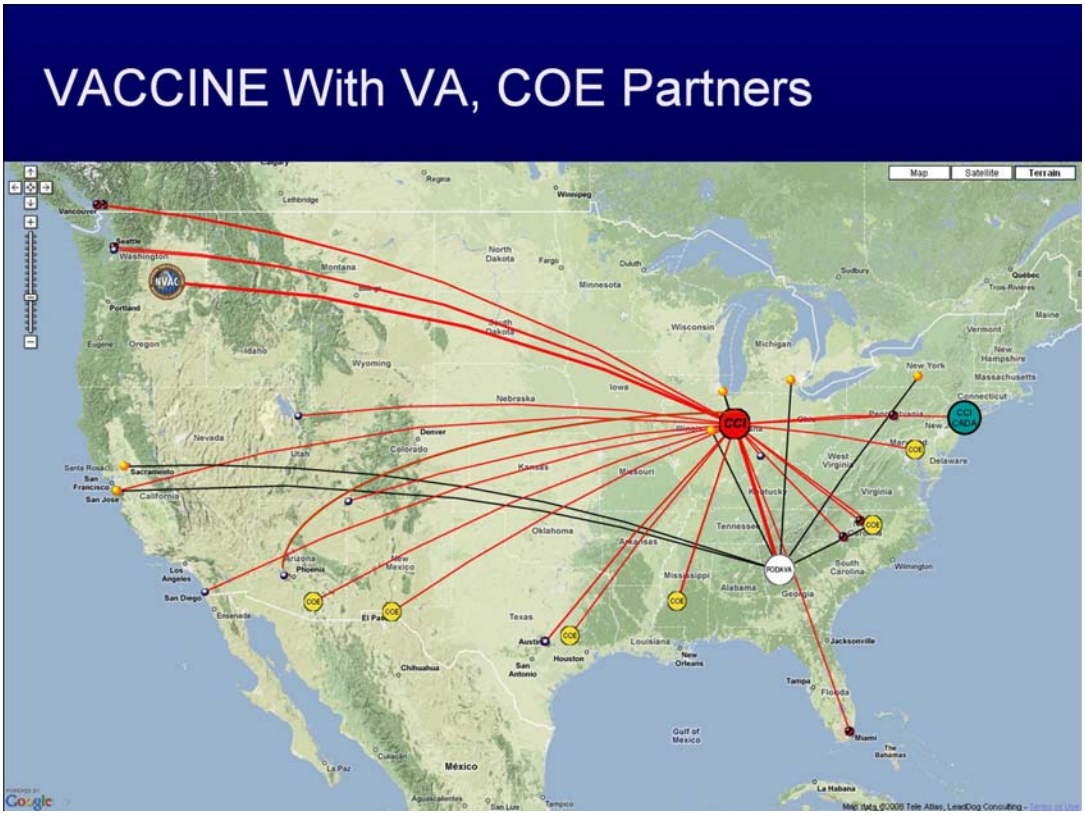


Figure 2: North American VACCINE team members and partners.

VACCINE's Accomplishments (July 1, 2009 – March 31, 2010)	
Mission Driven/Supplemental Research Projects Funded	15
Education Projects Funded	8
Seed Projects Funded	4
Peer-reviewed/Refereed Journal Articles	17
Conference Publications & Presentations	28
Other Reports	31
Scholarly Journal Citations of Published Reports	6
Student Theses	4
Patents	1
Graduate Students Participating in VACCINE Programs	**51**
Minority Students Participating in VACCINE Programs	15
Education Conference Participation	9
Education Presentations	5
HS-STEM Student Theses	3
HS-STEM Students Enrolled	11
K-12 Student Participation	120
Public Safety/Homeland Security Coalitions Established	5
State/Local Agency Memorandums of Agreement Signed	6
Leveraged Dollars	$21,557,914

Figure 3: Summary of VACCINE Accomplishments

II. Framework, Vision, Mission, and Strategy

The Visual Analytics for Command, Control and Interoperability Environments Centers' overarching goal is to establish itself as the preeminent asset to the myriad homeland security and public safety agencies who are our customers. In this context, we are developing novel tools to help the homeland security community make sense of and support decision making with the massive sea of text, geospatial, transaction, sensor, audio, video and all other data sources they have to deal with on a daily basis.

A. Framework

The VACCINE framework (Figure 4) is based on two key components:

1. Three foundational research areas that form our basis for addressing the seven DHS Command, Control and Interoperability mission directives:
   - Interactive Visual Analytic Foundations
   - Visually-Adapted Analytical Techniques
   - Investigative Analysis and Decision Making Environments

2. Mission-Driven Research, Education, and Seed Projects:
   - Integration and participation in important homeland security projects to provide real-world validation, testing, evaluation and a clear technology transition plan for our overall agenda.

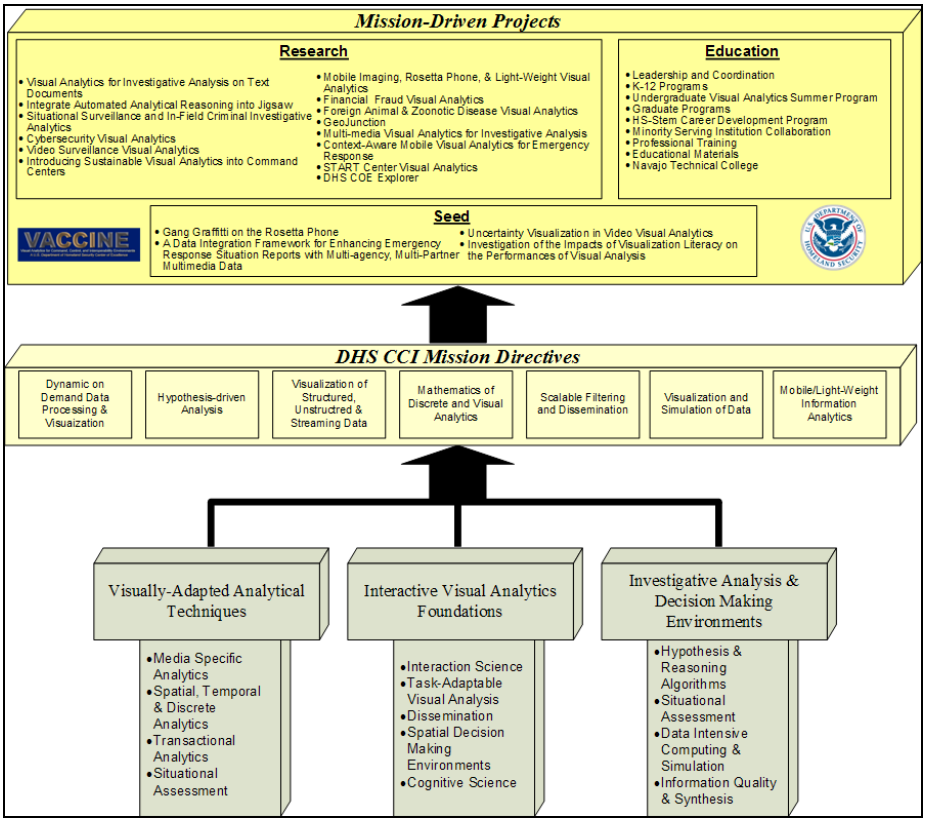


Figure 4: The VACCINE Framework

These two key components provide a tightly integrated short-term and long-term plan from theoretical foundations to validation and technology transfer to real-world DHS mission problems. Abstracting the key elements, generic problems, and common challenges across these projects provides focus and guarantees long-term success to our program. We are not just deriving ad-hoc solutions to problems, but addressing the fundamental research and academic questions, advancing the science of visual analytics, and creating a solid theoretical foundation for both future research and applications of that research to diverse homeland security challenges.

The Mission-Driven Research, Education, and Seed Projects are an evolving set of projects to explore and solve targeted, developing DHS mission needs. We have an annual internal seed project competition to foster the development of new ideas, including education-oriented and engagement-oriented projects. Successful seed projects will transition into our research or education portfolio and help evolve research and academic thrusts within each area as well. We also anticipate that our project portfolio will continue to expand as we provide solutions to other local, regional, and federal partnerships.

B. Vision

Using the best ideas from the top scientists, the VACCINE Center is conducting research that will advance the science of visualization and contribute to solving societal global challenges; educate students who are prepared to deliver; and actively engage and support our clients with a robust technology transition program.

C. Mission

Our mission is composed of the following four components:

1. Research, develop, and deploy interactive visual analytic environments for communicating and disseminating information and deriving insight from the massive homeland security data deluge of our digital world.
2. Provide visual analytic and scalable solutions for all 2.3 million extended homeland security personnel.
3. Achieve excellence in visual analytics and visualization sciences.
4. Educate homeland security stakeholders and the next generation of talent.

VACCINE Mission Foci

1. Innovation and Science Leadership – VACCINE is dedicated to researching, evaluating, and transitioning new methods and techniques for realtime, scalable, interactive visual-computational analysis, exploration, planning, hypothesis testing, management, and decision making from the data deluge faced by all homeland security personnel. We have assembled a team of the world's science leaders in visual analytics, media-specific analytics, geographic information science, information synthesis, data and knowledge management, scalable, data intensive computing, human-computer interaction, cognitive science, usability engineering, and communications to extend the science of visual analytics, as well as address the great challenges of CCI within homeland security.

2. Transition and Engagement – A key foundation of VACCINE's approach and success has been the engagement of partners to provide real-world grounding, impetus, validation, evaluation, and transition of our work. This engagement ensures the following two vital aspects of our work:
   - Cognizance of the variety of needs, issues, situations, and operational environments to ensure relevance of our activities and remove impediments for transition of technology, training, and education.
   - Clear mission need, feasibility, and technology transition plan addressing needs from the local to international level.

   We have local, state, federal, and international partners in Canada, Germany, the United Kingdom, Sweden, France, and the Netherlands. DHS funds are supporting our partners in Canada, Germany and the United Kingdom. We have formed several

Public Safety Coalitions constituting of multiple agencies. We are also working with corporate partners and national labs to transfer our technology from research demonstrations on real-world-scale problems to hardened, commercially deployable and viable applications.

3. Educational Innovation, Leadership, and Outreach – VACCINE is tasked and prepared to educate the next generation of homeland security professionals to harness the power of visual analytics and advanced computational tools to enable them to make effective decisions from the mass of multisource, multimedia data they will face in their careers. We are developing innovative educational techniques for visual analytics and decision making. We will continue to create the leading international repository and resource for visual analytics educational and training material, as well as practical lessons-learned/best practices CCI material. We are developing PK-12, undergraduate, and graduate visualization and analytic homeland security educational material and programs. We are leading technical education for visualization sciences integrated into the MSI network. Finally, we are working to develop and provide professional training and educational development materials and opportunities for homeland security professionals ranging from local emergency responders to state and federal officials.

Defining Characteristics

VACCINE is unique in its mission, research and approach to advancing the field of visual analytic and providing tools to solve important, world-scale problems. VACCINE is:

- Committed to our students and support their learning, both inside and outside the classroom.
- Focused on research that finds solutions to complex problems and provides insight into important societal issues.
- Engaged with our customers, both listening to their needs and anticipating and responding to emerging opportunities and challenges.
- Engaged in pursuing partnerships with a collaborative spirit and an understanding of where we have a distinct capacity to contribute.
- Dedicated to diversity, inclusiveness, and a global perspective.
- Dedicated to the highest standards of integrity, honesty, respect, shared responsibility, transparency, and mutual accountability.

D. Strategy

We have developed a center-wide strategy for success with measureable initiatives that is composed of components for each of our mission areas: research, management, education, and outreach.

Strategic Goal One - Research

VACCINE will serve society through creativity, expanded research, and state-of-the-art Innovations.

Initiatives

I. Create an innovative and effective research environment—frameworks, methods, and software—that advances the start-of-the-art in Visual Analytics. This will be accomplished through a careful balance of effort across our basic research and applied research agendas.

Actions

1) We will address the seven basic mission directives of the CCI Directorate by focusing on three foundational research pillars of visual analytics:
   - Interactive Visual Analysis and Decision Making Environments
   - Visually-Adapted Analytical Techniques
   - Information Integration and Management

2) We will apply a translational approach to our research methodology. In this manner, knowledge generated through scientific studies will mapped back onto the applications and evaluated in the field where it can impact the customer.

3) We will address our customers' needs through Mission-Driven Research Projects (MDRPs). The MDRPs will be an evolving set of projects to explore and solve targeted, developing DHS mission needs.

4) We will establish working advisory groups in "topic" communities (i.e., fire, law enforcement, emergency management) to enhance visibility and to secure valued counsel. These groups will provide real-world validation, testing, evaluation and a clear technology transition plan for our research agenda**.**

5) We will have an internal seed project competition to foster the development of new MDRPs, including education-oriented and engagement-oriented MDRPs. Successful seed projects will transition into our MDRP portfolio and evolve research thrusts within each area as well.

6) We will provide intellectual leadership for the field, publishing high-quality research in peer-reviewed conferences and journals, and speaking at research-focused international conferences, programs, and workshops.

II. Increase annual research expenditures on a consistent basis.

Actions

1) We will form strategic cross-cutting partnerships at the macro-scale including multi-disciplinary, multi-campus, DHS Centers of Excellence, international universities, and government and industry partners that will leverage existing funds and bring in new funding for high value research projects.

2) We will continuously surveil for additional visual analytics funding through US government funding sources (e.g., NSF, NIH, DARPA, DTRA, NGA), as well as corporate and international funding opportunities.

3) We will develop and promote an annual internal seed grant program for collaborative research that will bring on board new, high-impact faculty into the VACCINE team and offer significant potential for breakthrough science.

4) We will develop and nurture a corporate affiliate program where members can collaborate with VACCINE faculty and students through contractual arrangements or gifts.

Strategic Goal Two – Education

VACCINE will expand the pipeline of people with knowledge of the visualization sciences. Special emphasis will be placed on creating more opportunities for students to learn about visualization.

Initiatives

I. Expand learning opportunities to more students, faculty, and researchers throughout all educational programs. Programs will focus on specific areas of need, access to educational resources, certificate programs, and mentor pathways.

Actions

1) K-12
   - Develop standardized middle school modules focusing on homeland security and visual analytic fields.
   - Create an online high school visual analytics module.
   - Host summer K-12 teacher workshops that introduce visualization and analytics topics.
2) Undergraduate
   - Build on the Visual Analytics Summer Program at Purdue with recruitment of students from the entire CCI partner school network.
   - Expand undergraduate visual analytic course distribution among partner schools.
3) Graduate
   - Continue to develop the Visual Analytics Digital Library (VADL).
   - Create a Visual Analytics Masters Degree Program at one of the VACCINE partner schools.
   - Continue and expand the Affinity Research Pods throughout the VACCINE partner network.

II. Develop interdisciplinary professional training and outreach programs with the use of the existing educational materials and courses by adapting them so that they can be integrated into current professional training programs.

Actions

- Relationships have been cultivated with several institutions and VACCINE will work with each group in developing appropriate materials and training for their needs. Ultimately, the efforts with these organizations will provide the catalyst to integrate programs into other organizations. Our current partner institutions include the following:
  - Justice Institute of British Columbia (JIBC)
  - Federal Law Enforcement Training Center (FLETC)
  - Public Safety Academy of Northeast Indiana

III. Build a strong network of MSI educational and research programs both at community colleges and senior institutions through the MSI Cyberinfrastructure Empowerment Coalition (MSI-CIEC).

Actions

1) Develop involvement in Affinity Research Pods to include joint mentoring and implementation.

2) Recruit MSI students for the Visual Analytics Summer Program.

3) Involve HBCUs, Hispanic Institutions, and Tribal Connections, specifically the following:
   - Hispanic Association of Colleges and Universities
   - National Association for Equal Opportunity in Higher Education (NAFEO)
   - American Indian Higher Education Consortium (AIHEC)

Strategic Goal Three – Center Management and Growth

VACCINE will vigorously recruit, develop and engage a diverse set of employees and intra, inter, and external partners to provide real-world grounding, impetus, validation, evaluation, and transition of our work that will establish us as the recognized global leader in the Visualization Sciences.

Initiatives

I. VACCINE will work to vigorously recruit and retain top management talent when mission needs dictate and funding exists.

Actions

1) Our Leadership Board and Senior Advisory Boards will use their collective networks to identify potential applicants. In addition, the DHS COE network will be engaged as will appropriate marketing efforts.

II. Integration with Data Sciences co-lead. In order to have the greatest impact on the DHS mission, VACCINE and CCICADA, co-leads for the National Center for Command

Control and Interoperability, will work together to identify collaboration efforts within the entire spectrum as part of the strategic plan. VACCINE and CCICADA will determine opportunities that are beneficial to both centers, leverage their existing materials and network of partner schools for program sharing and dissemination, as well as reduce duplication of effort.

Actions

1) A continuous process will be in place where, periodically, high impact research and/or education projects with collaborative potential will be identified and coordination and work of these activities will commence.

III. Strengthen and expand collaboration with other DHS COEs.

Actions

1) All COE interactions will be centralized by the VACCINE lead. The Director will visit several COEs each year to explore collaborative opportunities and identify the best set of COE collaborations to engage.

IV. Diversity and Internationalization

Actions

1) We will expand our university partnerships through the substantial involvement of additional minority serving institutions.

2) We will expand our university partnerships to include European and US allied universities. We will initially explore international partnerships with the University of Konstanz and the University of Stuttgart in Germany and INRIA in France.

V. Enhance Visibility for the VACCINE Agenda

Actions

1) We will vigorously promote initiatives in high-value mission areas where the VACCINE team and international partner(s) will create a preeminent alliance.

2) We will develop and manage a marketing campaign addressing the VACCINE team. Media will include digital, print, and audiovisual processes.

3) We will work with our CCICADA team partners to promote collaborative research and education initiatives underway with our government partners so that future partners can be cultivated.

4) We will develop and maintain materials (digital, print, audiovisual) to publicize the Center and communicate its mission and efforts.

Strategic Goal Four – Technology Transition through Outreach

VACCINE will pursue public- and private-sector partners to enable intellectual property commercialization so that scientific breakthroughs and innovations can be transformed into useful technologies that can be made available to our global stakeholders.

Actions

1) We will learn about the variety of needs, issues, situations, and operational environments to ensure relevance of our activities and remove impediments for transition of technology, training, and education.

2) For each research project, we will identify a clear mission need, feasibility, and technology transition plan that will appropriately address needs whether from the local level through a continuum to the international level.

3) VACCINE will continuously develop local, state, tribal, federal and international homeland security government partners that will allow VACCINE technologies to be deployed in their organizations to be evaluated.

4) We will develop and work with corporate partners and national labs to transfer our technology from research demonstrations on real-world, real-scale problems to hardened, commercially deployable and viable applications.

III. Management Scope

To achieve the VACCINE objectives, we have devised a management plan with a strong organizational structure complemented by an expert leadership team. Our team has a strong collaboration history and experience that is essential to foster internal collaboration, as well as collaboration with DHS CCI, NVAC, other DHS CoEs, government agencies, field personnel, and other researchers. Our Center's management organization has three components: management staff, a leadership board, and a senior external advisory board.

- The management staff is intentionally lean in order to maximize research and education dollars while helping to minimize overhead. This team is responsible for overseeing the strategy, as well as the day to day operations of VACCINE, including interfacing with DHS, CCICADA and all external partners. The management staff consists of the following:

  - Center Director and Research Director – David Ebert
    David Ebert is a Professor in the School of Electrical and Computer Engineering at Purdue University, a University Faculty Scholar, a Fellow of the IEEE, and Director of the Visual Analytics for Command Control

and Interoperability Center (VACCINE), the Visualization Science team of the Department of Homeland Security's Command Control and Interoperability Center of Excellence. Dr. Ebert performs research in novel visualization techniques, visual analytics, volume rendering, information visualization, perceptually-based visualization, illustrative visualization, mobile graphics and visualization, and procedural abstraction of complex, massive data. Ebert has been very active in the visualization community, teaching courses, presenting papers, co-chairing many conference program committees, serving on the ACM SIGGRAPH Executive Committee, serving as Editor in Chief of IEEE Transactions on Visualization and Computer Graphics, serving as a member of the IEEE Computer Society's Publications Board, serving on the IEEE Computer Society Board of Governors, and successfully managing a large program of external funding to develop more effective methods for visually communicating information.

- Education Director – John Stasko

  John Stasko received the B.S. degree in Mathematics at Bucknell University in Lewisburg, Pennsylvania (1983) and Sc.M. and Ph.D. degrees in Computer Science at Brown University in Providence, Rhode Island (1985 and 1989). He joined the faculty here at Georgia Tech in 1989, and his primary research area is human-computer interaction. As the Education Director, he oversees and directs the education strategy of VACCINE.

- Managing Director – Timothy Collins

  Timothy Collins is the Managing Director of VACCINE. He oversees and implements all operational aspects of the Center including strategy, marketing, operations, finance, business development and human resources. He is an accomplished professional having a 25 year operations background in the public safety and security arena. He is a retired Indiana State Police Officer reaching the position of Region Commander and is a former Air Force Officer with a specialty in Command and Control. He is a graduate of the FBI National Academy and has an MBA from Purdue University. His business expertise includes program and project management, business development and strategic planning.

- Assistant Director of Education and Engagement – Marti Burns

  Marti Burns is the Assistant Director of Education and Engagement. She is responsible for educational initiatives ranging from K-12 programs through undergraduate and graduate level work, to professional training programs. She has worked with homeland security education for the past five years, leading an interdisciplinary team of

faculty in the development and implementation of new curriculum for master's level courses as well as the creation of curriculum for K-12 classrooms. In addition to education work, Marti has developed tabletop, functional and full scale exercises for area first responders. She received a B.S. in Industrial Engineering from Purdue University.

- Center Coordinator and Administrative Assistant – Jenny Kelly

Jenny Kelly serves as the information and communication manager for VACCINE.

Her duties include planning and scheduling meetings and appointments; organizing and maintaining paper and electronic files; managing projects; conducting research; and disseminating information to faculty, staff, and external partners. She managed a large scale multi-disciplinary medical practice in Illinois before moving with her family to Indiana. She began work as a legal assistant in a local Lafayette firm specializing in intellectual property, business transactions, litigation, estate planning, and employment law. In 2007, Jenny joined the Purdue Research Foundation working in the Office of Technology Commercialization before transferring to Purdue University.

- Research Pod Coordinator – Maureen Biggers

Dr. Biggers coordinates the education and research pod partnership program.

Biggers serves as Assistant Dean for Diversity and Education in the School of Informatics at Indiana University. She has been co-PI on three Broadening Participation in Computing projects, including the STARS Alliance, C-PATH diversity in curriculum, and is the Project Manager for the Alliance for the Advancement of African-American Researchers in Computing. She has worked in higher education administration for more than 30 years as a leader in both student and academic affairs.

- Assistant Director of MSI Programs – Richard Alo

Dr. Alo oversees the integration of minority serving institutions into VACCINE

research and education initiatives. Dr. Alo has served on several advisory and executive boards, including: Mathematical Association of America SUMMA Committee (Support for Undergraduate Minority Mathematics Awareness), Association of Departments of Computer and Information Science and Engineering at Minority Institutions (founding member), Coalition for Diversity in Computing (Chair), National Science Board NSF GRPA Committee, Multi Sector Crises Management Center (Board of Directors), and Houston Independent School District (HU-LINC Board). He chaired two national

conferences for ADMI and co-chaired the Richard Tapia Celebration of Diversity in Computing Symposium, Houston, 2001 and the Education Committee for Supercomputing in 2002. Dr. Alo was selected as the 2002 Educator of the Year by the Hispanic Engineer National Achievement Awards Conference.

- The VACCINE Leadership Board is responsible for the strategic decisions for the Center. This board is comprised of three permanent members plus three PIs from our universities on rotating two year terms, selected by the nineteen university PIs. The LB oversees the Seed Project review panel, advises on research, education, and outreach decisions and reviews and approves yearly Center plans.

    - David Ebert (permanent)
    - Timothy Collins (permanent)
    - John Stasko (permanent)
    - Pat Hanrahan (term)
    - Kelly Gaither (term)
    - Brian Fisher (term)
    - Alan MacEachren (term)

- The VACCINE Senior External Advisory Board provides strategic advice, information, and recommendations to the Center on basic and applied research activities, educational issues, and activities and operations related to visual analytics and command, control and interoperability in the homeland security environment. This board consists of the world's top experts in visual analytics and homeland security from the scientific, business, and government communities. This board has had one meeting thus far. It is anticipated that a minimum of two meetings will occur annually. Current membership includes:

    - Don Brackman
      President, National White Collar Crime Center
    - Chief John Buckman
      Past President, International Association of Fire Chiefs
    - Neville P. Clarke
      Special Assistant to the Vice Chancellor for Agriculture and Life Sciences and the Vice President for Research, Texas A&M University
    - Dieter Fellner
      Professor and Director, Technical University of Darmstadt & Fraunhofer Institute of Computer Graphics
    - William Fox
      Senior Vice-President, Global Anti-Money Laundering and Economic Sanctions Executive, Bank of America
    - David Kasik

Enterprise Visualization Architect, Boeing

- Daniel Keim
  Professor and Head, Information Visualization and Data Analysis Research Group, University of Konstanz, Germany
- Katherine A. Mitchell
  Director STEM Initiatives, American Indian Higher Education Consortium
- Mike Mitchell
  Director of Homeland Security Programs, Pacific Northwest National Laboratory
- Randall Murch
  PhD, Associate Director, Center for Technology, Security and Policy and Adjunct Professor, Virginia Tech
- Haesun Park
  Professor, Georgia Institute of Technology
- Jim Thomas
  Director Emeritus, National Visualization and Analytics Center
- Andrew Vallerand
  Director, Public Security Technical Program
- Leland Wilkinson
  Executive Vice President of SYSTAT Software
- Tom Wyss
  Indiana State Senator, DHS State and Local Officials Senior Advisory Committee for Homeland Security Advisory Council

- Tiered faculty participation. The VACCINE Center utilizes the best researchers in the field. Our research team is comprised of the following:

Florida International University	
Shu-Ching Chen	Computing and Information Sciences
Georgia Institute of Technology	
James Foley	Telecommunications
Ashok Goel	Computer Science and Cognitive Science
Carsten Görg	Interactive Computing
John Stasko	Interactive Computing
Indiana University	
Maureen Biggers	Diversity and Education
Jackson State University	
Jacqueline M. Jackson	Department of Computer Science
Loretta A. Moore	Department of Computer Science
Justice Institute of British Columbia	

Jack McGee	Public Safety
Murray Day	Emergency Management
Robert Walker	Emergency Management
Navajo Technical College	
Mark Trebian	Information Technology
North Carolina A & T State University	
Gerry Dozier	Department of Computer Science
Pennsylvania State University	
Guoray Cai	Information Science and Technology
Jin Chen	Geography
Lee Giles	Information Sciences and Technology
Frank Hardisty	Geographic Information Systems
Seogchan Kang	Plant Pathology
Alexander Klippel	Geographic Information Systems
Wang-Chein Lee	Computer Science and Engineering
Alan MacEachren	Geography
Prasenjit Mitra	Information Sciences and Technology
Donna Peuquet	GeoVISTA Center
Padma Raghavan	CyberScience
Anthony Robinson	Geography
Ian Turton	Geography
Purdue University	
Pamela M. Aaltonen	Nursing
Mimi Boutin	Electrical and Computer Engineering
William Cleveland	Statistics
Christopher Clifton	Computer Science
Edward Delp	Electrical and Computer Engineering
David Ebert	Electrical and Computer Engineering
Niklas Elmqvist	Electrical and Computer Engineering
Chris Foster	K-12 Programs
Arif Ghafoor	Electrical and Computer Engineering
Mourad Ouzzani	Computer Science
Sunil Prabhakar	Computer Science
Ahmed Elmagarmid	Computer Science
Simon Fraser University	
Richard Arias-Hernandez	Interactive Arts and Technology
Lyn Bartram	Interactive Arts and Technology
John Dill	Interactive Arts and Technology
Brian Fisher	Cognitive Science
Nathalie Prevost	Cognitive Science
Chris Shaw	Interactive Arts and Technology

Rob Woodbury	Interactive Arts and Technology
Stanford University	
John Gerth	Computer Graphics
Pat Hanrahan	Computer Science and Electrical Engineering
Jeffrey Heer	Computer Science
Swansea University	
Min Chen	Computer Science
The University of British Columbia	
Sidney Fels	Electrical and Computer Engineering
Edward Fortuno	Infectious & Immunological Diseases
Tobias Kollman	Infectious & Immunological Diseases
Rodger Lea	Media and Graphics Interdisciplinary Centre
Victoria L. Lemieux	Library, Archival, and Information Studies
Ronald A. Rensink	Psychology
University of Houston-Downtown	
Richard Aló	Computer and Mathematical Sciences
Ali Berrached	Computer and Mathematical Sciences
Ping Chen	Computer and Mathematical Sciences
Jianchao Han	Computer Sciences
Erin M. Hodgess	Computer and Mathematical Sciences
University of North Carolina-Charlotte	
Jianping Fan	Computer Science
Robert Kosara	Computer Science
William Ribarsky	Computer Science
KR Subramanian	Computer Science
William Tolone	Electrical and Computer Engineering
Jing Yang	Computer Science
University of Stuttgart	
Thomas Ertl	Computer Science
Daniel Weiskopf	Computer Science
University of Texas	
Kelly Gaither	Data & Information Analysis
University of Washington	
Daniel Dailey	Electrical Engineering
Mark Haselkorn	Human Centered Design & Engineering
Gael Tarleton	Port of Seattle
Virginia Tech	

Christopher Barrett	Computer Science
Stephen Eubank	Computer Science
Edward Fox	Digital Library Research Laboratory
Bruce Lawlor	Technology, Security, and Public Policy
Achla Marathe	Agricultural and Applied Economics
Madhav Marathe	Agricultural and Applied Economics
Arun Phadke	Electrical and Computer Engineering
Naren Ramakrishnan	Computer Science
Jeff Reed	Electrical and Computer Engineering
James S. Thorp	Electrical and Computer Engineering
Anil Vullikanti	Computer Science

IV. Research and Development Scope

The overall goal of VACCINE Center research is to create innovative and effective integrated data and visual analytic environments—frameworks, methods, and software—that advance the start-of-the-art in analyzing massive, heterogeneous, incomplete, temporally evolving homeland security data for anticipating, detecting, and responding to homeland security mission needs. Therefore, VACCINE is developing new dynamic visual analytics techniques based on cognitive and perceptual principles that increase the effectiveness of the entire analysis process; create precision information environments; create an information discourse environment; and enable visual analysis, knowledge synthesis, and extraction of insight for actionable decision making, event detection, management, and response.

For effective visual analytics, we are tightly integrating computational processing, data management and visual representation to efficiently present relevant information and solution alternatives to the user and to improve both data management and statistical data modeling. To ensure successful deployment, scalability, and usability of developed technologies, we are integrating, evaluating, and continuously refining our work through collaborative participation in Mission Driven Research Projects (MDRPs) ranging from emergency planning and response to health care monitoring to intelligence analysis. All of our MDRPs have the following key pervasive homeland security problem characteristics: massive, time evolving, heterogeneous data of varying quality; multiple-scales of problem solving, planning, management, and action; coordinated information analysis; information security and privacy challenges; and a wide variety of needs, skills, and computing resources of involved personnel. Each of our Mission Driven Research and Education Projects are reported in detail in Sections VI and VIII of this document.

To achieve our goal of creating this visualization and analytics environment, our research plan is based on three foundational research areas (previously discussed) that form our framework for addressing DHS CCI's homeland security mission directives: Interactive Visual Analytic

Foundations, Visually-Adapted Analytical Techniques, and Investigative Analysis and Decision Making Environments.

The involvement of end-user clients from the beginning in these research projects is critical to ensure real-world applicability, scalability, usability, and evaluation. VACCINE acknowledges that in order to build a "solution," it is imperative to begin with a "problem." That "problem" must start with our customers. In light of this, VACCINE has an active campaign to engage the end-user in the requirements process from the beginning. To date, we have formed five public safety consortiums which will oversee the research and development process to guarantee the usability and utility of our products.

1. Indiana Public Safety Consortium
2. Florida Division of Emergency Management's Region Seven Consortium
3. UNCC Consortium
a. Charlotte Mecklenburg Police Department Consortium
4. Washington State Consortium
5. Penn State Law Enforcement Consortium

The largest consortium, in Tippecanoe County, IN, represents over 250 law enforcement and emergency management personnel. This particular consortium is a VACCINE resource intended to be utilized by all VACCINE partner schools and the DHS CCI division of S&T.

The Tippecanoe County consortium has embedded a graduate student within the law enforcement community which gives us direct access to over 10 years of digital data. This data will allow us to conduct projects with the goals to coordinate and to improve evidence-based policing, prospective policing, and results-oriented policing. In addition, this consortium will work to coordinate and improve emergency management's planning, response, mitigation, and recovery efforts for any emergency, manmade or natural, in a direct or supporting role.

Having our customers directly integrated with our mission-driven research projects allows them to provide input, project instigation, and relevant feedback and transition paths for DHS technology transfer. The MDRPs are 18 to 36 months in duration and our initial set includes projects leveraging successful R&D from the former RVACs and new faculty, as well as several new initiatives. All of these projects are collaborative among the universities and specific personnel initially involved in each project are identified.

The MDRPs are intended to be an evolving set of projects to explore and solve targeted, developing DHS mission needs. We are not just deriving ad-hoc solutions to problems, but addressing the fundamental research questions, advancing the science of visual analytics, and creating a solid theoretical foundation for both future research and applications of that research to diverse homeland security challenges. The research infrastructure and focus on common research challenges and solution techniques provides a generic, robust infrastructure, as well as

tools for rapidly developing solutions for real-world, mission-critical homeland security problems.

Our research is extending the state of the art in many areas including the following:

- Theoretical work in the science of analytical reasoning
- Fundamentals of visual representations for multisource, uncertain, temporally evolving, and conflicting data
- Data representations, integration, quality assessment, cleaning, and reduction for massive datasets, streaming data, and multisource data integration
- Methods and novel visual analytic tools tailored to user, task, and device requirements
- Mobile visual analytics for real-time applications
- Novel spatiotemporal modeling, detection, and prediction techniques
- New multisource correlation and visual analysis techniques (e.g., linked animal and human health surveillance)
- Novel video analytics for tracking, event detection, and mobile video analytic applications
- Novel population distribution adaptive cluster detection and aberration detection techniques
- Novel synthetic data generation techniques for spatiotemporal data streams that match observed statistical properties of source data but are completely anonymized
- Adapted presentation methods for actionable decision making for a diverse set of users

V. Research Project Descriptions

MDRP 1: Visual Analytics for Investigative Analysis on Text Documents

Team

VACCINE:
Georgia Tech: John Stasko
Pennsylvania State University: Alan MacEachren, Prasenjit Mitra
Simon Frasier University: Brian Fisher

CCICADA:
University of Southern California: Ed Hovy

Abstract

The goal of Pennsylvania State University with this work was to enable analysts to forage for and make sense of place-relevant information extracted from text documents linked to formal geographic information. The visual analytic system developed can retrieve and process text documents, identify information of interest, present the information via an interactive visual interface, interactively record analyst-provided information and update the visually presented information to assist the analyst. Applications include monitoring the news media to track mass movement of people in response to events and monitoring Twitter feeds as input to investigative analysis. One component of the research has resulted in a prototype web application, SensePlace. The scientific goal of SensePlace is to visually and computationally support analyst sensemaking with text artifacts that have potential place, time, and thematic relevance to an analytical problem through identification and visual highlighting of named entities (people, places, times, and organizations) in documents, automated inference to determine document relevance using stored knowledge, and a visual interface with tightly coupled geographic map, timeline, and concept graph displays used to contextualize potentially relevant documents.

The team developed techniques and tools for investigative analysis on large collections of documents including both unstructured text and structured documents (e.g., spreadsheets). They developed systems that use interactive visualizations to help analysts with performing exploration, sense-making, and understanding tasks. If they could help analysts determine which documents to read first and to also find important, connected documents, then they could help them to make timelier and accurate assessments of the situations described within those documents. The project builds upon the Jigsaw system, and explores how new capabilities can be added to it. They also planned to disseminate the system for use by real clients, people, and organizations in various domains.

Technical Approach

Data collection methods used were compiling & sorting DB, data mining, and expert consultation. Analytic methods used were case studies, sampling, and statistical analysis. The

nature of research was applied, basic, consultation, coordination/integration, education, hybrid basic-applied, and hybrid applied-consultation.

Milestones

Milestones for VACCINE Period 1	**07/01/09 – 03/31/10**			
	Quarterly			
Development of text-layout views (e.g., WordTree)		X	X	
Design of infrastructure for geo-visual visualizations		X		
Implement geo-visual visualizations			X	X
Experimentation with variety of document clustering methods		X	X	
Development of visualizations of clustering methods			X	X
Experiments with event identification software systems to assess viability			X	
Implement visualizations of events identified in text				X
Complete implementation of 1st generation SensePlace application and create video demo of application use		X		
Submit paper for publication			X	
Create development plan for full GeoTemTA environment			X	
Integrate ideas and components from TexPlorer methods and SensePlace into GeoTemTA and create working prototype			X	X
Develop methods for representing and working with information/artifact uncertainty				X
Evaluation of developed techniques				X

Project Outcomes (Results, Accomplishments, Conclusions)

Publications

1. Carsten Görg, Jaeyeon Kihm, Jaegul Choo, Zhicheng Liu, Sivasailam Muthiah, Haesun Park, John Stasko. "Integrating Computational Analysis with Interactive Visualization in Document-Focused Visual Analytics Systems." Submitted to IEEE VAST 2010.
2. Jaiswal, A., D. J. Miller, et al. (2010). "Un-Interpreted Schema Matching with Embedded Value Mapping under Opaque Column Names and Data Values." IEEE Transactions on Knowledge and Data Engineering **22**(2): 291-304. {cross MDRP – 1 & 13}
3. Robinson, A. (2009). Needs Assessment for the Design of Information Synthesis Visual Analytics Tools. IEEE International Conference on Information Visualization. Barcelona, Spain, 353-360.{cross MDRP – 1 & 13}
4. Robinson, A. (2009). Visual Highlighting Methods for Geovisualization. 24th International Cartographic Conference. Santiago, Chile (CD only) {cross MDRP – 1, 3, 12 & 13}
5. Tomaszewski, B., J. Blanford, K. Ross, S. Pezanowski and A. MacEachren (submitted). "Supporting Rapid Sensemaking in Diverse Web Document Foraging." *Computers, Environment and Urban Systems*

6. Youn-ah Kang, Carsten Görg, John Stasko, "Evaluating Visual Analytics Systems for Investigative Analysis: Deriving Design Principles from a Case Study", Proceedings of IEEE VAST '09, Atlantic City, NJ, October 2009, pp. 139-146.
7. Youn-ah Kang, Carsten Görg, and John Stasko, "How can Visual Analytics Assist Investigative Analysis? Design Implications from an Evaluation." Accepted for publication in *IEEE Transactions on Visualization and Computer Graphics*.
8. Pike, William, John Stasko, Remco Chang, and Theresa O'Connell. Science of Interaction. Journal of Information Visualization (2009). *Work under UNC Charlotte VACCINE project.*

Other outcomes / impacts

The contribution at Penn State to research on visual analytics for investigative analysis on text documents has focused on geographic components of unstructured text information and the challenges of foraging for and sensemaking with that information. Geographic knowledge building from unstructured web sources starts with web document foraging during which the immense quantity, scope and diversity of web-based information create incredible cognitive burdens on an analyst's or researcher's ability to judge information relevancy. Determining information relevancy is ultimately a process of sensemaking.

The first half of the time period was devoted to preliminary research on visually supporting web document foraging and sensemaking and completing the implementation of the initial prototype that began under their Regional Visualization & Analytics Center funding. More specifically, they implemented the Sense-of-Place (SensePlace) analytic environment. They demonstrated a proof-of-concept by applying the tools to an analysis of using SensePlace to uncover potential population migration, geopolitical, and other infectious disease dynamics drivers (measles epidemics) in Niger. Their analysis allowed them to demonstrate how their approach can support analysis of complex situations along (a) multi-scale geographic dimensions (i.e., vaccine coverage areas), (b) temporal dimensions (i.e.,

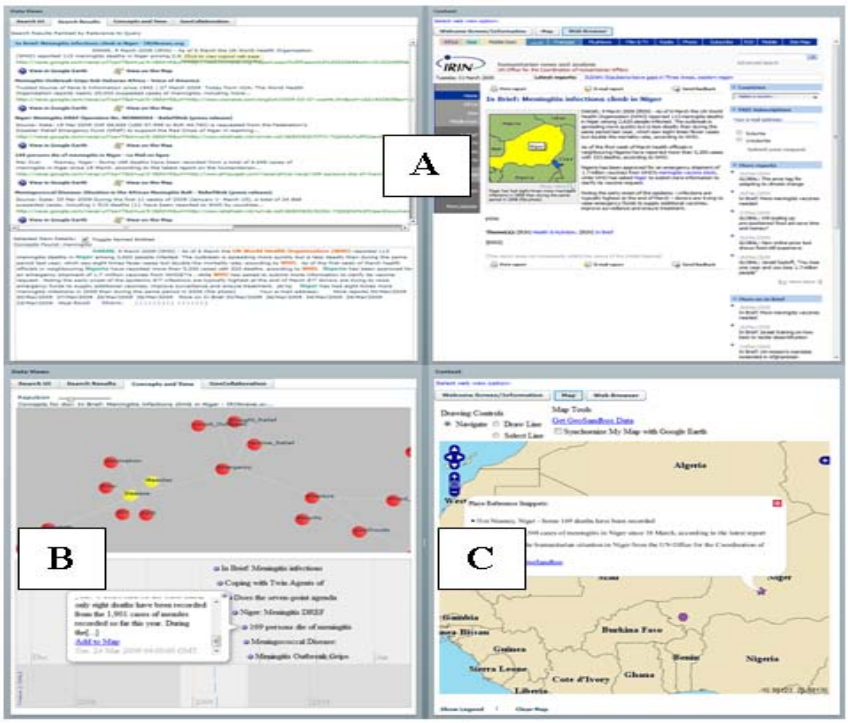


Figure 1: Screen shot of SensePlace.

seasonal population movement and migrations), and (c) diverse thematic dimensions (effects of political upheaval, food security, transient movement, etc.).

Figure 1 was a snapshot during a SensePlace analysis session. In this figure, the analyst was examining place references and thematic dimensions related to a story about a food crisis in Niger. In figure 2 below, the diagram illustrates the process used for foraging and sense making. The tool searched the web for text documents about a particular geographic location using the search terms entered by the user. Retrieved documents are displayed in the Results tab and analyzed by the user through the visual interface components (see Figure 1). The user views the articles by (i) viewing the abstract and highlighted keywords to determine relevancy, (ii) viewing where the article is throughout the map, and (iii) viewing the concepts affected and the sequence of events through the timeline. Information from relevant documents can be captured using a browser add-in (Zotero) and transcribed to a spreadsheet for later use. In addition, links to the articles and search terms can be saved using the Save Results Option in SensePlace (artifacts are saved to an XML and TXT file, respectively, which can be shared with colleagues or uploaded again if required in the future). Information from the articles/documents highlights new information and prompts new searches using new keywords. The process is repeated again until the user is satisfied that they have gathered sufficient information about the topic of interest. Information gathered through this tool can then be collated, synthesized and summarized by the user. Features/components contained within SensePlace are illustrated within the red box.

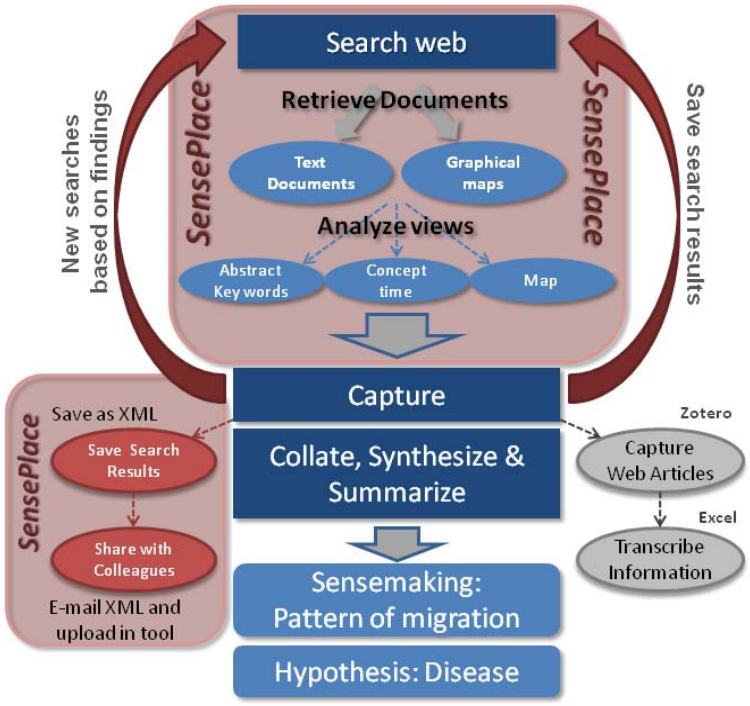


Figure 2: SensePlace foraging and sensemaking process.

Following completion of the initial SensePlace application, work focused on a second generation version that is more scalable and more adaptable to a range of information sources. In this work during the second half of the time period, they implemented a visual analytic interface that analysts can use to examine place, time, and concept components of information extracted from multiple text sources. As part of a test case study, the team compiled a database of Twitter feeds containing over 1 million Tweets and implemented tools to ingest GeoRSS feeds generated by a variety of sources. They implemented core parts of their system architecture including a distributed database, server technologies, and an initial user interface. They designed the

methods for entity and relationship extraction from Twitter feeds and developed a strategy for contextualization of the Twitter feeds by fetching feeds posted by the same user. They also did preliminary evaluation and adaptation of the back-end components.

Figure 3 below illustrates the Text Information Channel Backend for the second generation SensePlace that will support a wide variety of information feeds such as Tweets, GeoRSS feeds, HTML and plain text. In the milestone chart, this second generation tool was tentatively called GeoTemTA (for Geo-Temporal Text Analytics). This core architecture is being leveraged for use in MDRP 13 as well.

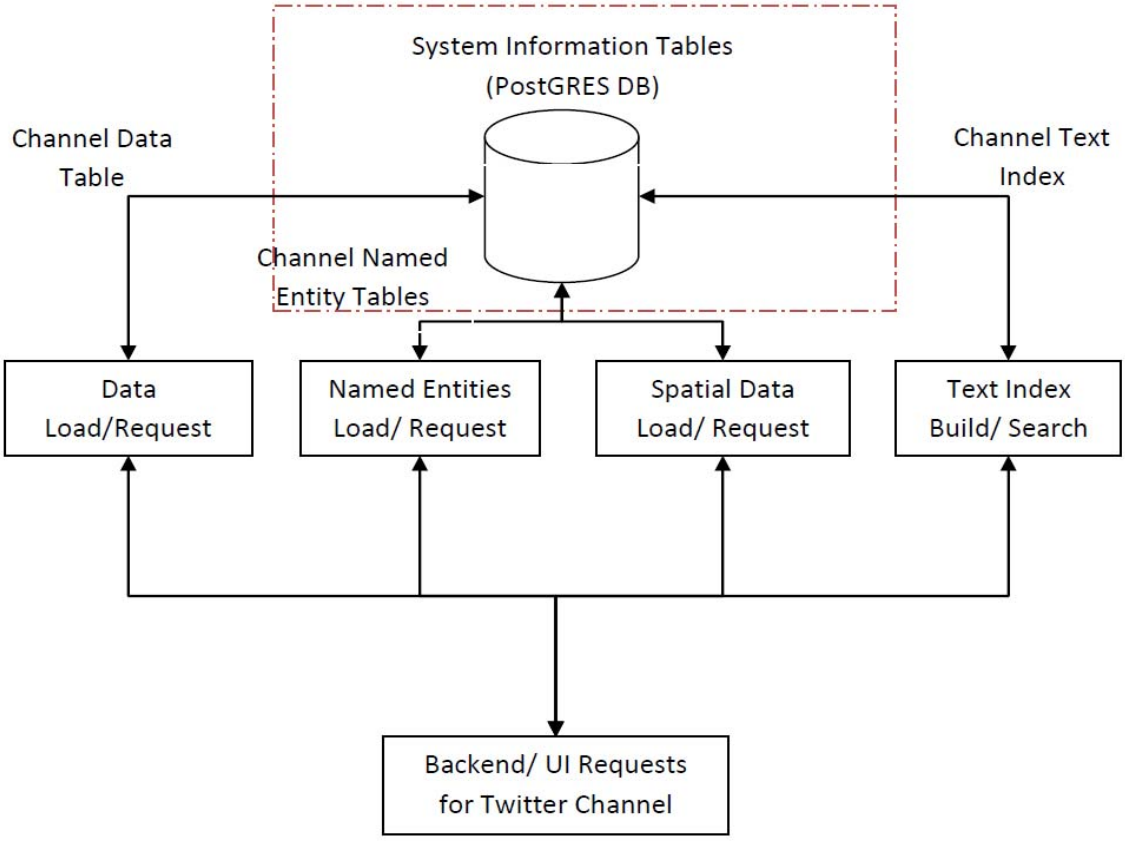


Figure 3: SensePlace Text Information Channel Backend Architecture

The system consists of a central database containing all information such as data tables, associated named entity tables, location tables and text search indices for each information channel which allow the UI and the backend text processing software to run. The architecture was designed such that a wide variety of information channels can be easily integrated and allow the user interface/processing software to quickly access and process this information. Figure 4 below illustrates the Twitter Information channel storage and processing backend. Each channel of information currently stored consists of four backend processing daemons.

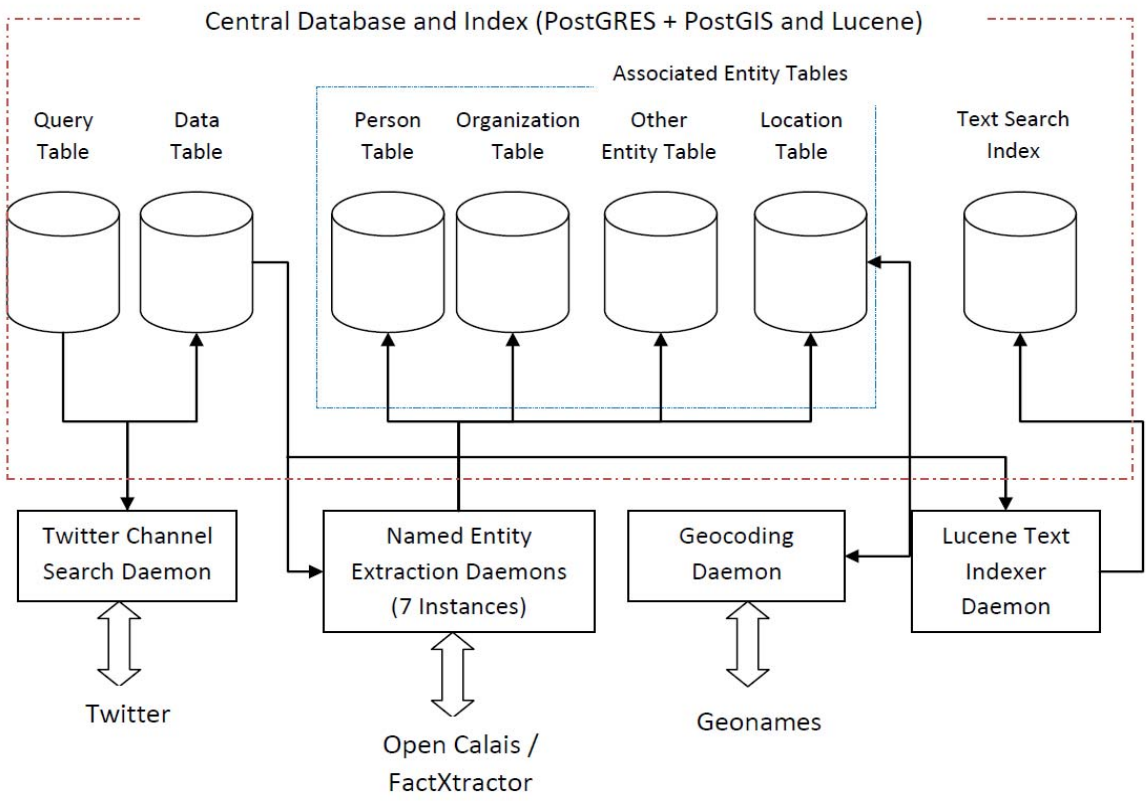


Figure 4: The SensePlace Twitter Information Channel Storage and Processing Backend

System components are described below:

1. Channel Search/Data Load Daemon: Each channel has an associated daemon which collects information of interest. For example, the Twitter channel search daemon currently uses a table which contains queries that are passed on to Twitter. Twitter then returns a set of Tweets which are then stored in the channel data table. Each channel data table typically contains single column of text (in this case it contains the text contained within the Tweets returned from a Twitter search) which is of interest. This column information is stored in the system information table as shown in Figure 1 is used by the text processing software.
2. Named Entity Extraction Daemon: The named entity extraction daemon extracts the named-entities contained within the text present for each channel and writes this information to the associated named-entity tables in the database. There are four broad categories of named entities that are extracted namely people, locations, organizations and others. The current daemon uses the Open Calais named-entity extraction web service. They are working towards integrating FactXtractor named-entity extractor as well. The named-entity extraction daemon is written such that it can be run on multiple machines on the same network allowing large amounts of text information to be processed in real time. The Twitter channel currently uses 7 machines to perform named

entity extraction. Examples of extracted locations and organizations are shown in Figures 4 and 5.

3. Geocoding Daemon: Once the named-entity extraction has occurred, a location table is generated which contains text information (e.g. “Haiti”) regarding locations contained within a channel (e.g. Figure 3). While this is suitable for text querying, however, such information is not of use when the user might want to perform spatial querying on this channel. For e.g., a user might want to use a bounding box in a map interface to restrict the information to that corresponding to the locations within this bounding box. The Geocoding daemon processes the location table (for all channels) and geocodes these locations with the point coordinates of the best matching location returned from Geonames.org web service. These point coordinates are stored within the location table as shown in Figure 6.
4. Lucene Text Indexer Daemon: As discussed in 1) each channel typically contains a single column of text in the channel data table. For example the Twitter channel data table contains one column which stores all the Tweets that were extracted by the channel search daemon. The Lucene text indexer daemon creates a Lucene text index for the channel text for full text querying. The index also stores dates where the lexicographic property of the stored string is used to allow text and date range (for e.g., Tweets about “swine flu” between dates “Jan 2009” and “Dec 2010”), text and day (for e.g., Tweets about “swine flu” on “Friday”), text and date (for e.g., Tweets about “swine flu” on date “10th Jan 2009”), text and day + date range type querying (for e.g., Tweets about “swine flu” between dates “Jan 2009” and “Feb 2009” occurring on days “Friday” and “Saturday”).
5. Channel Data Middleware: The middleware API layer provides an integrated access layer that allows all the information stored within the repository to be quickly accessed by the frontend. For example, Figure 7 shows the frontend where an analyst queried for “Earthquake”. The middleware layer allows the frontend to quickly access the information repository and display the information (relevant Tweets, associated persons, organizations, locations and temporal information, etc.) for consumption by a user.

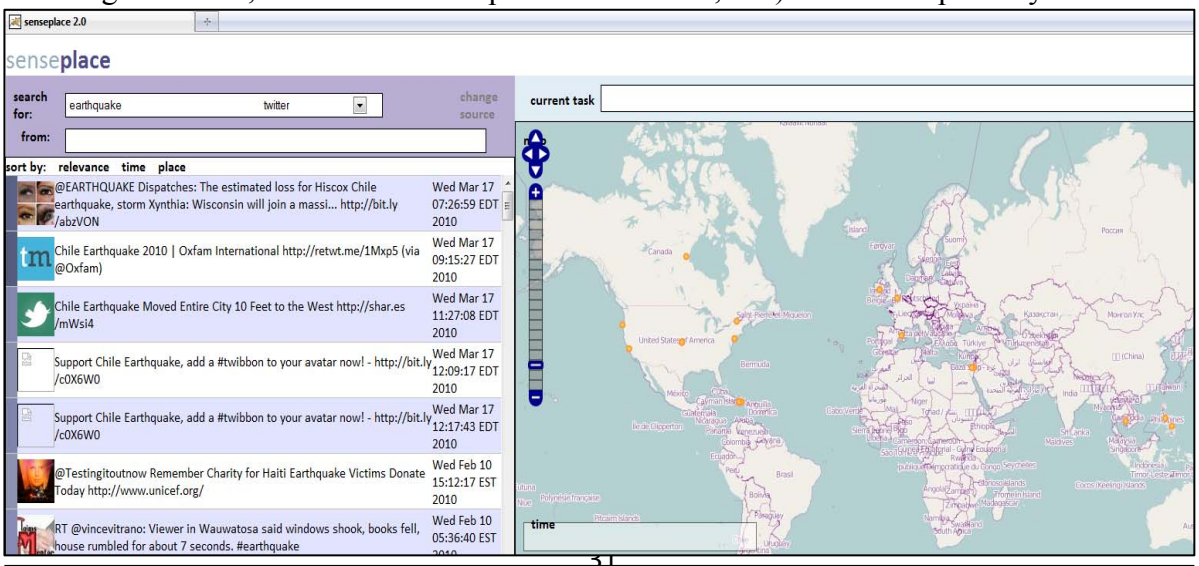


Figure 5: The Frontend showing a query interface returned Tweets and a map with Tweet locations.

Figure 5 and 6 show examples of Tweets that were loaded by the Twitter Channel Search daemon and the corresponding entities (Locations, Organizations and Persons) extracted by the Named Entity Extraction daemon. Figure 6 shows the location table for the Twitter channel and the spatial coordinates that were assigned to the locations by the Geocoding daemon.

	A	B	C	D
1	sysid	tweettext	name	type
2				
3	21	www.mrrlest8.com The H1N1 vaccine is still in short supply, as despite U.S. government assuran	united states	country
4	22	RT @Doc__D: Reading about h1n1 in pregnancyÃ‚Â»Ã‚Â» any cases at UWI?	reading	city
5	26	Survivor Tv Show ~~~ H1N1 Spread Prevention |WBNGTV: News Sports, Weather Binghamton, N	new york	provinceorstate
6	32	Mother of Kansas girl who died in Indy during FFA convention says daughter tested positive for H	kansas	provinceorstate
7	48	H1N1 Flu Virus: Government Of Canada Provides Travel Guidance: The Government of Canada to	canada	country
8	52	China: A/H1N1 accounts for 80 pct of China's total flu cases: official http://bit.ly/myCaN	china	country
9	60	H1N1 flu vaccine shortage impacting public flu clinics in Baltimore - http://tinyurl.com/yh5bpsd.	baltimore	city
10	61	Moroccan Health Minister said that 82 cases of H1N1 influenza have been diagnosed in a single d	casablanca	city
11	61	Moroccan Health Minister said that 82 cases of H1N1 influenza have been diagnosed in a single d	rabat	city
12	78	Ten H1N1 vaccination clinics in Toronto to serve 2.5 million people? GREAT PLAN! Looks like our	toronto	city
13	92	US Faces H1N1 Flu Vaccine Shortage - Voice of America: Voice of AmericaUS Faces H1N1 Flu	america	continent
14	99	Mt Sinai Hospital in Toronto is trying to avoid the criticism of not communicating to the public, as	toronto	city
15	112	Planes, Traines And The H1N1 Flu Virus, Canada Travel Guidelines: The Government of Canada t	canada	country
16	121	US Faces H1N1 Flu Vaccine Shortage - Voice of America.. http://bit.ly/1RtUFo #h1n1	america	continent
17	132	CVS pharmacies just announced they have H1N1 vaccines in Florida, Nevada, & North Carol	florida	provinceorstate
18	132	CVS pharmacies just announced they have H1N1 vaccines in Florida, Nevada, & North Carol	nevada	provinceorstate
19	132	CVS pharmacies just announced they have H1N1 vaccines in Florida, Nevada, & North Carol	north carolina	provinceorstate
20	134	First H1N1 Related Death Confirmed in Madison County today...another death also reported from	madison county	provinceorstate
21	134	First H1N1 Related Death Confirmed in Madison County today...another death also reported from	mccreary county	provinceorstate
22	142	10-Point Checkup: H1N1 flu in Placer County widespread, vaccine upon way http://tinyurl.com/yk	placer county	provinceorstate

Figure 6: Example Tweets and the extracted locations and types

sysid	tweettext	name	type
1058	HHS waives certain Medicare, Medicaid, CHIP requirements for H1N1 flu: http://ow.ly/xkDA #fb	medicare	organization
1063	'Further rise' in swine flu cases (BBC News).. http://bit.ly/KgkSr #h1n1	bbc	company
1066	Emergency approval from FDA for new swine flu test (AP via Yahoo! News).. http://bit.ly/1LK7lm #h1n1	fda	organization
1066	Emergency approval from FDA for new swine flu test (AP via Yahoo! News).. http://bit.ly/1LK7lm #h1n1	yahoo!	company
1070	North Canton Company Fights Against H1N1 with New EPA Registered Hard Surface Disinfectant, Formulat	environmental protection agency	organization
1077	North Canton Company Fights Against H1N1 with New EPA Registered Hard Surface Disinfectant, Formulat	environmental protection agency	organization
1096	The line-up at NY Civic Centre at 11:48 for #h1n1 shot. Cops said 4 - 5 hr wait. http://pic.gd/d7921f	ny civic centre	organization
1108	PA Department of Health: National Delays in H1N1 Flu Vaccine Distribution Highlight Need for Continued Pr	pa department of health	organization
1112	News Flu H1N1: North Canton Company Fights Against H1N1 with New EPA Registered Hard Surface Disin	environmental protection agency	organization
1113	News Flu H1N1: China to lift swine flu-inspired ban on US pork - The Associated Press http://ow.ly/15Y7f6	associated press	company

Figure 7: Example Tweets and extracted organizations and types

Figure 7 shows a temporal analysis of Tweets for a broad keyword “earthquake” that may be of interest to an analyst. Tweet counts can be very useful for discovering events as Figure 8 shows. For example, they found two abnormal spikes in Figure 9 corresponding to the dates 02/10/10 and 03/02/10. On those two dates, earthquakes struck the Northern Illinois region and Chile respectively. The current indexing mechanism provides an easy way to extract and visualize such information, so an analyst may be able to find relevant information within the vast stores of irrelevant information. In addition, a future direction for work is an automated event detection mechanism which would essentially discover such abnormal spikes in information related to a broad keyword of interest (for e.g. Earthquake, Hurricane, Tornado, Disease, etc.) followed by automated analysis of information present within these spikes. For e.g., consider the spike in Tweet counts for the “Chile Earthquake”. A 2-gram analysis of all Tweets in this spike should have a very high frequency of the words “Chile” and “Earthquake” appearing together in Tweets.

Based on the high frequency of appearance of "Chile" and "Earthquake", the system can automatically use these as keywords for a more directed search of relevant information to the Chile Earthquake. This should provide a better filtering mechanism so as to remove irrelevant information, as well as provide an automated way for alerting an analyst towards an event of interest. They are currently working on an implementation of such a system.

	A	B	C	D	E	F	G
1	locid	name	type	sysid	geocoded	geonameid	coordinates
2							
3	19401	pakistan	country	97002	t	1168579	(30,70)
4	19402	florida	provinceorstate	97004	t	4155751	(28.7505408,-82.5000976)
5	19407	united states	country	97008	t	6252001	(39.76,-98.5)
6	19408	france	country	97008	t	3017382	(46,2)
7	19416	turkey	country	97018	t	298795	(39,35)
8	19417	delhi	city	97019	t	1261481	(28.6128189621009,77.2311401367188)
9	19418	delhi	city	97021	t	1261481	(28.6128189621009,77.2311401367188)
10	19419	delaware	provinceorstate	97021	t	4142224	(39.0003906,-75.4999224)
11	19420	india	country	97022	t	1269750	(20,77)
12	19421	india	country	97025	t	1269750	(20,77)
13	19422	wales	country	97027	t	2634895	(52.5,-3.5)
14	19423	wales	country	97029	t	2634895	(52.5,-3.5)
15	19337	baltimore	city	96943	t	4347778	(39.2903848,-76.6121893)
16	19350	united states	country	96956	t	6252001	(39.76,-98.5)
17	19353	cuba	country	96958	t	3562981	(22,-79.5)
18	19355	united states	country	96959	t	6252001	(39.76,-98.5)
19	19356	france	country	96959	t	3017382	(46,2)
20	19368	leogane	city	96971	t	3722286	(18.5108333,-72.6338889)
21	19371	united states	country	96973	t	6252001	(39.76,-98.5)
22	19377	united states	country	96978	t	6252001	(39.76,-98.5)

Figure 8: Example locations extracted from the Tweets and their Geonames ids and latitude, longitude coordinates.

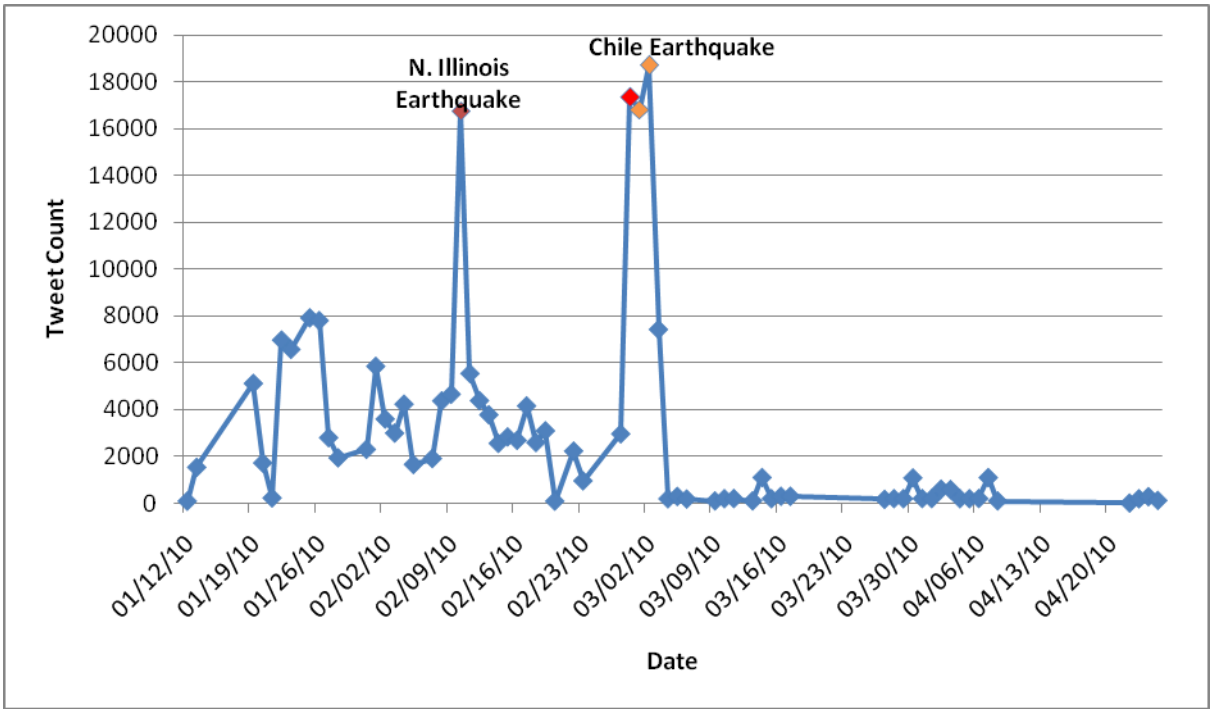


Figure 9: Tweet count as a function of date for the keyword "Earthquake". Two earthquakes "Chile Earthquake" and "N. Illinois Earthquakes" are events that can be statistically detected (spikes) based on daily average Tweet counts

The SensePlace application was specifically focused on fostering geographically-centered sensemaking through identification and visual highlighting of named entities (people, places, times, and organizations) in documents, automated inference to determine document relevance using stored knowledge, and a visual interface with tightly coupled geographic map, timeline, and concept graph displays. They used a case study (focused on uncovering potential population migration, geopolitical, and other drivers for infectious disease dynamics in Niger) to demonstrate how our approach can support analysis of complex situations along (a) multi-scale geographic dimensions (i.e., areas where vaccine coverage is non-existent or limited both within Niger and the surrounding countries), (b) temporal dimensions (i.e., population movement within Niger due to seasonal migrations related to food and livelihood), and (c) diverse thematic dimensions (effects of political upheaval, food security, transient movement, etc.). The current re-designed system increases the scalability of the system by orders of magnitude allowing the analyst to examine a large text corpus. They also enhanced the system to be able to handle streaming text data with high efficiency. As a case in point, they showed how the system can use Twitter data and RSS feeds to enable visual analytic exercises.

In terms of evaluation, last spring and summer the team at Georgia Institute of Technology performed a study with a simulated intelligence analysis session and compared the use of Jigsaw to other more traditional techniques such as using pen and paper or using desktop search. Jigsaw generally seemed to support more thorough analysis, but most importantly the study helped them identify different investigative strategies that people follow with visual analytics tools and it helped them to derive design principles for visual analytics tools. They published a paper at VAST 2009 about this study and are publishing a journal paper in *IEEE Trans. on Visualization and Computer Graphics,* as well.

In terms of the system, they had goals of doing more with geovisualizations, visualizing events, and visualizing uncertainty. As they worked more with the system, they decided to instead focus more on the database capabilities, entity identification, implementing new views of the documents and entities, and developing further computational analysis capabilities. All of these directions were extremely successful. The system now stores document analysis information in a database on disk, so it can scale to much larger document collections. They added and refined the entity identification capabilities. They added new Circular Entity graph and Document Grid views. Finally, they are making excellent progress adding new document analysis capabilities to Jigsaw. They implemented automated document clustering by theme, and they implemented document similarity analysis. They also built document summary capabilities, so they can give one sentence from each document that most typifies it. Finally, they began to explore sentiment analysis capabilities.

They created a new release (0.3) of the system and began distributing it to people and organizations who want to use it. They have a growing and wide-ranging client set now and continue to work on expanding it. Articles about Jigsaw have appeared in the DHS R-Tech newsletter and Fraud Magazine, a magazine for investigators of money laundering and fraud cases. They are also exploring new document domains for analysis such as bioinformatics, business markets, on-line reviews, and academic papers.

Jigsaw relevant:

New views (visualizations) were added to the Jigsaw system including a WordTree view and a Circular Graph view.

A significant amount of time was spent in creating a distribution version of the system that could be given to academic colleagues, as well as, various government and corporate interested clients. In order to do so, a major rework of the system architecture was performed, changing how documents are stored during analysis. They moved to a database model of document storage, so they could interface with much larger document collections.

In January, the 0.3 version of the system was created including documentation, tutorials, example data, etc.

Articles about the system appeared in newsletters and magazines, and they helped bring additional interested parties to VACCINE.

In the third quarter, the research focused on adding further computational analysis capabilities to the system including document clustering. These algorithms provide more analytic power to Jigsaw and, from the initial use of the algorithms, appear to enable many new types of analyses. A paper was submitted to the VAST Conference about these aspects.

MDRP 2: Integrate Automated Analytical Reasoning into Jigsaw

Team

VACCINE:
Georgia Institute of Technology: John Stasko, Ashok Goel

Abstract

The objective of this project is to develop and integrate more automated analysis and reasoning capabilities into the Jigsaw system. Presently, as a visual analytics system, Jigsaw is heavy on the visualization aspect and the analyst explores the document collection manually. While VACCINE believes there is much value in this, the team would like to integrate more automated reasoning and analysis capabilities into the system as well. So, for instance, the system might suggest hypotheses about the documents, it might help the analysts confirm/refute hypotheses they have developed, or it might simply suggest avenues to explore more. The team will explore how Jigsaw can be combined with the STAB system also created at Georgia Institute of Technology, which provides these types of automated analysis capabilities. Such a task will be challenging, because the two systems were implemented in completely different environments and with very different data models. Here, the team plans to see if these two approaches can be combined, and then will use an example document collection as a case study.

Technical Approach

Data collection method used was case studies, and the nature of the research was the analytic method and the basic method.

Milestones

Milestones for VACCINE Period 1	**07/01/09 – 03/31/10**			
	Quarterly			
Survey and research existing evidence marshalling tools		X	X	
Design advanced evidence marshalling for Jigsaw			X	X
Implement evidence marshalling capabilities into Jigsaw				X
Design how to present STAB analysis results in Jigsaw		X		
Implement STAB analysis results views in Jigsaw			X	X
Design of how to use Jigsaw to generate hypotheses into STAB			X	X
Overall integration of Jigsaw and STAB		X	X	X

Project Outcomes (Results, Accomplishments, Conclusions)

Other outcomes / impacts

Progress was slow during the first two quarters of the project, but the team made good strides in the final quarter. They presently have a working version of a system that integrates Jigsaw with STAB. On the one hand, an analyst can have automated reasoning performed in STAB and the results can be examined (item by item) in Jigsaw. On the other hand, Jigsaw explorations can

lead to hypotheses that can be entered into STAB for automated checking and analysis. The team made a video of the combined systems in March. They have not authored any publications about this project yet, but are working on a draft of one now.

Many internal changes to the data structures and models within STAB were necessary to accomplish the integrated changes between STAB and Jigsaw. In the resulting system, analysis can proceed in different fashions. In one, the analysts can run STAB's core automated reasoning capabilities. Its output and results are now presented in a manner that coordinates with Jigsaw. So, for instance, the analyst can select an individual mentioned in one of the hypotheses developed by STAB, and explore that individual's information within Jigsaw. This acts as a kind of sanity check for what STAB produces. In another fashion, an analyst can explore documents as they usually would in Jigsaw, but they can use a new interface to send hypotheses to STAB for its automated analysis on those hypotheses (e.g., STAB might find evidence that confirms or contradicts the hypothesis).

To demonstrate these capabilities, the team created a case study of the combined system running on the VAST Conference Contest dataset from 2006. Events within the documents were manually coded, because they do not have any software that does automatic event identification reliably.

The team created a video of this system integration in June for Joe Kielman's request for videos. They outlined a paper about this project, but have not yet begun writing the paper.

As for the evidence marshalling capabilities within Jigsaw, that project is in progress. In the final quarter, the team continued with the design of this component and began implementing it. Some initial usage of the component and feedback was gathered from that and sent back to redesign some of the pieces. Further development of this component is still underway.

MDRP 3: Situational Surveillance & In-field Criminal Investigative Analytics

Team

VACCINE:
Pennsylvania State University: Alan MacEachren, Frank Hardisty
Purdue University: David Ebert, Mimi Boutin, Ed Delp, Bill Cleveland
Virginia Polytechnic Institute and State University: Chris Barrett, Stephen Eubank, Madhav Marathe

CCICADA:
Rutgers University: Fred Roberts

Abstract

The over-arching goal for the first year was to develop methods for combining syndromic surveillance data with detailed synthetic social network based epidemic simulations for improved state assessment, support of insight building and consequence analysis for decision makers.

VACCINE developed novel methods for enabling event animation in "map mashups" and for more efficient web-map client server interaction. In addition, advances in spatio-temporal analysis methods have been achieved and integrated into the GeoViz Toolkit. Discussions have been initiated with the Harrisburg Department of Police about transitioning our CrimeViz web-mapping tool for representation and tracking of crime incidents.

Work in this project focused on development and application of new exploratory visual analytic (EVA) mapping methods to support criminal investigative analytics. The team developed situational surveillance visual analytics tools to process large volumes of geographically referenced data searching for events of interest. This work will integrate the Visualization Sciences team's work into the Data Analytic team's work for the Port Authority of New York/New Jersey (PA NY/NJ). Emphasis was on methods and tools for making sense of and analyzing numerical and categorical data that contain geographic and temporal references, thus on understanding and visualizing the space-time dynamics on the phenomena investigated and on identifying and tracking change over time. Methods developed will be implemented in component-based software that support the mixing and matching of the team's advances with those of the other COE teams working on this topic. Both web-based and desktop tools will be implemented and tested, in support of a range of potential applications.

Technical Approach

The team used bench tests, bioassay, compiling & sorting DB, data mining, expert consultation, field monitors, sampling (tissue), and survey to collection data. They also used Biometrics, case studies, econometrics, genomic testing, modeling, molecular biology/genomics, risk assessment, sampling, and statistical analytics as their analytic methods. Lastly, for the nature of research they used applied, basic, consultation, coordination/integration, education, hybrid basic-applied, and hybrid applied-consultation.

Milestones

Milestones for VACCINE Period 1	07/01/09 – 03/31/10			
	Quarterly			
Initial development of spatiotemporal visual analytic methods for scaling to large datasets		X	X	
Evaluation of spatiotemporal visual analytic methods on health surveillance and crime surveillance data			X	X
Development and integration plan for deployment of tools in PA NY/NJ complete			X	X
Submit paper on design and implementation of web-map tools for crime data analysis		X		
Submit paper on efficient web-map client-server interaction involving large volumes of spatiotemporal incident data			X	
Generalize the system reported on above through a more robust system architecture to support more flexible web-map client-server interaction involving large volumes of spatiotemporal incident data – produce demonstration of this extension				X
Release version of GeoViz Toolkit with new ESDA methods for investigating spatial patterns and events in crime data				X
Work with colleagues at Data Science team at Rutgers to produce cross-site work plan detailing coordination/collaboration & coordinate with potential users				X

The team's current progress has included interacting with the Tippecanoe County Police Department and obtaining workspace for VACCINE researchers to directly interact with officers in the field. This will allow the team to improve technology transition and aid in tool development.

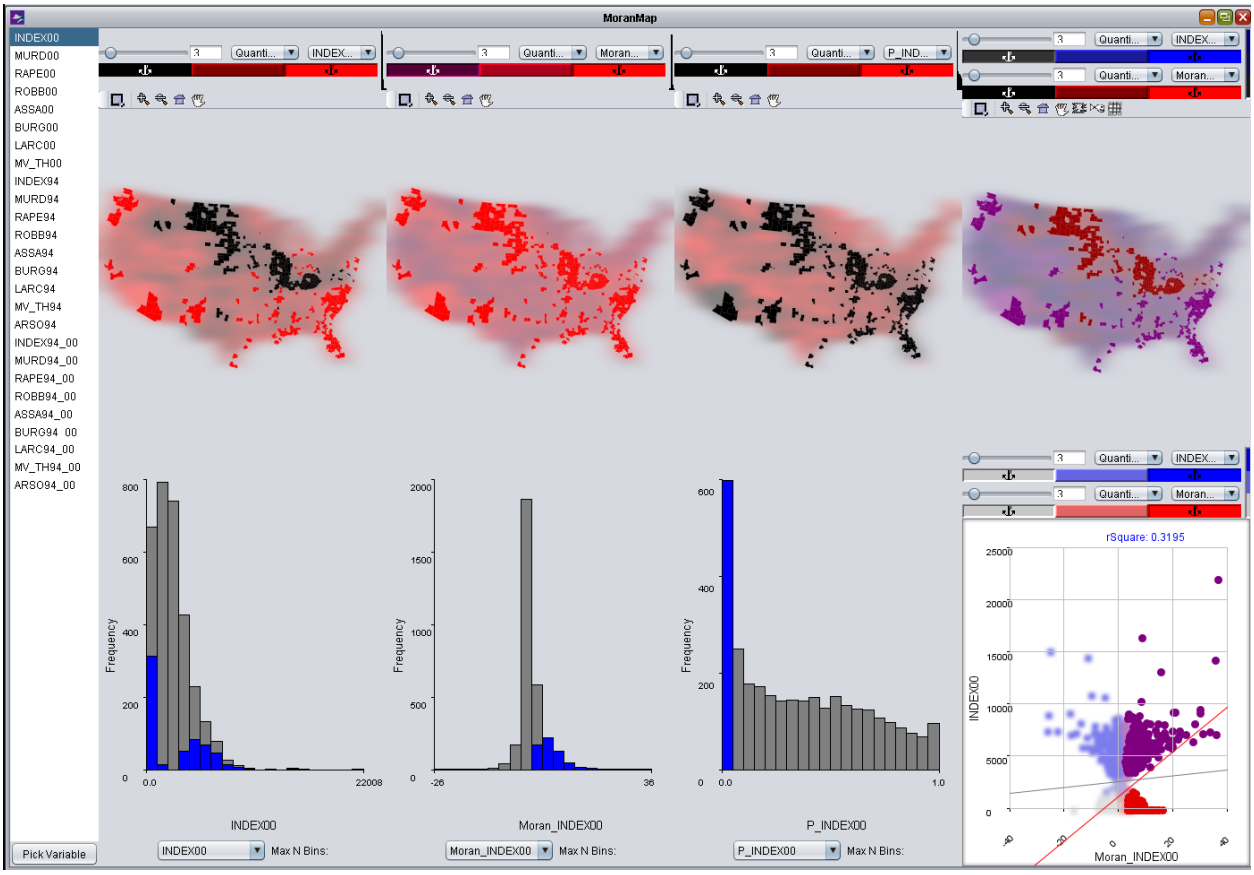
The results for the first year include the release of two software packages, the GeoViz Toolkit, GeoVISTA, CrimeViz, which will enable better understanding of phenomena relevant to crime investigation, and VALET. The team has also implemented a tool that supports scaling of web mapping tools to relatively large event data sets. This tool, HerbariaViz, has been tested with data from a herbaria collection (because the data were available and representative of the kinds of data of interest). Developments with each are outlined briefly.

GeoViz Toolkit

VACCINE researchers at Penn State's GeoVISTA Center developed novel methods for discovering spatial and spatio-temporal patterns in crime and related demographic phenomena using the GeoViz Toolkit as a platform. The GeoViz Toolkit is an open-source set of geographic visualization and analysis components. The advances made for analyzing crime data involved developing methods for interactive exploration of spatial structure in crime and demographic data, and their implementation as interactive tools that take advantage of previously developed components.

In particular, the team developed spatio-temporal versions of the Local Moran's I spatial statistic, and they incorporated that statistic into an interactive tool within the GeoViz Toolkit. They applied these methods to national crime trends, and were able to both identify known

patterns of crime and uncover some unexpected trends that bear further analysis. These ESDA advances were incorporated into a newly released version of the GeoViz Toolkit.



GeoVISTA CrimeViz

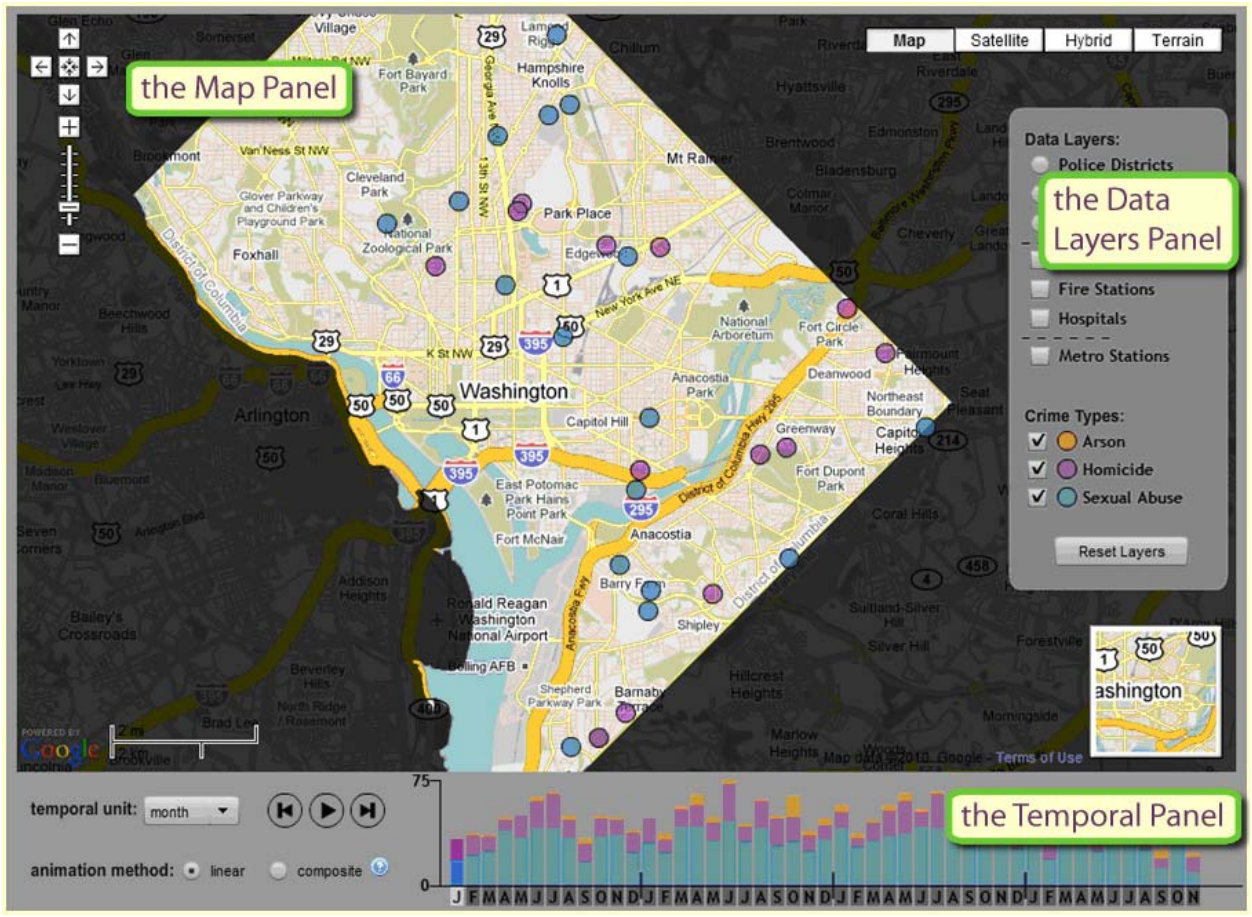
The contribution of CrimeViz is that it is an easy-to-use tool for police officers and other crime analysts to understand both spatial and temporal trends in crime data. This tool is being transitioned into use by the Harrisburg Police Department. The contribution of the GeoViz Toolkit is that it implements a number of multivariate data analysis tools, which will enable better understanding of both crime incidence and the socio-political factors which relate to crime incidence.

CrimeViz is an extensible web-based map application that supports exploration of and sensemaking about criminal activity in space and time.

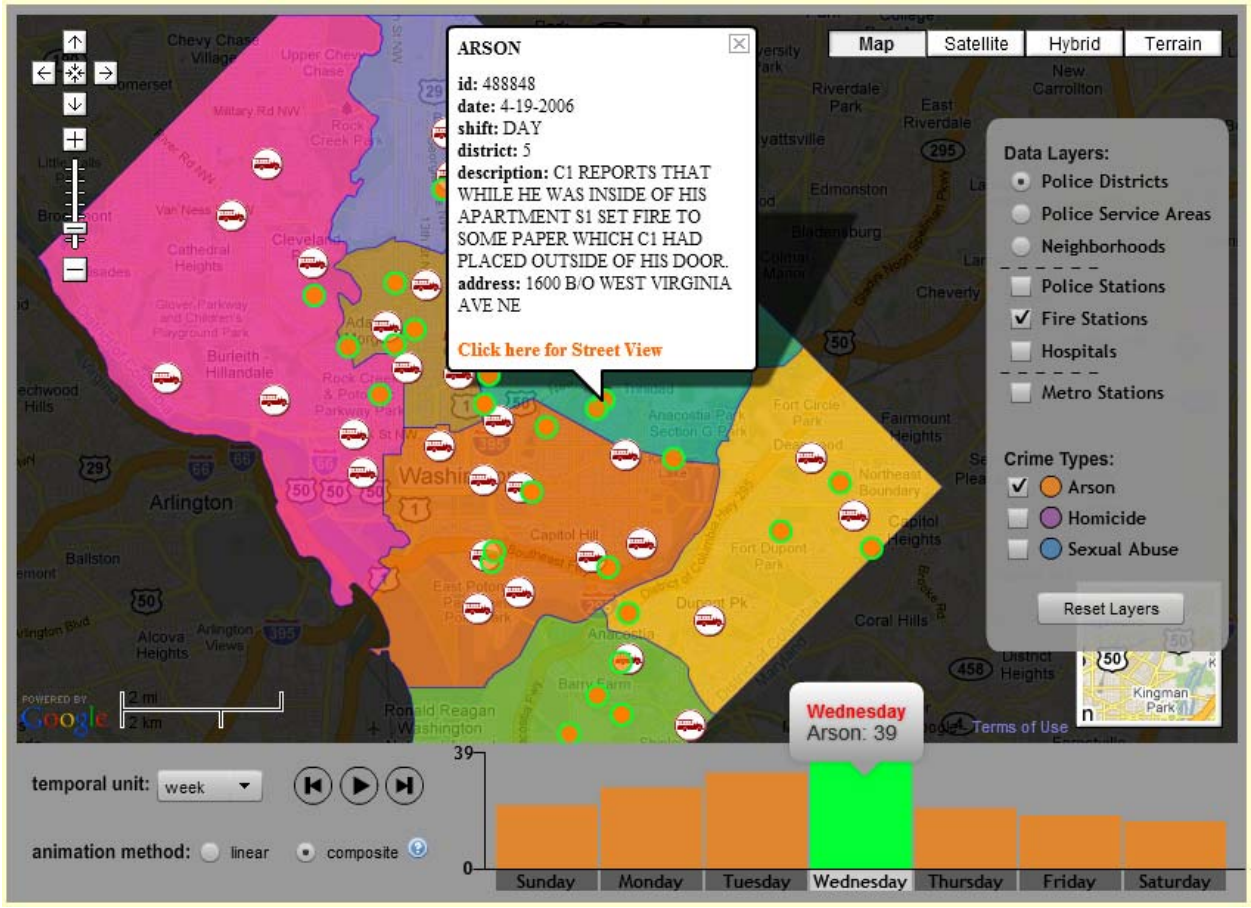
The current prototype illustrates the potential of the tool by visualizing a rich dataset of violent crimes published to the web in near real-time by the District of Columbia Data Catalog (http://data.octo.dc.gov/). Utilizing a suite of geovisual analytics tools, analysts can identify

complex spatio-temporal patterns of crime in the District of Columbia, allowing analysts to unlock important insights about the crime incident dataset.

The current implementation of CrimeViz includes a central interactive map (using the Google Maps service), interactive filtering by crime type, linear and composite animations, an interactive temporal legend that doubles as a frequency histogram, and a set of togglable map layers.



The team is employing a user-centered design approach to guide development of the GeoVISTA CrimeViz concept. An initial usability assessment of version one of the applications revealed interface and mapping problems, as well as other bugs. The results were used to revise the CrimeViz prototype substantially. A typical screen in an analysis session that highlights key features is shown below. Our next step in the process (year two) is to gather more information about the current practice of crime mapping and analysis through interviews, surveys, and hands-on interactive sessions with law enforcement personnel. Feedback from these activities will allow us to identify core features currently available in other tools that must be implemented in GeoVISTA CrimeViz, as well as tasks that current tools do not support.



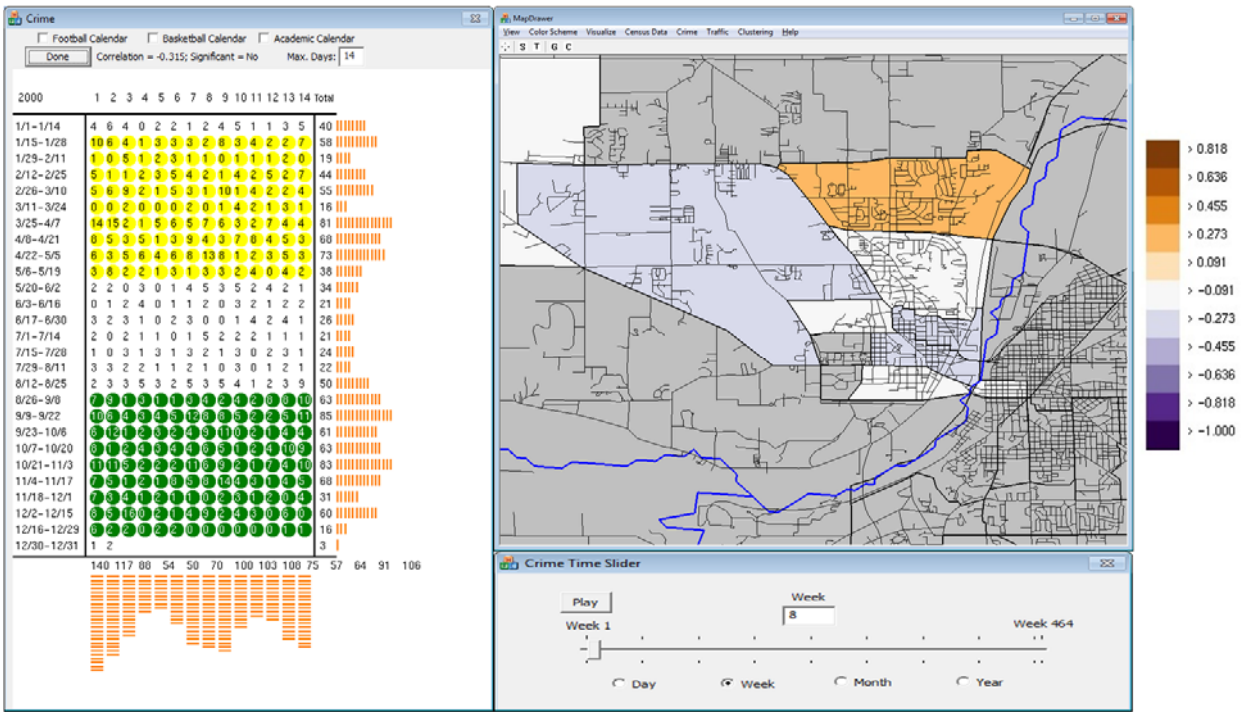
The research thus far indicates that most medium to small municipal police departments lack analytical tools and training to explore and make sense of their crime incident datasets. The GeoVISTA CrimeViz concept provides these crime analysts with an extensible, easy-to-use tool to support spatiotemporal crime analysis and mapping.

As a component of the effort to develop CrimeViz, we have developed a set of methods to support event animation in web-based "map mashups". This has produced a code library that extends the Google Maps API to support well-designed interactive cartographic applications (see the two publications in Cartographic Perspectives below for details).

VALET: Visual Analytics Law Enforcement Technology

We have developed VALET, a Visual Analytics Law Enforcement Toolkit for analyzing spatiotemporal law enforcement data. VALET provides users with a suite of analytical tools coupled with an interactive visual interface for data exploration and analysis. This system includes linked views and interactive displays that spatiotemporally model criminal, traffic and civil (CTC) incidents and allow officials to observe patterns and quickly identify regions with higher probabilities of activity. Our toolkit provides analysts with the ability to visualize different types of data sets (census data, daily weather reports, zoning tracts, prominent calendar dates, etc.) that provide an insight into correlations among CTC incidents and spatial

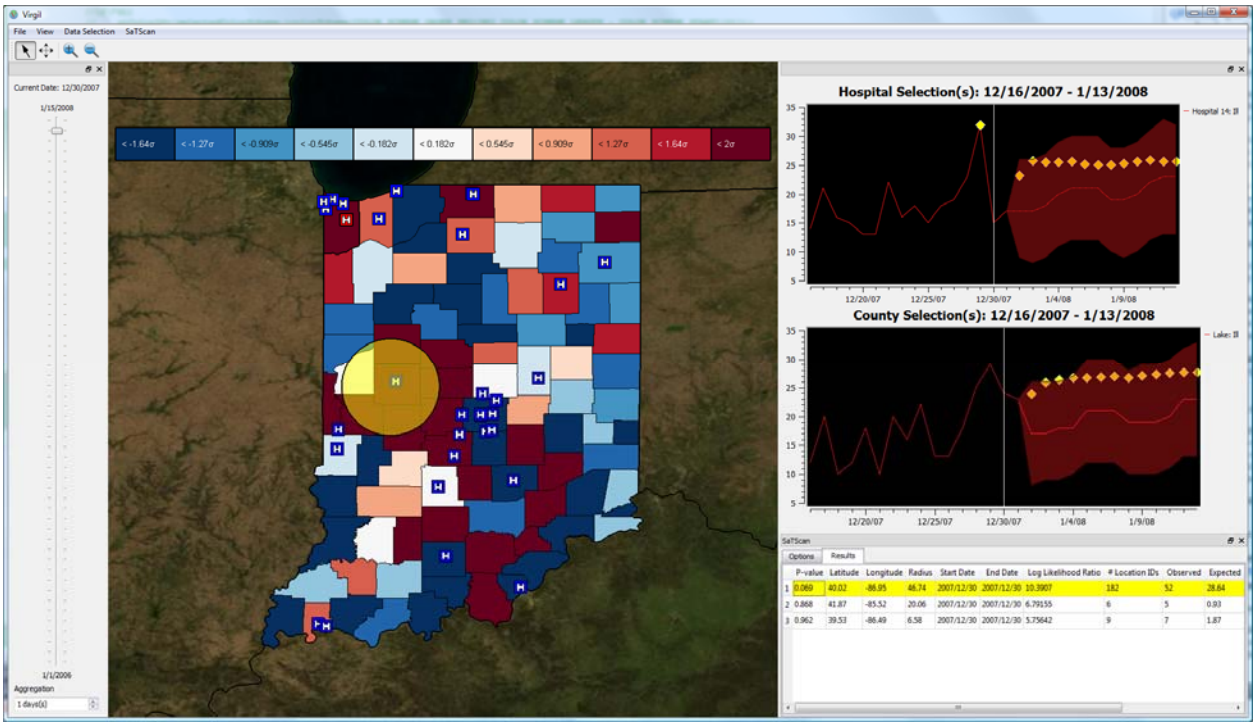
demographics. In the spatial domain, we have implemented a kernel density estimation mapping technique that creates a color map of spatially distributed CTC events that allow analysts to quickly find and identify areas with unusually large activity levels. In the temporal domain, reports can be aggregated by day, week, month or year, allowing the analysts to visualize the CTC activities spatially over a period of time. Coupled with the spatial mapping view, our system provides analysts with a time series view of the data. Furthermore, we have incorporated spatiotemporal prediction algorithms to forecast future CTC incident levels within a 95\% confidence interval. Such predictions aid law enforcement officials in understanding how hotspots may grow in the future in order to judiciously allocate resources and take preventive measures. Our system has been developed using actual law enforcement data and is currently being evaluated and refined by a consortium of law enforcement agencies.



Visual Analytics for Syndromic Surveillance

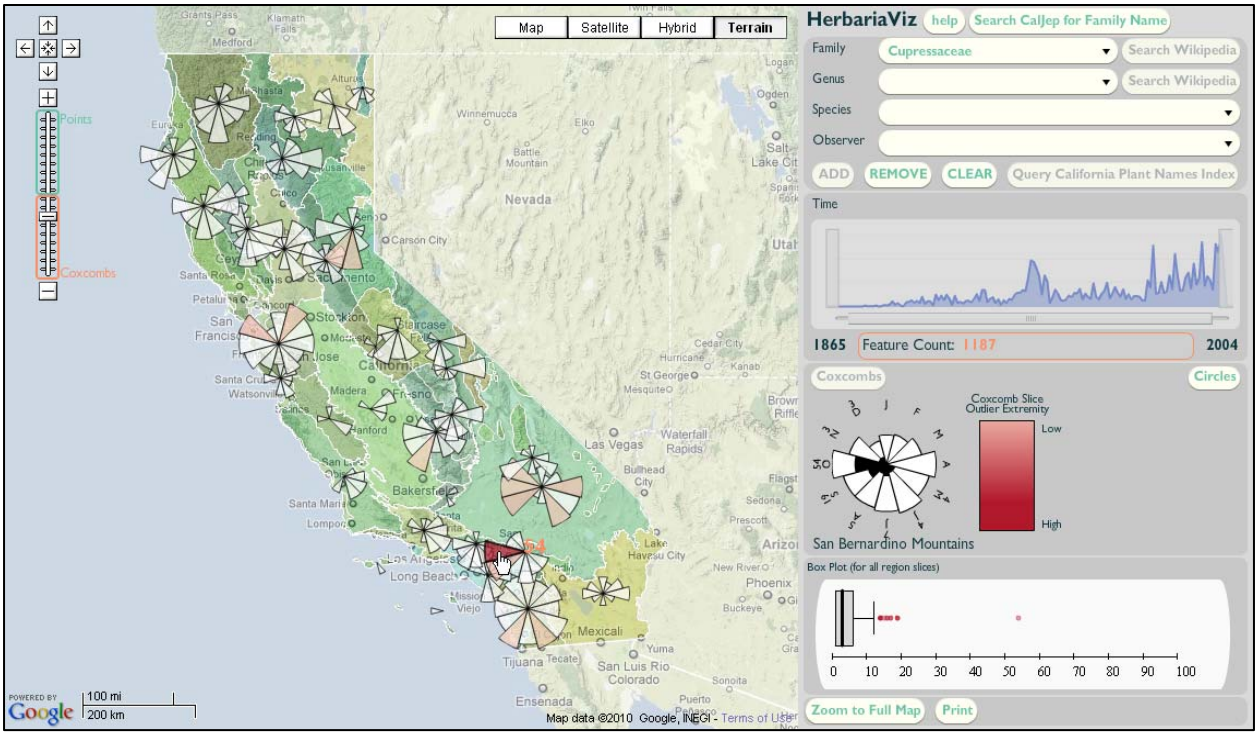
Recently, the detection of adverse health events has focused on pre-diagnosis information to improve response time. This type of detection is more largely termed \emph{syndromic surveillance} and involves the collection and analysis of statistical health trend data, most notably symptoms reported by individuals seeking care in emergency departments. Currently, the Indiana State Department of Health (ISDH) employs a state syndromic surveillance system called PHESS (Public Health Emergency Surveillance System), which receives electronically transmitted patient data (in the form of emergency department chief complaints from 73 hospitals around the state at an average rate of 7000 records per day. To help in analyzing this data, we have created a visual analytics system to provide users with the means to explore trends

in their data. In spatiotemporal data (such as syndromic surveillance data), analysts are searching for regions of space and time with unusually high incidences of events (hotspots). In the cases where hotspots are found, analysts would like to predict how these regions may grow in order to plan resource allocation and preventative measures, and predict where future hotspots may occur. To facilitate such forecasting, we have created a predictive visual analytics toolkit that provides analysts with linked spatiotemporal and statistical analytic views. Our system models spatiotemporal events through the combination of kernel density estimation for event distribution and seasonal trend decomposition by loess smoothing for temporal predictions. We provide analysts with estimates of error in our modeling, along with spatial and temporal alerts to indicate the occurrence of statistically significant hotspots. Spatial data are distributed based on a modeling of previous event locations, thereby maintaining a temporal coherence with past events. Such tools allow analysts to perform real-time hypothesis testing, plan intervention strategies, and allocate resources to correspond to perceived threats.



HerbariaViz

In addition to CrimeViz, we have carried out related research jointly with MDRP 12 to develop a method for efficient web-map client-server interaction and data aggregation involving relatively large volumes of spatiotemporal point data and a symbology and symbol scaling method for representing those data in a web client map application. As a test application, the team used a set of non-crime "incident" data available to us. Specifically, the data consisted of 377,000 spatially-referenced plant samples collected in California between 1860 and 2007. Each record included species information along with the data and place of collection plus most identified the person doing the collection. Thus, while not about crime, these data are analogous to data on crime, terrorism, and similar incidents.



Three research objectives were addressed in the component of the research: (1) develop a method for efficient web-map client-server interaction involving relatively large volumes of spatiotemporal point data, (2) develop a symbology and symbol scaling method for representing those data in the client, and (3) develop an interface for client-server interactions and data exploration. The figure above illustrates the interface to the methods and tools developed and implemented in a tool called *HerbariaViz*.

TOPIC 1: Structural feature extraction of relational networks to support computational epidemiology

It is now well established that social contact networks play a critical role in the progression of epidemics in urban regions. The team analyzed the graphical and relational properties of large, realistic urban social contact networks. These networks are dynamic, labeled and pose new computational challenges. In our study, we focused on two representative synthetic social contact networks -- Miami and Seattle. The research focused on two fundamental issues:

- First, focus on a central question in Network Science – relationship between contact network structure and spread of diseases over these networks. Then, identify new structural measures that provide qualitatively different insights about epidemics.
- Second, we focus on mining interesting subgraphs in dendrograms – a dendrogram is a random subgraph of a social contact network, containing the edges that are used for

disease transmission. Many questions in epidemiological analysis – such as "on average, who is most likely to infect seniors?" or "are there disease transmission pathways involving children?" can be formulated as subgraph mining questions. These questions can be viewed as problems about the occurrence of labeled subgraphs in dendrograms.

There is a lot of work on extracting important features in complex networks. In biological networks, this approach has been used to identify motifs, which may have important structural and functional implications. However, it turns out that because of the stochastic disease transmission models, the above questions cannot be formulated and solved using standard techniques. The subgraphs in question are associated with some probability, which depends not only on the probabilities of the edges in the subgraphs, but also where the subgraph is located, in relation to the initially infected node. The team's contribution is two-fold:

1. Formulating the epidemiological questions in terms of probabilistic subgraph counting problems.
2. Developing efficient computational methods for solving the resulting subgraph counting problems. In general, counting fixed sized subgraphs in a given graph is a well-known "hard problem" (formally speaking it is #P-hard). Thus the work focused on developing fast heuristics: these heuristics have controlled trade-offs between solution quality and running time.

Reasoning about interesting relational patterns can lead to important insights about the spread of epidemics, and these insights can be used for developing novel intervention strategies and public health policies. This forms the basis of the ongoing and year 2 work.

The resulting solutions can also be used as a part of visual analytic tools for public health policy planners and researchers. For example: i) the most frequent disease transmission pathways can be overlaid on a geographic map when the coordinates of each person is known and such information can help the policy planners; ii) graphical representations (e.g., a bar chart) of the counts of the various important (and frequent) subgraph patterns in the dendrogram can help intervention designers.

TOPIC 2: Visual Analytics for Epidemiology

VACCINE developed an automated process for generating movies of the geographic spread of contagion over time through a network of interacting agent. The movies show the prevalence of the contagion over time in small blocks covering the region of interest. Currently, the data can be based on where the infected agents live, or where the agents contracted the contagion, and can show either new infections, the agents who are currently infected, or the total infections since the beginning of the epidemic, although any geo-location tied to the agent can be used. While currently aimed at epidemics, the system is general and can be used for any geo-located, time

varying data. The movies can either show a single dataset, or 2 or 4 datasets, which are synchronized in time.

This tool is useful in several different ways. First, as a model check, it is useful for developers and modelers to be able to visualize the geographic component of a models output. It serves as a basic check that the software is acting as expected, and provides a starting point for tracking down mistakes in either the software itself, or the model's specification. Second, it provides a useful way to compare the outcomes of different interventions on the spatial spread of contagion, which are not apparent in plots of prevalence by day. For instance, it is easy to see the difference between a high-localized prevalence and a lower, more geographically distributed prevalence, even though the total prevalence in the population may be the same.

Project Outcomes (Results, Accomplishments, Conclusions)

Publications

1. Auer, T., MacEachren, A.M., McCabe, C. and Pezanowski, S. submitted: HerbariaViz: A web-based client-server interface for mapping and exploring flora observation data. Ecological Informatics. {joint work with MDRP 12}
2. Chen, J., MacEachren, A.M. and Peuquet, D. 2009: Constructing overview + detail dendrogram-matrix views. IEEE Transactions on Visualization and Computer Graphics 15, 889-896. {joint work with MDRP 12}
3. Hardisty, F. and A. Robinson (Accepted). "The GeoViz Toolkit: Using component-oriented coordination methods to aid geovisualization application construction." International Journal of Geographic Information Science. {joint work with MDRP 12}
4. Klippel, A., F. Hardisty and C. Weaver (2009). "Colour Enhanced Star Plot Glyphs – Can Salient Shape Characteristics be Overcome?" Cartographica 44(3): 217-231.
5. MacEachren, A.M. and Pezanowski, S. (2009). "Geovisualization: Leveraging the Opportunities of Geographic Information." Adobe Developer Connection Education Developer Center: Articles from educators: Adobe: http://www.adobe.com/devnet/edu/articles/macEachren_pezanowski.html.
6. Maciejewski, Ross, Rudolph, Stephen, Hafen, Ryan, Abusalah, Ahmad, Yakout, Mohamed, Ouzzani, Mourad, Cleveland, William S., Grannis, Shaun J., Ebert, David S. A Visual Analytics Approach to Understanding Spatiotemporal Hotspots. IEEE Transactions on Visualization and Computer Graphics, 16(2): 205-220, March/April 2010.
7. Roth RE, and Ross KS 2009 Extending the Google Maps API for Event Animation Mashups, Cartographic Perspectives, Special Digital Issue, No. 64, Fall 2009, 21-31.
8. Roth RE, and Ross KS 2009 Extending the Google Maps API for Event Animation Mashups: Tutorial. Cartographic Perspectives, Special Digital Issue 32-40

Other outcomes/impacts:

Using public health epidemiology as a guiding example, the team had developed novel computational methods for supporting visual and data analytics of large-scale socio-technical systems. A key feature of our work is its generality – indeed the methods are not just restricted to

public health epidemiology and social contact networks, but generalizes well to other diffusion processes and other classes of socio-technical networks.

Virginia Polytechnic University participated in collaborations with other Homeland Security Centers of Excellence. They have initiated collaborations with Dr. Jim Koopman who is a member of the CAMRA team and Dr. Joshua Epstein who is a member of the PACER team. Both of these collaborations have focused on developing agent-based models to study infectious disease spread. Dr. Koopman's interest is in fomite-based diseases while Dr. Epstein is interested in studying the socio-behavioral adaptations during an epidemic.

In addition, the team has recently initiated collaborations with researchers at Chalmers University (Prof. Devdatt Dubhashi) in Sweden to develop synthetic populations and networks for Sweden to support public health epidemiology. They have submitted a proposal to MSB in Sweden that has recently signed a MOU with DHS in the US – they believe that the proposal takes a step towards establishing important research relationships that were called out in the MOU. If funded, this interaction will provide VACCINE researchers the opportunity to interact with researchers in Sweden and Chalmers in particular on problems of interest to DHS.

Pennsylvania State University collaborated with other Homeland Security Centers of Excellence as well. They created a web-delivered version of the START Global Terrorism Database (GTD) with geographic data visualization and analysis tools built in; this version saw use in college courses covering Homeland Security.

1. The CrimeViz tools are described in a project web site, which include a link to a video on the home page: http://www.geovista.psu.edu/CrimeViz/

2. The CrimeViz application is online at: http://www.geovista.psu.edu/DCcrimeViz/app/

3. The HerbariaViz application is online at: http://www.geovista.psu.edu/herbaria/v3/

4. A HerbariaViz video is available at: http://www.youtube.com/user/GeoVISTACenter#p/u/16/ok3Ha2j_Pmk

5. The web-map map mashup methods developed are available as open source tools at: http://code.google.com/apis/maps/documentation/flash/reference.html.
6. The GeoViz Toolkit is an open source suite of tools that are available from the GeoVISTA Center (at no cost): http://www.geovista.psu.edu/grants/cdcesda/software/. Those interested in accessing the full code for their own development work can do so at: http://code.google.com/p/geoviz/

VACCINE activities resulted in a suite of innovative methods for handling spatial data and extracting information from them, all implemented as interactive tools that can be accessed over the internet. These included new exploratory spatial data analysis (ESDA) methods in the GeoViz Toolkit, the CrimeViz tool which allows easy geographic visualization of crime

trends and is being transitioned into use by the Harrisburg Police Department, and the HerberiaVIZ tools for spatial data aggregation.

MDRP 5: Cybersecurity Visual Analytics

Team

VACCINE:
Purdue University: Bill Cleveland
Stanford University: Pat Hanrahan, John Gerth

Abstract

While the frontline of cybersecurity is likely to always be computer-driven -- leveraging pattern recognition, machine learning, and other highly automated techniques, these approaches are not sufficient. The adversaries in network security are intelligent and highly-motivated humans who continuously evolve new attack techniques to avoid any fixed defense. Network security analysts play a key role in proactively identifying and designing responses to new attacks or exposed weaknesses. Here, engaging the powerful pattern recognition abilities of the human visual system is essential in the analysis of the massive amounts of data involved. Thus, a focus of this project has been on developing tools for the frontline analysts. However, the complexity and scope of cybersecurity analytics requires the development of tools which move beyond simply visualizing the network traffic data. Successful tools must also provide a framework for developing higher-level semantic models to provide a simplifying level of abstraction, enabling greater analytic sophistication. To that end, VACCINE is coupling our tools to insights gained from basic research on network traffic data elements.

Technical Approach

The technical approach used was compiling & sorting DB and field monitors for the data collection. Case studies, modeling, and statistical analysis for analytic methods. For the nature of research, hybrid basic-applied and hybrid applied-consultation were used.

Milestones

Milestones for VACCINE Period 1	**07/01/09 – 03/31/10**			
	Quarterly			
Report on current network traffic modeling techniques		X		
Visual interface for display of traffic models			X	X
Toolkit for semantic data integration			X	X
Demonstration of prototype supporting visual specification and evaluation of models				X
Plan and milestones for next phase of project				X

VACCINE is currently engaged in deploying one of their earlier visual tools, *Isis*, as part of the daily workflow for analysts at US-CERT Operations in Arlington, VA (major support for this work comes via a separate grant from DHS S+T in collaboration with Pacific Northwest National Laboratory). During this funding year, the team has created user-documentation and conducted preliminary training exercises with analysts. In addition, they are providing design

advice, evaluation criteria and support to US-CERT in the acquisition of an Enhanced Analytic Database Capability (EADC), which will enable the high-speed data retrieval for interactive use of the tools over dozens of terabytes of data.

As part of the evaluation for the EADC, the team has written code for *Isis* to enable it to connect with 4 additional database systems including a new 'sqlsilk' capability being developed for US-CERT data sources by the SEI at Carnegie Mellon. The team has been active participants in creating the specifications and schema for all the EADC database systems. The vendor testing is scheduled to run from May-July 2010 with acquisition and installation projected for 4Q 2010.

Project Outcomes (Results, Accomplishments, Conclusions)

Publications

1. (submitted to RAID 2010) "Detection of SSH Keystroke Packets in TCP Connections"
2. (submitted to InfoVis 2010) "An Extension of Wilkinson's Algorithm for Positioning Tick Labels on Axes"
3. (submitted to UIST 2010) "Interactive End-user Data Integration for a Geospatial Web"

Other outcomes/impacts

There are two major results for this year. First, VACCINE has made substantial progress in deploying one of their tools, *Isis,* at US-CERT Operations by convincing the group to acquire the necessary database infrastructure for interactive visual analytics (details below). The team has acquired independent funding for much of this transfer work, and hopes to have it completed within another year's time. If this effort is successful, they hope not only to have transferred technology to a DHS frontline unit, but also, and perhaps more importantly, to have learned what is involved in making such a transition from research to operations.

The second result is the first fruit of our effort to understand the fundamentals of network traffic. The team has been able to use visual analytics and statistics to develop a principled understanding of SSH traffic. From that understanding, they have been able to develop an algorithm, which identifies packets carrying SSH keystrokes even though all SSH traffic is encrypted. This algorithm operates on the packet dynamics of *any* TCP connection. It has been constructed in a streaming fashion in order to make it practical for use on high-speed links.

Perceived data sensitivity causes major delays in data acquisition and quality as data must often be anonymized before sharing and the obfuscation process is problematic as it can easily destroy information critical to analysis. DHS is attempting to address this issue through it's PREDICT data initiative which may be available in 2011. However, the team has made arrangements through local institutions to collect data for this project that enables it to proceed relatively unhindered while the larger questions are being worked out.

MDRP 6: Video Surveillance Visual Analytics

Team

VACCINE:
Purdue University: Edward Delp
University of North Carolina at Charlotte: Jianping Fan, Bill Ribarsky

Abstract

The goal of this project is to automatically determine the threat behavior exhibited by individuals or groups of people by observing them via surveillance video. We will accomplish this by extending some of our previous work in real-time object tracking. We will first track the individual(s) and observe how they move in the scene. For example are they walking in patterns such as circling a particular position? Are they trying to minimize their cross-sectional area by squatting? Are they turning their heads a lot to see if someone is looking at them? We will catalog a group of these types of behavior patterns to be used to analyze the video. One other issue that needs to be addressed is the determination of the quality and type of the video surveillance system that is required to be able to do this type of analysis

Technical Approach
This research employs analytic methods, including modeling, sampling, and statistical analysis. The nature of the research is hybrid basic-applied.

Milestones

Milestones for VACCINE Period 1	**07/01/09 – 03/31/10**			
	Quarterly			
Acquire video sequences and complete one threat behavior		X		
Evaluation of the type/quality of video needed for analysis		X	X	X
Complete analysis on 2-4 threat behaviors			X	X
Integrate threat analysis into other projects			X	X

Privacy has been recognized as an important topic in surveillance videos, and the technological developments in this area are ongoing. Unfortunately, most existing work on privacy protection for the visual information mainly focus on the individuals (individual video objects) in the videos and treating them independently. It has become more evident that people may use social network or object co-occurrence associations to infer the privacy of the privacy-sensitive video objects even if they are blocked from the videos. The objective of this project is to contribute to the discussions about multiple types of privacy disclosures for visual information and providing solutions to contest the context-based inference when the privacy-sensitive individuals are blocked or hidden in the surveillance videos. Contesting the context-based privacy inference is a topic to be severely under-researched.

In order to contest the context-driven privacy inference, we first partition the privacy disclosures for visual information (i.e., surveillance videos) into two types: (1) *identification of individuals in videos and disclosing their privacy*; (2) *integration of social contexts or other contexts to identify the privacy (i.e., hidden information)* of individuals (i.e., video objects) even if they are blocked in the surveillance videos. There are some existing works to handle the first type of visual privacy disclosure by blocking the privacy-sensitive video objects, but there is no existing work to deal with the second type of privacy disclosure (i.e., context-driven privacy inference). This privacy dependency (i.e., context-based privacy inference) is a problem that may seriously affect the effectiveness of our systems for visual privacy protection.

To protect the privacy of the individuals in a surveillance video, the given surveillance video is first partitioned into multiple objects and background. Obviously, we should realize that supporting such the object-based video content representation is not a trivial task for current computer vision technologies. After the video objects are detected automatically, privacy sensitivity of these video objects is usually analyzed, and the video objects are further partitioned into two groups: *privacy-sensitive objects* versus *privacy-insensitive objects*. Usually, the privacy-sensitive video objects are blocked from the surveillance video and the privacy-insensitive objects are remained in the surveillance video. The reasons for blocking only the privacy-sensitive video objects is that people believe releasing the privacy-insensitive video objects may not induce privacy disclosure. However, this is not true at all. Because of the associations (i.e., co-occurrence associations and social contexts) between the video objects, releasing the privacy-insensitive video objects may allow dishonest users or public to integrate the object co-occurrence associations and other information sources to infer the privacy-sensitive objects in the surveillance video even they have been blocked.

Based on these observations, the research focuses on: (a) after the video objects are detected from the surveillance videos, we further identify their associations which may induce privacy disclosure; (b) we generate the object co-occurrence associations (which may induce privacy disclosure) by using some information sources which are publically available; (c) developing new video transformation or blocking algorithms for contesting the privacy disclosure induced by such the context-based privacy inference as shown in Fig. 1.

Many objects may appear in the surveillance videos simultaneously, but their importance may be different significantly and people may pay different efforts to protect their privacy. Blocking only the privacy-sensitive video objects from the surveillance videos may be insufficient to protect their privacy effectively because their associations with the privacy-insensitive video objects (which are released in the videos) may be exploited for inferring their privacy.

VACCINE's research took account of the association-based privacy inference to: (a) partition the given surveillance video into multiple objects; (b) analyze the co-occurrence associations between the objects by using other information sources such as company organization structure, parent-children relationships, et al.; (c) such the object co-occurrence associations will be used to analyze the privacy sensitivity of other residue video objects (except the privacy-sensitive video

objects) in the surveillance video as shown in Fig. 1, and the privacy-insensitive objects which have strong co-occurrence associations with the privacy-sensitive video objects (which have been blocked from the given video) may also become privacy-sensitive and will also be blocked from the given surveillance video. Therefore, the new definition of visual privacy depends on two issues: (a) privacy of individual video objects; and (b) the strength of their co-occurrence associations with the privacy-sensitive video objects.

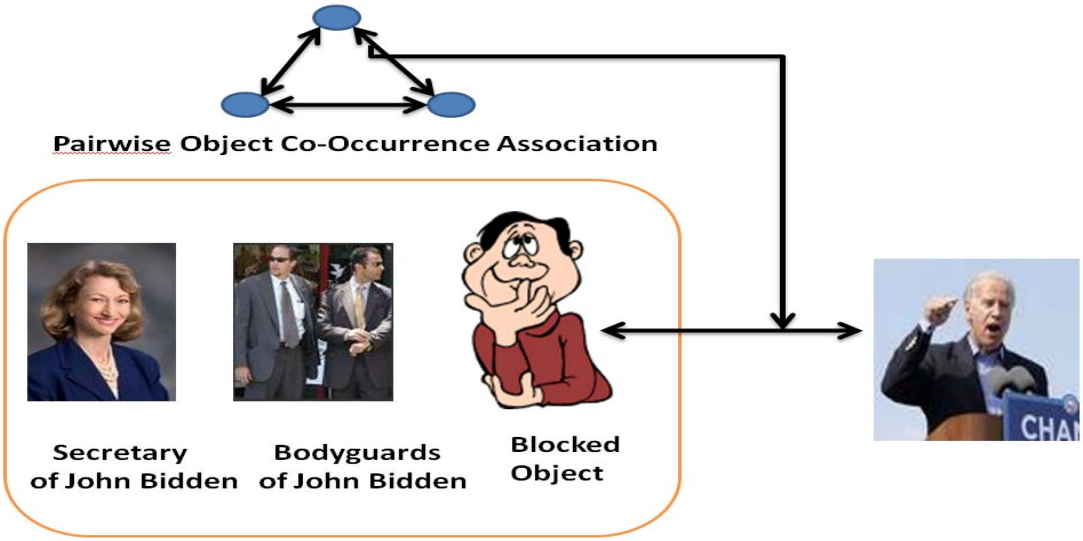


Figure 1: Privacy-insensitive video objects may become privacy-sensitive when their frequent co-occurrence objects are privacy-sensitive.

The team treated the issue (i.e., contesting the privacy disclosure while maintaining the utility of surveillance videos in certain levels) as an optimization problem by solving the following equation:

$$\max\{U(video)\}$$

$$subject\ to\ R(video) < T$$

Where U(video) is the utility of surveillance video after the privacy-sensitive videos are blocked, R(video) is the privacy disclosure risk when the surveillance video with current level of privacy protection is released, T is an acceptable privacy disclosure threshold for a given task.

In a certain domain of surveillance videos, we define the utility of surveillance video U(video) as:

$$U(video) = \frac{I(object - remain)}{I(object - tatal)}$$

Where I(object-remain) is used to define the number of intended objects which remain in the surveillance video, I(object-total) is used to represent the total number of objects of interest in the surveillance video.

The privacy disclosure risk R(video) consists of three types: (a) identifying the individuals directly from the surveillance video; (b) identifying the individuals by integrating the context networks and the blocked video; (c) predicting the appearance of the individuals with certain accuracy bound by integrating the context networks and the blocked video. Thus, the privacy disclosure risk R(video) is defined as:

$$R(video) = a.R(individual) + b.R(context) + c.R(bound)$$

Where a+b+c = 1, R(individual) is used to represent the privacy disclosure because of identifying the individuals directly, R(context) is used to define the privacy disclosure because of the integration of context networks and the blocked video, R(bound) is the privacy disclosure because of the prediction of the appearance of the privacy-sensitive objects in a surveillance video with certain accuracy bound. R(individual) is defined as:

$$R(indivisual) = C, if\ individual\ is\ identfied;$$

$$R(individul) = 0, otherwise$$

R(context) is defined as:

$$R(context) = \sum_{i=1}^{N} \frac{1}{\log\left(1000(X_i - Z_i)\right)}$$

Where X_i and Z_i are used to represent the privacy-sensitive objects and the privacy-insensitive objects in the surveillance video, and |X – Z| is used to characterize the diffirence between the original video object X and the predicted object Z.

R(bound) is defined as:

$$R(bound) = \frac{1}{\log\left(1000\ bound\right)}$$

Where the bound is used to representation the accuracy bound for predicting the appearance of the privacy-sensitive video objects in the surveillance video. Thus privacy protection for the surveillance videos is finally transformed into an optimization problem.

Human-Intended Object Detection

In this project, the team has developed multiple innovative approaches for intended video object detection for video privacy protection application: (a) integrating camera metadata for intended object detection; (b) integrating human attention model for intended object detection; (c) integrating camera motion and object-camera motion coordination for intended object detection;

and (d) integrating the specific task for intended object detection because the definition of the intended objects largely depends on the given task.

As shown in Fig. 2, digital videos taken by human beings significantly differ from the videos that are taken by a surveillance camera or a vision sensor on a robot, e.g., human beings may intentionally capture videos to express his/her feeling or record a memorial video event. Such a creative video capture process is accomplished by adjusting two factors: (1) the parameters setting of a camera; and (2) the position between the camera and the objects of interest or events of interest. To enable automatic understanding and interpretation of the semantics of videos, it is very important to take all these factors into account. Unfortunately, most existing algorithms for video processing and understanding focus on only the video content itself while completely ignoring these two important factors for video capturing.

Photographers may follow the basic rules of esthetic sentiment that shared among the majority of civilized human beings to make the videos attractive, beautiful or even astonishing in all circumstances even in tragedy. For example, people often save certain amount of space for the sky in outdoor videos to make it look balanced. The interesting subjects are often placed in the middle of the picture and keep it integrated. If environmental luminance is not sufficient to make a clear picture, photographers may use a flash and they may make sure that the interesting subjects take proper exposure while the surroundings are neglectable.

The human intention is buried under the pixels we can see from videos and the camera metadata we can read from head files. In order to enable automatic video understanding (i.e., intended object detection), it is very important to develop new frameworks by integrating the visual content of the videos and the camera metadata. Unfortunately, most existing techniques for video processing and understanding have completely ignored the camera metadata while focusing on only the visual content of the videos.

From the above interpretations, one can observe two important issues: (1) not all pixels are born equal and some of them are more important than the others in a video; and (2) the camera metadata can indicate the photographer's intention which may help us detect the intended objects of the videos easily.

The intended objects (i.e., objects of interest) can be defined as a single semantic object or a group of semantic objects in the videos. Such the intended objects can catch the viewers' interest and represents the semantics of the videos, which is critical for video privacy protection while maintaining the utility of videos at certain levels (i.e., ideally the intended objects are most important for the video). Obviously, such the intended objects in a video can provide an alternative way to interpret the semantics of the video more compactly. The supporting intended object detection becomes very attractive and important for both automatic video understanding and video privacy protection.

Based on these observations, the team has done the following research: (a) developing a novel framework by integrating camera metadata and statistical analysis of video content to detect the intended objects more precisely; (b) develop a new video privacy protection framework which is able to maintain the utility of video at certain levels while protecting the video privacy effectively.

Camera metadata, including exposure time, flash, object distance and focal length extracted from EXIF format (\url{http://www.exif.org/Exif2-2.PDF}), as well as the low-level cues. A typical scenario of video/photo capture is illustrated in Fig. 2. A butterfly, which is the object of interest to the photographer, appears in the right season at the right time in the right place (it's unlikely to appear in winter on the snow at night). The photographer wears camouflage so that he can approach the object to achieve a better conformation. The parameter of the digital camera is set and the photo is captured intentionally. The camera records: (a) the objects and the scene in pixels that we can see; and (b) the underlying settings and the environment condition (such as GPS value and luminance etc.) in the metadata that can be read from EXIF files. The final photo intensively blurs the foreground and the background by setting larger aperture so that the interesting object can vividly stand out.

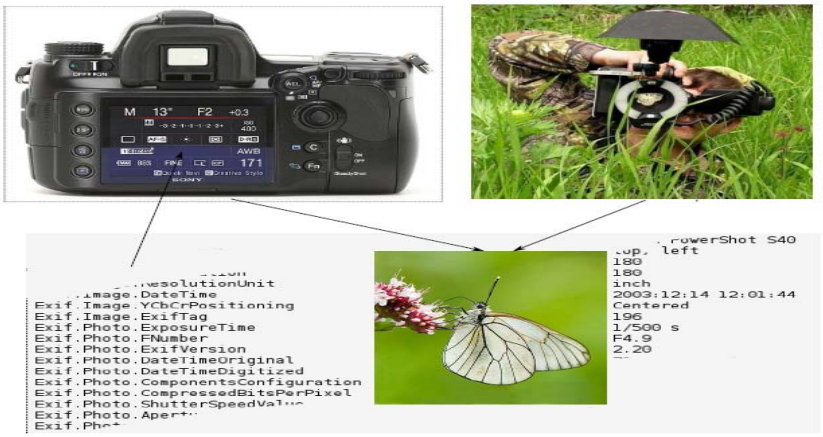


Figure 2: The scenario for video/photo capturing.

Through controlling the camera metadata, photographers can intensively capture the objects or the scenes into photos/videos, which he/she wants to share with others.

Therefore, the camera metadata can somehow reflect the intentions of the photographers and the focus of the photos. Based on these understandings, camera metadata may play an important role in automatic detection of intended objects in a video.

In this project, the team has developed a new approach to explore the camera metadata for automatic detection of the intended objects in videos. They have obtained that the following metadata are most helpful for intended object detection: the exposure value (EV), object distance and date/time. Although the GPS can be recorded in the EXIF file and it can provide the information where the video is taken (which is also critical for video privacy protection because

location is a privacy-sensitive issue). Because the object distance parameter is not available for camera metadata in currently available models, the team will focus on exploiting the EV value for intended object detection.

Many digital cameras nowadays can select different mode of focal points such as central, face, or dynamically set anywhere using the touch screen or a joystick like device. With this information, one can know where the intended objects are in the video. Based on the assumption that the objects in videos have temporal and spatial structures, data/time can help to predict the objects that can only be seen at a certain time period. For example, the sunset and night scene can be seen at the afternoon and night.

The exposure value (EV) that is available in most of the camera models on the market today can be useful for intended object detection, such that the exposure value may be directly provided in the metadata or can be calculated from the exposure time *t* and the aperture f-number *N*:

$$EV = \log \frac{N^2}{t}$$

This value can be set with exposure prior or aperture prior with the other one automatically adjusted by the user or both set automatically in certain mode. The EV value is a function of luminance:

$$EV = log \frac{L \times S}{K}$$

Where $\boldsymbol{L \times S}$ is the luminance, $\boldsymbol{S}$ is the ISO speed and $\boldsymbol{K}$ is the reflected-light meter calibration both $\boldsymbol{S}$ and $\boldsymbol{K}$ can be treated as constant. Given the EV, the team can have more accurate measurement of the luminance. The luminance is closely related with the scene and its value is set according to the scene for a traditional camera.

Based on these observations, our approach for intended object detection and video understanding is straight forward: (a) given the segmentation of a frame of a video, the team will first integrate the visual features of the image and the camera metadata to train the classifier for detecting the regular objects and intended objects in an image; (b) the detected objects and intended objects will be tracked among frames.

The algorithm for intended object detection consists of the following key steps: (a) images are first segmented into multiple objects of interest; (b) a number of objects of interest are labeled as the training samples to learn the classifiers for intended object detection; (c) both the visual features of the video objects and the camera metadata are integrated to train a SVM classifier for intended object detection. (d) the detected intended objects are tracked among frames automatically.

Duplicate or Near-Duplicate Image Detection

In this project, the team has developed a multi-step approach for image duplication detection, which can significantly speed up duplicate detection from large-scale image collections: (a) large-scale images are first partitioned into multiple clusters according to their visual similarity contexts, where only some global visual features are extracted and used for image clustering; (b) pairwise image matching is further performed for detecting the duplicate or near-duplicate images by using SIFt feature matching as shown in Fig. 3.

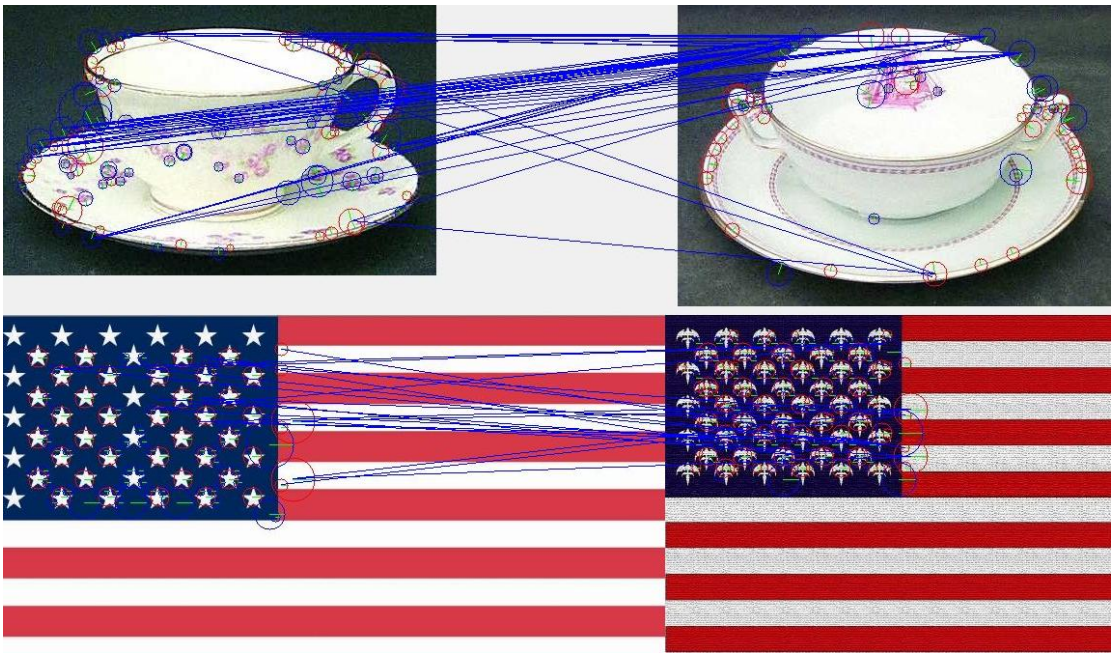


Figure 3: Pairwise image matching for duplicate and near-duplicate image detection.

Project Outcomes (Results, Accomplishments, Conclusions)

Our research for taking account of the association-based privacy inference is to: (a) partition the given surveillance video into multiple objects; (b) analyze the co-occurrence associations between the objects by using other information sources such as company organization structure, parent-children relationships, et al.; (c) the object co-occurrence associations will be used to analyze the privacy sensitivity of other residue video objects (except the privacy-sensitive video objects) in the surveillance video, and the privacy-insensitive objects which have strong co-occurrence associations with the privacy-sensitive video objects (which have been blocked from the given video) may also become privacy-sensitive and will also be blocked from the given surveillance video. Therefore, our new definition of visual privacy depends on two issues: (a) privacy of individual video objects; and (b) the strength of their co-occurrence associations with the privacy-sensitive video objects.

For human-intended object detection, the team has done the following research: (a) developing a novel framework by integrating camera metadata and statistical analysis of video content to

detect the intended objects more precisely; (b) develop a new video privacy protection framework which is able to maintain the utility of video at certain levels while protecting the video privacy effectively.

Publications

1. Ka Ki Ng and E. J. Delp, "New Models For Real-Time Tracking Using Particle Filtering," Proceedings of the SPIE/IS&T Conference on Visual Communications and Image Processing(VCIP), San Jose, CA, January 2009.
2. K. K. Ng and E. J. Delp, "Object Tracking Initialization Using Automatic Moving Object Detection," Proceedings of the SPIE/IS&T Conference on Visual Communications and Image Processing (VCIP), January 2010, San Jose, CA.
3. Z. Li, J. Fan, "Stochastic contour approach for automatic image segmentation", Journal of Electronic Imaging, vol.18, no.4, 2009.
4. Luo, Qingshan, Xiaodong Kong, Guihua Zeng, Jianping Fan: Human action detection via boosted local motion histograms. Mach. Vis. Appl. 21(3): 377-389 (2010).

Conference Publications

We have submitted 3 conference papers that are under review.

MDRP 7: Introducing Sustainable Visual Analytics into Command Center Environments

Team

VACCINE:
University of Washington: Mark Haselkorn and Daniel Daily
Purdue University: David Ebert and Tim Collins

FAZD:
Texas A&M: Jim Wall

Abstract

MDRP7 focuses on the development, fielding and maintenance of next-generation command and control environments. In their current stages of development, DHS command and control system initiatives such as *Virtual USA* and *Watchkeeper* focus primarily on delivering a common operating picture, a common communication platform and, occasionally, analytic support for key decisions. MDRP7 develops strategies and field demonstrations that extend these evolving capabilities by placing them in the context of two additional capabilities: (1) support for the ways the multi-stakeholder security community has chosen to work together (in the case of DHS systems, the NIMS Incident Command System) and (2) the sustainability of these systems based on stakeholder ownership, trust and, most importantly, the ability to achieve continuous improvement through actual use.

The critical strength of NIMS/ICS is that it supports the integrated efforts of a community of diverse stakeholders managing a complex, evolving situation from diverse perspectives involving many different roles, responsibilities, decisions and actions. While NIMS/ICS provides a blueprint, the necessary level of coordination and communication can be difficult to achieve, especially given the many different agencies, organizations and units involved, each with its own mission, jurisdiction, processes, databases, systems and cultures. MDRP7 is working on next-generation CCI systems (what some call "Precision Information Environments") that will provide critical information and communication capabilities within an environment that supports the unique, yet coordinated efforts of the diverse safety and security community.

To support the many perspectives and roles within the security community, VACCINE is describing methods and strategies for developing next-generation CCI systems that enable in-field customization of CCI environments. This allows each user group to further adopt their system perspective and capabilities to their particular roles and responsibilities, as well as to the evolving nature of the situation being managed. To accomplish this, MDRP7 is developing an "adaptive development" strategy that increases stakeholder ownership and trust by giving user groups the capability of guiding and achieving continuous system improvement even after the environment has been fielded. One key result is a more sustainable system that does not have to be replaced as evolving requirements emerge over time.

Technical Approach

The data collection methods used included expert consultation, field monitors, and survey. The analytic methods used were case studies and modeling. The nature of research was coordination/integration and hybrid basic-applied.

1. Demonstration video available at http://citdweb.tamu.edu/dreams/CGDS_Overview.wmv
2. Webinars to various DHS clients and partners (e.g. see attachment III_2)
3. Interim products from the JHOC demonstration (e.g. attachments III_3_a and III_3_B)

Milestones

Milestones for VACCINE Period 1	**07/01/09 – 03/31/10**			
	Quarterly			
Research, document, and assess current processes for development and introduction of technology into command center environments		X	X	X
Recommend changes to current processes that would improve how VA systems: (1) support command center awareness and decision-making; (2) increase the likelihood that VA systems improve over time rather that degrade, and (3) increase the likelihood that new systems are integrated into the current environment and contribute to stakeholder community building rather than create additional barriers			X	
Demonstrate potential VA tools for JHOC operations		X	X	
Define data dissemination needs, processes, and multi-jurisdictional authorities to deploy interactive VA systems in diverse command and control environments			X	X
Design VA system following information from above tasks, producing a design document and initial prototype VA components			X	X

Project Outcomes (Results, Accomplishments, Conclusions)

Publications

1. Benson, A.L., Biggers, K., Wall, J., and Haselkorn, M.P. (2010) "Adaptive Development of a Common Operating Environment for Crisis Response Management." In *Proceedings of the 2010 International Conference on Information Systems for Crisis Response and Management (ISCRAM2010)* [Fully-refereed, blind, 45% acceptance].

Follow-on funding/related projects

Further development will occur under a Phase I Small Business Innovative Research on Precision Information Environments awarded April 2010, Department of Homeland Security Science & Technology, $100,000 (lead Truestone LLC).

Other outcomes / impacts

Teaming with FAZD (Texas A&M) and VACCINE (Purdue), developed a demonstration of a next-generation CCI system. The team delivered that demonstration as part of a regional field exercise involving numerous stakeholder agencies, over 300 personnel, on water assets, and the

regional command center. From this experience, as well as literature analysis and discussions with key government personnel involved in CCI development, the team has drawn a number of conclusions as to how next generation CCI systems can be better designed, developed, fielded, and maintained. Additional details of the results, accomplishments and conclusions are provided below.

Development of the PSICOE

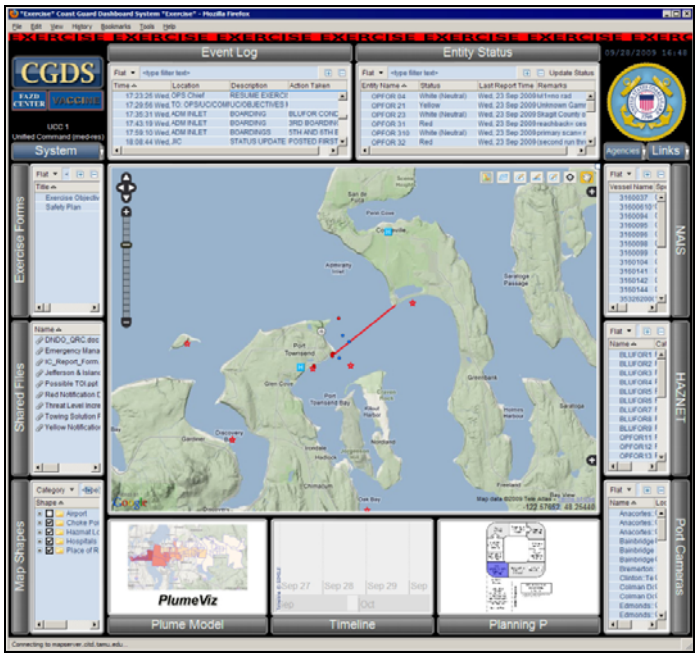
The Puget Sound Interagency Common Operating Environment (PSICOE) was adapted from the Dynamic Preparedness System (DPS—developed by FAZD at the Texas Center for Applied Technology), which is an internet-based, adaptive software framework. Working with FAZD, VACCINE and regional security stakeholders led by the Coast Guard, the team populated the framework with relevant informational components. To do this, they considered the nature of the decisions being supported, the roles of the people making those decisions, and the tradeoff between simplicity of data capture versus integration with the PSICOE environment and other components. For example, personnel in the exercise wanted a component to provide weather information, but this component did not need to be integrated with other information sources. Providing this component in an independent window that opened a local internet weather service was extremely simple. On the other hand, working with Purdue and VACCINE, they provided a Plume model that was fully integrated into the PSICOE interface and GIS map layer. This proved extremely powerful to decision makers, but also involved a great deal more effort than the independent weather component. They also provided pre-set configurations of components and interface settings tailored to various communities of practive, called *profiles*. The four profiles created for the AMSC exercise were: (1) field, (2) planning, (3) unified command and 4) media/public information. The transformation of DPS to PSICOE was accomplished in six weeks, and demonstrated a set of basic information, analytic and communication capabilities for an on-water exercise involving small vessels with potential radioactive threats.

The COE was further tailored to regional conditions through an analysis of key decision making communities and the decisions to be supported during the target exercise. An analysis of the information environment and needs was conducted with the Area Maritime Security Committee (AMSC) and five key decision making questions were identified to drive the development of components of the PSICOE. These five questions/decisions were: (1) Where are the threatening vessels and what is their status? (2) What is the nature of the threat? Is there a dispersal device? (3) What should we do with the identified threat? Is there a place and/or condition (e.g. scuttled) that would lower the threat risk? (4) How are we managing the commercial and pleasure vessel traffic? (5) What are we telling the public? The pre-configured user profiles were used to support the stakeholder roles in answering these questions.

Demonstration of the PSICOE during a Regional Exercise

A regional scenario was designed to exercise the Coast Guard's implementation of NIMS/ICS in the detection of and response to small vessels transporting radioactive material, possibly with a detonation device, in the Puget Sound maritime arena. Three choke points were established at

inlets to the Sound and each was monitored by a branch of field operations units. Blue force units were tracked using a transponder data stream which was displayed on the PSICOE. Participating Coast Guard vessels designated potentially threatening "Red force" units as triangular entities on the GIS display. Commercial shipping traffic was monitored through the NAIS open-source tracking system. These and many other data components from a wide range of disparate sources were accommodated within the COE framework.



A screen capture of the PSICOE Unified Command profile during the AMSC exercise

The PSICOE was an adjunct experiment to the field exercise and not intended to be a primary communication system, nor was it incorporated into the interagency action plan. It was included in the exercise as a demonstration and presentation of the COE, rather than as a fully functional information and communication system. For this reason, participants received minimal training in using the PSICOE prior to the exercise. The university teams trained the Coast Guard participants in how to use the PSICOE for approximately 2 hours, 2 days prior to the exercise. Participants were taught only those features that were anticipated to be of use for the various command roles. For example, training of on-water field participants focused on how to place a shape representing a potentially threatening vessel on the map component, mark that entity as unknown/cleared/threatening, and share that component and information with other selected units. Other training covered selecting a profile, entering event log entries, using the plume model component, broadcasting data to other units, sharing files with other units, and using the resource/entity status component. The entity status component was intended to track vessels and report the current status of that entity.

During the exercise, a university team member was stationed with each of the participating units in the command center. The units in the field were able to request assistance in using the PSICOE through cell phone contact with the university teams. For several of the participating units, such as Unified Command, the university team representative became a dedicated PSICOE operator throughout the exercise. In other units, such as the public media joint information center (JIC), participants operated the PSICOE themselves.

A pivotal moment during the exercise occurred when the covered communications radio channel became unavailable to Coast Guard units on the water. The exercise was nearly cut short; however, the Unified Command decided to continue when field units began improvising by using the PSICOE event log and entity status components as a communications workaround for the remainder of the exercise. The PSICOE abruptly became a focal point of the exercise, because units relied upon it for information and communication.

Because no procedure for using the PSICOE communication components had been established prior to the exercise, the field units began improvising ways to use the PSICOE as their primary communications channel. Two field units used the event log to report which vessels they had boarded and what the subsequent vessel status was. The other field unit reported this information using the entity status component. Interestingly, without specific training and using the flexible COE with different strategies, the units were able to successfully complete the exercise. This demonstrated a proof-of-concept for the COE's flexibility, effectiveness, and ease of use.

Once the simulated nuclear material had been discovered during the exercise, the Unified Command used the PSICOE to brainstorm and visualize possible courses of action to mitigate the threat. The plume model became a primary visual analytics tool for this effort. The UC used the PSICOE as a decision-support tool while devising plans which would minimize projected deaths and illnesses if the nuclear device detonated. In this manner, the UC decided on a plan to move the vessel in question into open water, where a crane and barge could be used to raise the vessel for inspection and neutralize the simulated radioactive threat.

While creative and successful, this operation of the Unified Command was outside of regular ICS procedures. In their ad hoc communications solution, field units provided all available information to the UC, whereas under usual conditions the Situation Unit, Planning, and Operations teams gather and sift information before presenting to UC. Under normal ICS conditions, these groups periodically report the situation on the ground and suggest plans for the next operational period to the UC, which is responsible for vetting and eventually approving a suggested plan. As a spur-of-the-moment workaround, PSICOE upset this procedure during the exercise by allowing field units to provide all information to all units within the command center.

Lessons from the Exercise

One lesson learned from the exercise is that an initial set of procedures should be developed jointly between the client and development organizations during the preliminary development

period of the COE. Ideally, the COE should be introduced to the client organization in tandem with training on procedures to use the environment. In addition, a set of guidelines should be introduced at the same time for developing further applications and making interface changes. This is intended to preserve the common operating picture of the system while allowing flexibility for individual user groups to develop the COE to their needs.

Second, introducing the PSICOE independent of the incident action plan upset the established ICS information gathering, processing and sharing procedures. Unintentionally as part of the improvised communications workaround, the PSICOE shared all available information with the Unified Command, circumvented the Situation Unit, Planning and Operations sections to evaluate information and formulate plans. Unintended consequences, such as this disturbance in the procedures of the client organization, should be explored with an eye towards understanding how the power of a next-generation information and communication environment can affect, for good or bad, existing organizational processes and strategies.

This issue of information sharing also illustrated the complexity of the regional information system as a whole. The ICS structure and procedures currently in place are meant to provide the Unified Command with regular information feeds which have been filtered for relevance and distilled by the other ICS units. However, given the access to all information (through an environment such as the PSICOE), the members of the UC, who are generally also the decision making bodies of the Sector and other participating agencies, may not relinquish this complete view in favor of the distilled information picture prescribed by ICS procedures. System developers and planners must work as facilitators to help resolve these kinds of tensions and to align new capabilities with the overall strategy of the organization.

Challenges

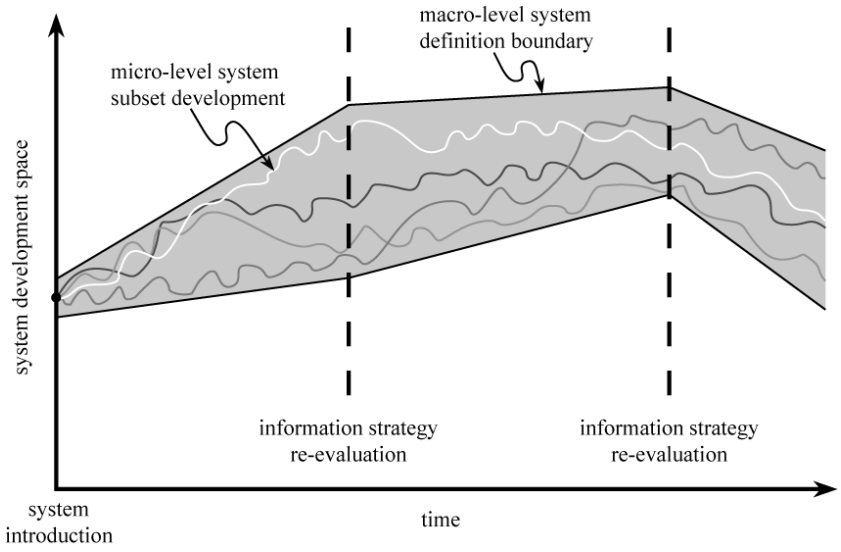
Several prerequisites exist for applying a new, more participatory design approach in a crisis response management organization. First, the organization's leadership must support the system development and foster ownership of the COE within the organization. Success depends upon how much the user groups take ownership of the system and continue to cultivate the COE. The client organizations must lead in guiding the evolution of the system to meet their needs while using the expertise of the development organization to facilitate that progress.

Second, the system development organization and the client organization must establish a long-term, trusted relationship to successfully cultivate the COE. This service business model for technology development is already being adopted in other contexts, such as internet services, but has yet to be fully recognized as a superior method for creating sustainable CCI systems.

Third, to successfully implement a new development paradigm, usability experts, designers and system analysts must relinquish some of their roles as creators and owners of systems and take up the challenge of facilitating design and development of systems owned by users. Such professionals will be required to augment their skill sets to accommodate this change in roles.

Conclusions

VACCINE proposes *adaptive development* (AD) as a unique and superior development paradigm for COEs in the context of mission-driven organizations. AD does not address development methods directly, but instead presents a holistic design and development model based on *ongoing* and *contextual* development of the system. AD maintains a two-level development process driven by the community of practice: *macro-level system definition* and *micro-level system development*. AD is based on two principles of practice: (1) the bulk of system development occurs "live" within the client organization after a base architecture and capability has been introduced to the target environment, and (2) system analysts facilitate user-driven development of the COE application layer according to the client's overall information strategy and constraints.



Adaptive development process model

AD addresses the deficiency in current development processes for creating sustainable COEs. The greatest opportunity AD presents is that it may allow us to create and manage systems which last 10, 15 or more years while becoming more useful over time. This presents clear incentives based on costs of creating new systems from scratch every few years, retraining, mistakes and inefficiencies due to technology turnover.

AD also presents the opportunity to create more useful and successful COEs. Through a two-tiered development process, one level focusing on aligning the COE with the information strategy of the organization and the other level focusing on developing flavors of the system for individual user groups, AD provides a more holistic and comprehensive development process than preceding models.

Interagency collaboration is another area where AD is likely to provide greater benefits than other system development models. This is because AD classifies changes in trusted partners, selective information sharing, and rapid application development as routine systems development. The process's continuous development strategy is geared to accommodate rapid changes in relationship and information sharing within and among organizations.

Further study is needed to develop the concept of AD for mission-driven security communities. The PSICOE project is an ongoing effort which can continue to reveal the challenges and successes of using an AD development approach within the United States Coast Guard and other DHS agencies.

Finally, it should be noted that the AD approach to COEs presented here represents a future direction that builds on the important current situational awareness/information sharing effort known as Virtual USA. Most importantly, the team adds the support of an inter-agency ICS management strategy and a framework for future Precision Information Environments (PIE). The PIE concept, introduced by the DHS Basic/Futures Research thrust area (under Dr. Joe Kielman), will combine an overwhelming number of diverse data sources and advanced visual analytic tools with information tailored to the users' roles, needs and cognitive capacities.

MDRP 9: Mobile Imaging, Rosetta Phone, and Light-Weight Visual Analytics for in-Field Analytics

Team

VACCINE:
Purdue University: Ed Delp, Mimi Boutin

Abstract

The Rosetta Phone is a hand-held translation device for foreign text currently being developed by Purdue in collaboration with Next Wave Systems, LLC. The goal of this device is to enable an individual with little or no prior knowledge of foreign languages to recognize and translate foreign text in real time and without the use of a network connection. This particular project focuses on developing the methodology and software needed to enable the Rosetta Phone to automatically read and translate Spanish documents.

Technical Approach

The data collection method used was data mining. The analytic methods used were case studies and statistical analysis. The nature of research used was hybrid basic-applied.

Milestones

Milestones for VACCINE Period 1	**07/01/09 – 03/31/10**			
	Quarterly			
Identification of representative document type for development and testing purposes		X		
Picture database acquisition		X	X	
Development of reading software		X	X	
Development of translation software		X	X	
Deployment of reading and translation software on portable device				X
Testing and performance evaluation				X

Project Outcomes (Results, Accomplishments, Conclusions)

We have developed and deployed a Spanish language translator on a mobile phone that can read and translate a limited set of Spanish language image documents.

Other outcomes/impacts

The team decided to focus their efforts on Spanish magazines and newspapers.

For development and testing, they acquired 10 pictures of a Spanish magazine and a Spanish newspaper (both from Spain) using a 6 Mpx digital camera (Casio EXZ110). All pictures were acquired indoors and without flash. The articles they chose included single columns and double column articles, as well different font styles, sizes and colors.

Different open source OCR software tools were tested, including OCRpus, Tesseract, OCRad, GOCR. The most accurate one was found to be Tesseract. The next most accurate tool is OCRad. OCRad also has faster speed and lower memory requirements than Tesseract. However, the team decided to use Tesseract, as the accuracy of OCRad was very poor on some of the test pictures.

There are two main classes of machine translation software: rule-based and statistics-based. After surveying the available existing open source translation tools, the team chose to test Apertium, a rule-based translation engine using markov models, as well as Moses, a statistics-based translation software. Both types of software were found to be adequate. But overall, Apertium seems to be best candidate, as it combines lightness, a wide range of topics, and can be used to translate between similar languages and dialects, such as Spanish-Catalan.

Using Tesseract and Apertium, they implemented a command-line-based system to read and translate Spanish articles on the Nokia N900.

Note: A commercial software for translation of text found on pictures for the Nokia N900 was recently released. The software is called Photo Translator and uses a network connection to translate the text. The team installed this software on one of our Nokia N900 and tested it. For the moment, the function “capture a picture” is not enabled. Unfortunately, all the images they tried gave them lecture errors. It thus appears that the software is not quite ready for prime time.

MDRP 11: Financial Fraud Visual Analytics

Team

VACCINE:
University of North Carolina at Charlotte: William Ribarsky, Robert Kosara
Purdue University: Niklas Elmqvist

Abstract

This project's goal is to develop and evaluate new visual analytics tools to explore and discover behaviors (suspicious behavior, emerging risk behavior) in financial transaction data. To uncover suspicious behavior, we are developing a set of tools based on a hierarchical data model for interactive exploration even with millions of transactions over up to a year's time period. This approach uses multiple views of the transactional data to pinpoint suspicious behavior that would not be apparent with just one view. The analyst can then drill down to the particular account or related accounts (having the same keyword signatures) for detailed investigation. Emerging risk analysis is a much broader area that seeks to uncover emerging risk in the bank's transactional, credit card, and loan portfolios that may not be adequately accounted for. The effect of the economic downturn and job losses in specific industries on unsecured loans for specific groups of customers is an example of such emerging risks. Tools such as those used for fraud analysis, geographic analysis tools, and tools to include and correlate multiple dimensions of behaviors and ratings are being developed or applied to this analysis. Emerging risks are important factors for resilience of financial institutions and the financial system.

Technical Approach

This research employs analytic methods, including modeling, sampling, case studies, econometrics, risk assessment, and statistical analysis. The nature of the research is hybrid basic-applied and applied. Data collection methods include compiling & sorting DB, data mining, expert consultation, and field monitors.

Milestones

Milestones for VACCINE Period 1	**07/01/09 – 03/31/10**			
	Quarterly			
Initial discussions with banks and government partners	X			
Initial Prototype of FraudWatch		X		
Presentation and discussion of FraudWatch; further implementation			X	X
Publication on FraudWatch				X

Project Outcomes (Results, Accomplishments, Conclusions)

Publications

1. Javed, W., Elmqvist, N. Stack Zooming for Multi-Focus Interaction in Time-Series Data Visualization. In *Proceedings of the IEEE Pacific Visualization Symposium 2010*, pp. 33-40, 2010.
2. Chang, Remco, Caroline Ziemkiewicz, Roman Pyzh, Joseph Kielman, and William Ribarsky. Learning-based Evaluation of Visual Analytics Systems. ACM BELIV 2010.
3. Chen, Yang, Jing Yang, and William Ribarsky. Toward Effective Insight Management in Visual Analytics Systems. Proc. IEEE Pacific Graphics 2009, pp. 49-56.
4. Dou, Wenwen, William Ribarsky, and Remco Chang. Capturing Reasoning Processes through User Interaction. To be published. IEEE EuroVAST 2010.

Other outcomes/impacts:

Based on VACCINE's previously developed WireVis tools, the team is developing the FraudWatch tools and methods. Although WireVis focused on money laundering in wire transactions, FraudWatch has a larger purview, applying visual analytics to explore large amounts of financial transactional data over a range of types and discover suspicious activity. The tools then permit detailed analysis of the suspicious activity and enable automated search (through the signatures and keywords of the target transactions) for other transactions with similar properties.

Working with Agus Sudjianto, Senior Vice President and head of the analytics groups at Bank of America (BoA), the team has begun a project that involves predictive risk analytics, especially of emerging risks. The BoA approach is heavy on automated methods and statistical techniques, but due to the very large data and complex analyses they realize that humans must be involved in attaching meaning to the results, discovering insights, and drawing conclusions. This is where interactive visualization and visual analytics comes in. In addition to bank transaction data, the team is looking at new data, such as credit card transactions. There will be a period of months where they will jointly develop ideas and methods followed by presentation to senior management. After that the project could get a lot bigger on the bank's side.

This work ties in with the financial analytics work in the DHS project, but it is broader. The VACCINE team is bringing the FraudWatch tools into this arena, where the tools will be augmented for predictive risk analytics. In addition, the team is developing a new set of complementary tools and methods, with BoA help, that will look at credit ratings, amount of secured and unsecured loan activity, types of loans, geographic distribution to bring out the economic climate and health of a region, and other factors to determine detailed, time-dependent risk profiles that are flexible enough to account for unforeseen factors revealed in the data analysis. These factors can then be inserted for more robust predictive risk analysis.

The team should also be able to build off of the knowledge gained and methods developed here in establishing relationships with other banks and financial services companies. Finally, BoA has

developed a rather rigorous curriculum for training new employees from non-financial backgrounds in what they need to know in financial analytics. This can be quite useful input for their educational programs; in particular, they plan to develop a financial informatics and analytics Professional Science Masters that can be a significant part of the VACCINE educational offerings.

As planned and in addition to the team's work with Bank of America, they have had further discussions with Wells Fargo/Wachovia and with TIAA-CREF. They are planning a Financial Analytics Workshop with Agus Sudjianto of Bank of America. This will probably be in Winter, 2011. Invitees will include banks and insurance companies (including BoA, Wells Fargo, and TIAA-CREF), leading university researchers, and significant representatives from government agencies. During the workshop they will establish a research agenda in financial analytics that will correlate with the VACCINE agenda, and will also encompass broader needs in risk analytics for government and the financial industry and customer analytics for the banks. In addition, the workshop will provide an opportunity to present the latest relevant tools and methods from visual analytics and other relevant fields that can serve as a foundation for new research.

The partnership with Bank of America is giving the team access to analysis, customer data, and business approaches that is unprecedented for a university. They are working with the BoA analytics group, which has access to all BoA transactional, corporate, credit card, and other data for its worldwide operations. The analytics group uses a combination of advanced statistical and database techniques. The analytic group also appreciates the value of visualization for the complex reasoning problems they face based on exploration of large scale multivariate data. This is a perfect partnership for the development of visual analytics tools.

They have placed a Ph.D. student as a full-time employee of the bank and a faculty member of the VACCINE team as a consultant. With this arrangement, they are getting access to bank data and analytic tools. Although the Ph.D. student is a bank employee, his R&D activities are supervised by the UNCC team and the head of the BoA analytics group. Thus the team has full access to the tools and analyses the student develops and will use them for the tools developed in this project. As initial results from the project are presented to BoA management, they expect it to grow substantially.

A graduate research assistant has also been hired and is developing simulated financial transaction data. BoA is helping the team on this task, and they have also formed a partnership with Mark Whiting and his group at PNNL to use their threat generator to simulate large sets of financial transactions. Some of this data will have suspicious activity inserted in them.

Highlight:

Working with Agus Sudjianto, Senior Vice President and head of the analytics groups at Bank of America (BoA), we have begun a project that involves predictive risk analytics, especially of emerging risks. The BoA approach is heavy on automated methods and statistical techniques, but due to the very large data and complex analyses they realize that humans must be involved in

attaching meaning to the results, discovering insights, and drawing conclusions. This is where interactive visualization and visual analytics comes in. In addition to bank transaction data, we are looking at new data, such as credit card transactions. There will be a period of months where we will jointly develop ideas and methods followed by presentation to senior management. After that the project could get a lot bigger on the bank's side. We are planning a fellowship program that could support several Ph.D. students in visualization and/or analytics. These students will work intimately with developers in the BoA analytics group and with their thesis advisors, developing and pursuing relevant thesis topics.

This work ties in with the financial analytics work in the DHS project, but it is broader. We are bringing the FraudWatch tools into this arena, where they will be augmented for predictive risk analytics. In addition, we are developing a new set of complementary tools and methods, with BoA help, that will look at credit ratings, amount of secured and unsecured loan activity, types of loans, geographic distribution to bring out the economic climate and health of a region, and other factors to determine detailed, time-dependent risk profiles that are flexible enough to account for unforeseen factors revealed in the data analysis. These factors can then be inserted for more robust predictive risk analysis. This visual analytics-based analysis will address resiliency in the financial system, providing a new dimension of interest to DHS.

Finally, BoA has developed a rather rigorous curriculum for training new employees from non-financial backgrounds in what they need to know in financial analytics. This can be quite useful input for our education programs; in particular, we plan to develop a financial informatics and analytics Professional Science Masters that can be a significant part of the VACCINE educational offerings.

MDRP 12: Foreign Animal and Zoonotic Disease Visual Analytics

Team

VACCINE:
Purdue University: David Ebert, William (Bill) Cleveland
Pennsylvania State University: Alan MacEachren, Frank Hardisty

FAZD:
Texas A&M (FAZD): Newville Clark, David Hartley

Abstract

This work focuses on developing and applying new exploratory visual analytics (EVA) methods for spatial and space-time foreign animal and zoonotic disease data (e.g., data related to avian influenza, rift valley fever, swine flu, West Nile virus). The VACCINE team will extend their methods for spatial cluster and pattern analysis and their implementation. A key goal is to provide tools that help analysts clarify the spatial (and spatio-temporal) structure of foreign animal and zoonotic disease phenomena, thus leading to better disease outbreak prediction. In complementary work they will explore use of geographic information systems (GIS) to integrate disease, environmental, demographic, and other geographic data necessary to understand disease-environment and disease-human relationships. Initial work will focus on creating an initial prototype visual analytic system for Rift Valley Fever (RVF) tailored to two user groups: epidemiological modelers and decision makers.

VACCINE efforts resulted in a suite of spatial analysis methods implemented in the GeoViz Toolkit that will lead to better understanding of diseases of zoonotic origin. These methods include interactive cluster detection, interactive spatial structure exploration (using Moran's I), and space-time spatial structure exploration (using LISTA-Viz). These methods will be applied against the West Nile virus dataset to investigate virus distribution over space and time.

Grant activities also included sorting and aggregating Pennsylvania mosquito data to allow for efficient data viewing through web-based application at weekly/monthly/yearly intervals. This has also been done for daily climate information. The year with the worst WNV-positive mosquitoes as well as the most number of reported human cases was used for preliminary testing of a web application.

Technical Approach

The data collection methods used were compiling & sorting DB, and expert consultation. Statistical analysis and modeling were used for analytic methods. The nature of research was hybrid basic-applied.

Milestones

Milestones for VACCINE Period 1	07/01/09 – 03/31/10			
	Quarterly			
Development of epidemiological visual analytic environment prototype for RVF		X	X	
Development of linked epi-economic visual analytic prototype			X	X
Design of decision maker interface and decision factors for RVF planning and response				X
Prepare plan outlining approach to GIS-based analysis of PA West Nile Virus data		X		
Prepare plan for interface enhancements to support usability and utility of the integration of Proclude into the GeoViz Toolkit		X		
Release version of GeoViz Toolkit with enhanced Proclude usability.			X	
Generate demo and video illustrating the GeoViz – spatial cluster detection advances and their application to representative data.			X	
Prepare paper on integration of GeoViz and Proclude for submission.				X
Prepare development plan for Space-Time Local Moran's I methods				X
Report on methods developed for analysis of WNV data				X

Project Outcomes (Results, Accomplishments, Conclusions)

Publications

1. Auer, T., MacEachren, A.M., McCabe, C. and Pezanowski, S. submitted: HerbariaViz: A web-based client-server interface for mapping and exploring flora observation data. Ecological Informatics. {joint work with MDRP 3}
2. Brigantic, R.T., Ebert, D.S., Corley, C.D., Maciejewski, R., Muller, G.A., and Taylor, A.E. Development of a Quick Look Pandemic Influenza Modeling and Visualization Tool. ISCRAM2010: 7th International Conference on Information Systems for Crisis Response and Management, 2010.
3. Chen, J., MacEachren, A.M. and Peuquet, D. 2009: Constructing overview + detail dendrogram-matrix views. IEEE Transactions on Visualization and Computer Graphics 15, 889-896. {joint research with MDRP 3}
4. MacEachren, A.M. and Pezanowski, S. (2009). "Geovisualization: Leveraging the Opportunities of Geographic Information." Adobe Developer Connection Education Developer Center: Articles from educators: Adobe: http://www.adobe.com/devnet/edu/articles/macEachren_pezanowski.html. {overview online paper that presents work related to MDRP 3, 12, & 13}
5. Robinson, A. (2009). Visual Highlighting Methods for Geovisualization. 24th International Cartographic Conference. Santiago, Chile (conference CD, no page #s) {paper introduces a visual display method relevant to Penn State tool development in MDRP 1, 3, 12, and 13} http://www.geovista.psu.edu/publications/2009/ACR_ICA_2009.pdf

Technology Transitions

1. The GeoViz Toolkit is distributed as free Open Source software. The team has been active in outreach to public health communities focused on both research and practice to introduce them to the tools.
2. The GeoViz Toolkit was introduced to public health practitioners in a Tutorial given at the URISA GIS in Public Health Conference. Providence, RI, June 5-8, 2009.

Other outcomes/impacts

Development of extensions to the GeoViz Toolkit to support spatial and space-time analysis
In VACCINE's work on component-oriented coordination methods to support geovisual (and more generally visual) analysis, the team has developed a new strategy for visualization coordination through the combination of several recent advances in software engineering: automatic introspection of objects, software design patterns, and reflective invocation of methods. This strategy is called Introspective Observer Coordination and demonstrates how this method can allow end users to create their own visualization toolkits on the fly, solve challenges associated with coupling spatial analysis methods to geovisualization methods, and make geovisualization applications sharable between users.

In order to identify autocorrelation in space and time simultaneously, the team has implemented a spatio-temporal version of the Local Moran's I statistic, and claim two advances. First, they exploit the fact that there are a limited number of topological relationships present in the data to make Monte Carlo simulation computationally tractable, and thereby bypass the "curse of dimensionality" which has prevented widespread application of this method in the past. Second, the team developed a tool (LISTA-Viz) for interacting with the spatio-temporal structure uncovered by the statistics.

In addition to the above, Hardisty and Robinson developed a plan to iteratively refine the Proclude GeoViz Toolkit interface to enhance its usability and utility, to complete one round of evaluation and interface enhancements.

Visual analytics, GIS, and spatial analysis applied to West Nile Virus Data

In addition to the team's core method and tool development research, they have pursued a case study application that supports work of the Pennsylvania West Nile Virus monitoring program. A multi-year data set (of 35,000 insect traps and results of testing for WNV) over multiple years has been obtained from the State of Pennsylvania. Analysis of the data revealed that thirteen species contained WNV. Of these, two key species may be involved in the transmission of West Nile virus in Pennsylvania with an additional two competent vectors. Distribution of these species varies in space and time within a year and between years with species composition

changing within a season and between years. Data is in point form and has been aggregated to municipality level to improve computation time.

The data covers 9 years of information with mosquitoes collected at a variety of traps, at different intervals within a single location and at different intervals across the different trap locations. For example, sampling may occur on a weekly basis at a single location during the peak season and less regularly at the start and end of the season and/or certain locations once the problem has been reduced. In addition, frequency of sampling at the same location may change. Of the total sampled locations, 1400 have been sampled annually while others have been sampled intermittently. Additional complexities inherent in the data are the traps. A number of traps have been used each attracting a specific species and/or a species at a specific life-stage (e.g. host-seeking females seeking their first blood meal versus host-seeking females that have already had at least one previous blood-meal and may have acquired the virus).

Preliminary analyses were performed on the data using the GeoViz Toolkit Proclude module, but it is not yet clear if significant "hotspots" were identified. Due to the complexity of the data, the next stage is to run a series of time-series analyses against species-trap type to determine how species-dominance may change over time. This will be run at different aggregated scales and using different time slices. Initial investigations will include the use of the Morans-I statistic time-series analysis described below.

Environmental factors such as temperature play a key role in the development of mosquitoes as well as the ability of the mosquito to transmit the virus once it has been acquired. Daily temperature and precipitation data were also obtained for Pennsylvania for the time-frame of the mosquito data (2001-2009). A degree-day model has been used in Wyoming and California to predict West Nile virus occurrence over time. A PhD student has been comparing accuracy of degree models and thresholds in predicting West Nile virus in Pennsylvania at four locations. Preliminary results show that using temperature may be sufficient in predicting the transmission of West Nile virus in one location of Pennsylvania and against one species. Initial results were consistent against 2 years. Based on these results, potential collaborators have been consulted about a possible grant proposal to another funding agency that would leverage our DHS work.

Three scenarios were developed to aid in the development of web-based tools for the exploration and communication of data about WNV. These included providing (a) basic information for the general public to view current risk, (b) comprehensive level of information with some interaction possible, and (c) comprehensive level of information with data interaction and statistical analysis. Medium and comprehensive levels are aimed at researchers and policy makers with a more in depth understanding of the WNV-mosquito system as well as statistical analysis.

The VACCINE team developed a suite of spatial analysis methods in the GeoViz Toolkit that will lead to better understanding of diseases of zoonotic origin. These methods include

interactive cluster detection, interactive spatial structure exploration (using Moran's I), and space-time spatial structure exploration (using LISTA-Viz). These methods will then be applied against the West Nile virus dataset to investigate virus distribution over space and time.

In addition, the team has sorted and aggregated the Pennsylvania mosquito data to allow for efficient data viewing through web-based application at weekly/monthly/yearly intervals. This has also been done for the daily climate information. The year with the worst WNV-positive mosquitoes as well as the most number of reported human cases was used for preliminary testing of a web application.

Visual analytics and decision making tools applied to pandemic modeling

Other scenario analysis tools include our work in pandemic modeling and visualization. PanViz technology provides public health officials with a suite of visual analytic tools for analyzing the spread of a pandemic based on Gaussian mixture models. This tool allows officials to analyze various decision points (school closure, media reports and strategic national stockpile medicinal releases) and their impact on the spread of the pandemic. In order to help public health officials better understand these charges, we have developed the PanViz toolkit. Through the application of the Indiana State Pandemic Influenza Planning Tool developed by partners at the Pacific Northwest National Laboratory, we simulate a pandemic outbreak originating in Chicago, IL with attack and mortality rates similar to the 1918 pandemic. Our PanViz tool allows officials to track the spread of influenza across the state of Indiana and implement various decision measures at any time during the pandemic. These decision measures can be toggled on and off to allow users to better understand their effects on different county populations. Demographic filtering is also available for various age ranges, and interactive manipulation of model parameters allow users to create various levels of pandemic severity in order to assess various situations. By using this tool, officials can analyze resources and decision.

Visual analytics and decision making tools applied to Rift Valley Fever modeling

Along with modeling pandemics using a Gaussian mixture model, other work has focused on differential equation spread models, particularly focusing on the spread of Rift Valley Fever. Our work in rift valley fever modeling provides a framework in which scientists, decision makers and healthcare officials can analyze potential threats to various environments. Information about disease transmission and infected areas is displayed in a linked geo-spatiotemporal environment allowing for the analysis of disease spread over time. Users are interactively able to analyze the loss of life, illness and economic impact across affected geographic areas. While recent work has focused on accurately modeling disease spread, little work has been done in developing interactive decision support tools for analyzing the future course of the outbreak and evaluating potential disease mitigation strategies. The absence of such tools makes it difficult for researchers, analysts and public health officials to evaluate response measures within outbreak scenarios. As such, our research focuses on the development of an interactive decision support environment in which users can explore epidemic models and their impact. This environment

provides a spatiotemporal view where users can interactively utilize mitigative response measures and observe the impact of their decision over time. The environment also provides decision history visualization and navigation tools that support the comparison of mortality and infection rates corresponding to different response measures at different points in time. Our tool provides decision history support visualizations to help analysts identify locally optimal combinations of response measures. In this manner, our tool allows for simultaneous decision comparison and minimizes the risk of wasting resources while supporting an effective mitigative campaign.

MDRP 13: GeoJunction: Collaborative Visual-Computational Information Foraging and Contextualization to Support Situation Awareness

Team

VACCINE:
Pennsylvania State University: Alan MacEachren, Prasenjit Mitra

CCICADA:
University of Southern California: Ed Hovy

Abstract

This project focuses on developing, implementing, assessing, and transitioning methods and tools to (a) help analysts, crisis managers, and policy makers find and contextualize relevant multimodal information quickly and (b) to support the situation awareness (SA) that is required to interpret evidence and make decisions in a continually changing environment. Specifically, GeoJunction is conceived as a knowledge-enabled, place-time aware, and computationally enhanced web-based environment that supports visually-enabled: (a) document query filtered by place, time, person/organization, and concept; (b) collaborative information foraging, entity and relation extraction, and contextualization; (c) exploration of connections between social and geographic networks; and (d) situation assessment and monitoring with continually updating information.

Technical Approach

The data collection methods utilized were compiling & sorting DB, data mining, expert consultation, and survey. Analytic methods utilized were case studies, sampling, and statistical analysis. The nature of research was consultation, hybrid basic-applied, and hybrid applied-consultation.

Milestones

Milestones for VACCINE Period 1	**07/01/09 – 03/31/10**			
	Quarterly			
Release working web prototype of the faceted place-time-concept document query and review interface		X		
Complete at least one MS or PhD research proposal related to this effort		X		
Produce a video-demo for use in recruiting potential users to participate in human-centered design activities as the project moves forward			X	
Submit initial paper for publication detailing web interface for faceted place-time-concept document query & review			X	
Design architecture for generalizing the system to add support for collaborative information foraging, entity and relation extraction, and contextualization			X	
Add capabilities to support a wider range of document types and a wider range of geographic information in the map component of the interface			X	X
Work with colleagues at UNCC, Purdue, and USC, produce cross-site work plan				X

detailing coordination/collaboration & coordinate with potential users				

Project Outcomes (Results, Accomplishments, Conclusions)

Publications

1. Cai, G. and B. Yu (2009). "Spatial annotation technology for public deliberation." Transactions in GIS 13: 123-146. http://spatial.ist.psu.edu/cai/Cai_TGIS_09.pdf
2. Hardisty, F. (2009). "GeoJabber: Enabling Geo-Collaborative Visual Analysis." Cartography and Geographic Information Science 36(2): 267-280. {joint work with MDRP 3}
3. Jaiswal, A., D. J. Miller, et al. (2010). "Un-Interpreted Schema Matching with Embedded Value Mapping under Opaque Column Names and Data Values." IEEE Transactions on Knowledge and Data Engineering 22(2): 291-304. {joint work with MDRP 1}
4. MacEachren, A.M. and Pezanowski, S. (2009). "Geovisualization: Leveraging the Opportunities of Geographic Information." Adobe Developer Connection Education Developer Center: Articles from educators: Adobe: http://www.adobe.com/devnet/edu/articles/macEachren_pezanowski.html {this online overview highlighted Penn State work from MDRP 3, 12, and 13}
5. MacEachren, A.M., Stryker, M.S., Turton, I.J. and Pezanowski, S. 2010: HEALTH GeoJunction: Visual-computationally Enabled Browsing of Health Publications and Their Contents. International Journal of Health Geographics 9(23). http://www.ij-healthgeographics.com/content/pdf/1476-072X-9-23.pdf
6. Yu, C., A. M. MacEachren, B. Yarnal, and D. J. Peuquet. 2009. Integrating scientific modeling with a GeoAgent-based representation of human-environment interactions for supporting dynamic hazard management: a drought example in Pennsylvania, USA. Environmental Modelling & Software 27 (7):1501-1512. {joint work with NSF-funded Human-Environment Regional Observatory project}

Follow-on funding/related projects

The National Science Foundation invited a proposal for a contract project ($25,000) to adapt methods in our initial Health GeoJunction application for application to analysis of their grants database. The objective is to demonstrate the potential of our place-time-concept filtering and sensemaking support tools to support the work of program officers at NSF.

Other outcomes / impacts

Research during the 9 months of year one focused on four linked activities. First, the VACCINE team completed the development of, as well as a publication and video about, Health GeoJunction (an application that supports quick filtering and place-time-concept sensemaking with research abstracts). The second activity was research carried out as part of a course project to investigate usability and utility of GeoDeliberator (a web-based collaborative deliberation tool that implements map-based asynchronous strategies). The third activity involved a Masters degree project that focused on the development of collaborative geographic query tools to support different-place work with maps in contexts such as planning for resilience or crisis

response. The fourth activity was in coordination with the team's work on MDRP #1, they developed a system architecture to support capture, storage, and retrieval of text artifacts to be made accessible to users through different visual interfaces for different application situations.

Health GeoJunction

The initial prototype tool, Health GeoJunction was focused narrowly on quick filtering and place-time-concept sensemaking with research abstracts. The volume of health science publications was and still is escalating rapidly. Thus, keeping up with developments was becoming harder as was the task of finding important cross-domain connections. When geographic location was a relevant component of research reported in publications, these tasks were more difficult because standard search and indexing facilities have limited or no ability to identify geographic foci in documents.

To address this need, Health GeoJunction provides a web application that supports analysts in the task of quickly finding scientific publications that are relevant geographically and temporally, as well as thematically. More specifically, Health GeoJunction (Figure 1) is a geovisual analytics-enabled web application providing: (a) web services using computational reasoning methods to extract place-time-concept information from bibliographic data for documents and (b) visually-enabled place-time-concept query, filtering, and contextualizing tools that apply to both the documents and their extracted content. The emphasis in this component of the research, specifically, is on strategies for visually-enabled, iterative, facet-like, place-time-concept filtering that allows analysts to quickly drill down to scientific findings of interest in PubMed abstracts and to explore relations among abstracts and extracted concepts in place and time. The approach enables analysts to: find publications without knowing all relevant query parameters, recognize unanticipated geographic relations within and among documents in multiple health domains, identify the thematic emphasis of research targeting particular places, notice changes in concepts over time, and notice changes in places where concepts are emphasized.

PubMed is a database of over 19 million biomedical abstracts and citations maintained by the National Center for Biotechnology Information; achieving quick filtering is an important contribution due to the database size. Including geography in filters is important due to rapidly escalating attention to geographic factors in public health. The implementation of mechanisms for iterative place-time-concept filtering makes it possible to narrow searches efficiently and quickly from thousands of documents to a small subset that meet place-time-concept constraints. Support for a more-like-this query creates the potential to identify unexpected connections across diverse areas of research. Multi-view visualization methods support understanding of the place, time, and concept components of document collections and enable comparison of filtered query results to the full set of publications.

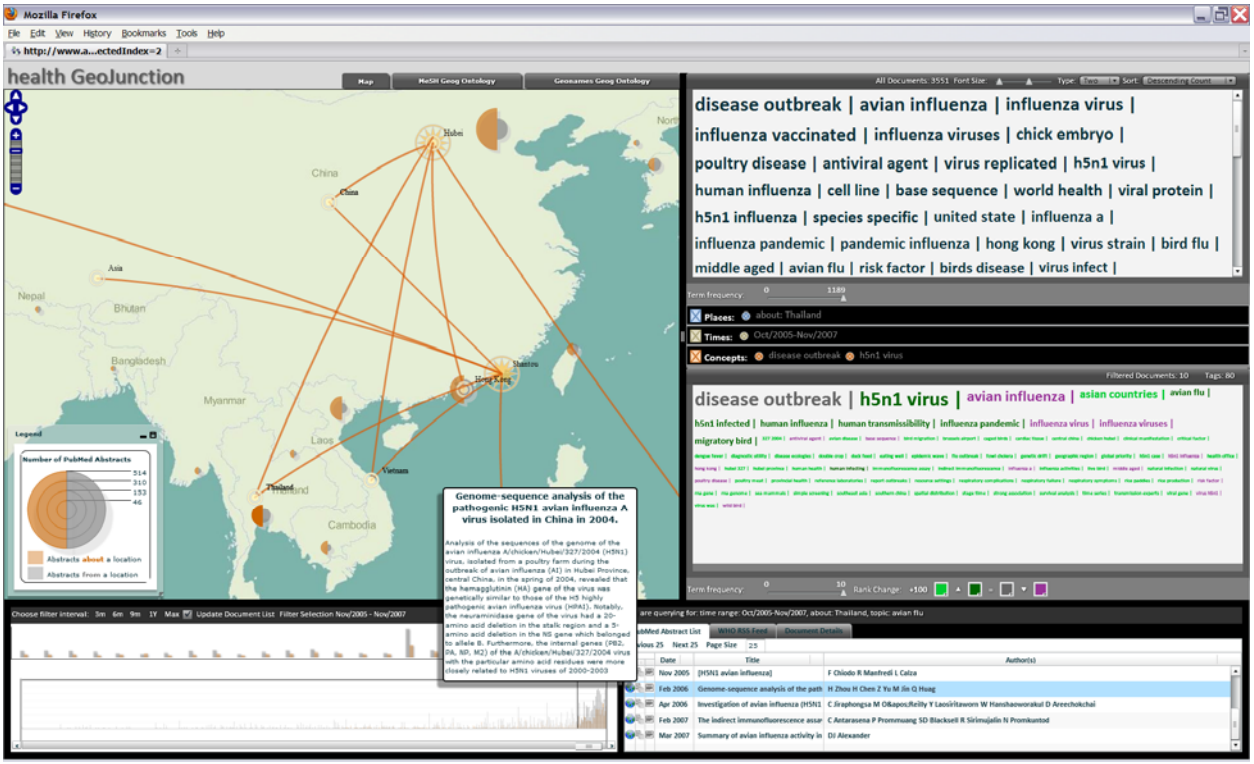


Figure 1. Shows the result after the user clicked on the "about" side of the Thailand symbol on the map, filtering the result to only those judged (based on MeSH or GeoJunction feature extraction tools) to be about Thailand, and the user clicked on "disease outbreak" in the top tag cloud to further filter to the subset of documents about Thailand that are also about disease outbreaks. Not surprisingly, H5N1 virus has moved up to be the third most frequent term in this set of documents (now including only 23 of the original set). Then, the user identified two papers of interest. The geographic footprints of both are depicted on the map and the abstract for one is highlighted.

GeoDeliberator Assessment

In a controlled study (conducted without VACCINE funding as part of a Geovisual Technology Use & Usability course, under its own IRB), the team investigated the extent that a web application, the GeoDeliberator (Figure 2), facilitated the construction and use of map annotations as visual artifacts for mediating asynchronous deliberation about a geographic problem. A multi-view interface design provided linked geographic, temporal, and conceptual views of discussion contributions made as map annotations, comments georeferencing marks placed on the map. Employing a quasi-asynchronous protocol in a lab environment, the team assessed the usability of the system; while participants performed a number of decision-making activities. During each session the means for real-time communication was removed and pairs of stakeholders were asked to use a map-based tool to articulate issues and identify the issues most relevant to a geographic problem solving scenario.

Results of the assessment outlined will be applied to guide the design of follow-up studies assessing the utility of the system for aiding the process of deliberation. The design decisions assessed in this study are an initial implementation of approaches that allow group work with relevance to a number of group activities including: visual query of geographic information across space-time-conceptual dimensions; collaborative, geographically contextualized information foraging; exploring social network interactions through deliberation artifacts; monitoring activities and situation assessment.

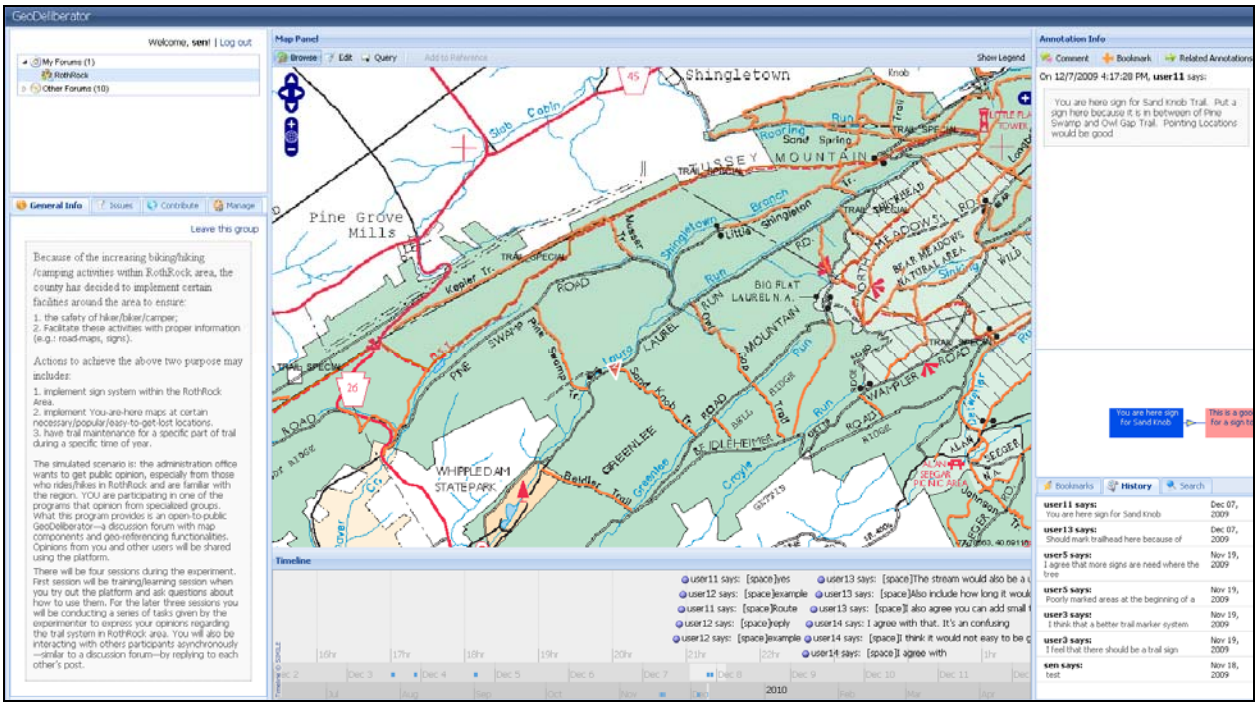


Figure 2: The GeoDeliberator interface shows map annotations over a trail network map (middle); a time series view of the discussion (bottom); a history of recent activity (lower right); a network view of contributions and responses (right, middle).

Collaborative Spatial Query

Processes of knowledge-building, situation assessment, planning and decision-making are frequently accomplished by groups of remotely located people working together. Research in the field of geocollaboration has been aimed at discovering ways to support this type of collaborative work using GIS technologies. Past research in geocollaboration has typically started with an assumption that all relevant data have been assembled and represented on a shared map. However, that is often not the case, particularly in domains like crisis response and risk assessment. A key task often faced by collaborators is to find the right data to meet the task at hand, thus, there is a need for tools that support synchronous collaborative spatial data queries.

The goals of this research were (a) to create the Spatial Queries in Synchronous Collaboratories (SQSynC) software to enable remote, collaborative spatial query in real-time requiring only an

internet connection and web browser and (b) to analyze the methods used by geographers to construct the queries in this environment to develop a deeper understanding of the process.

The SQSynC software prototype is a web-based application implemented using Adobe Flex, Google Maps and PostgreSQL. SQSynC can support both synchronous and asynchronous collaborative interactions between the users and the interface (Figure 3). In synchronous mode, the software behaves as if there is only one interface that every user is concurrently interacting with, even though each user is remotely connected through individual computers. For instance, if one person were to pan and zoom the map, every other user would see the map change. In asynchronous mode, the software provides each user with a distinct, private interface that no other users can see or interact with. To enable more direct collaboration, a second, public tab exists. Users can move items from their private tab over to the public tab in order to share them with the other users. In both modes, a visual query interface allows users to quickly select categories of information from large and distributed data repositories and to apply filters to the multiple data categories quickly.

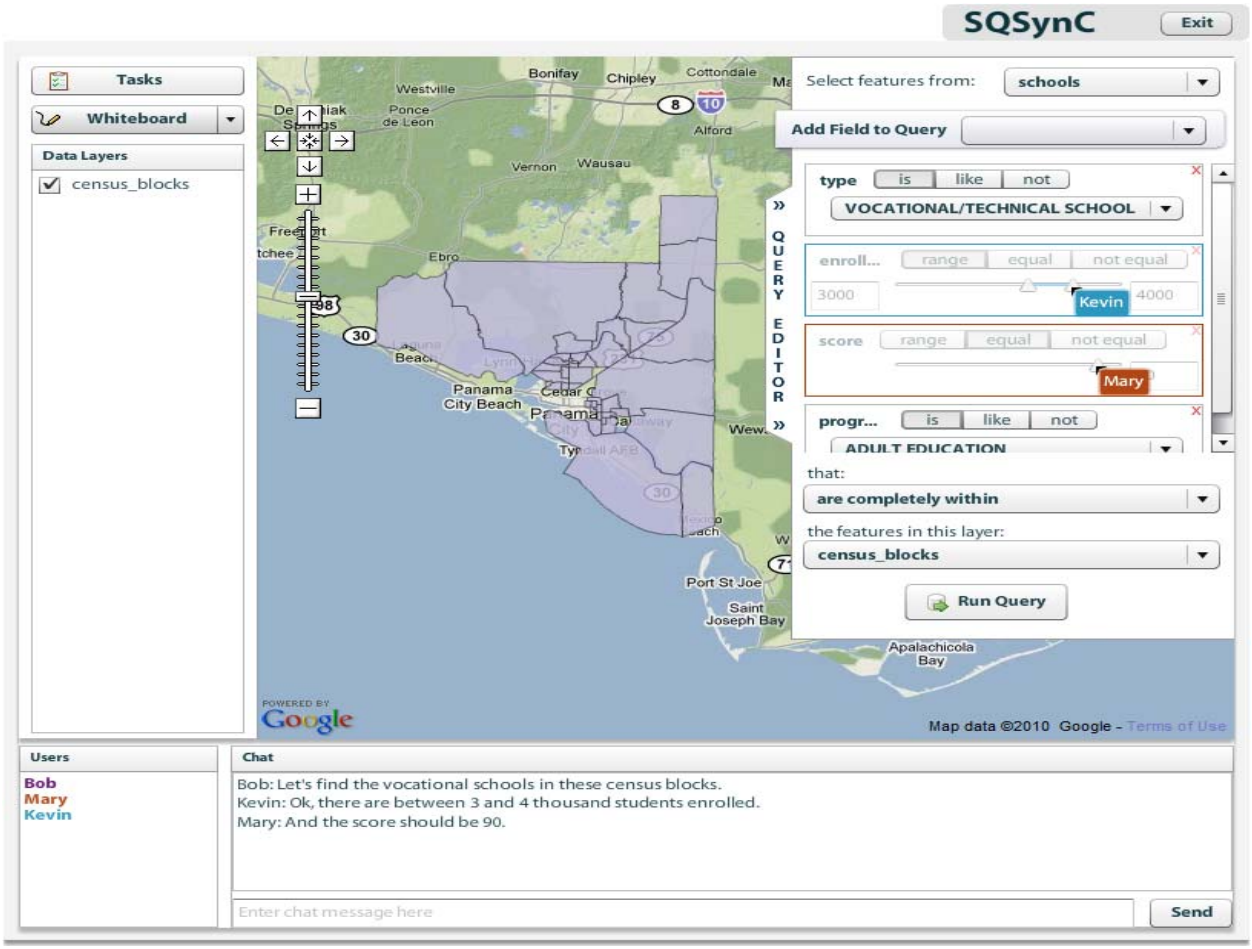


Figure 3: SQSynC running in synchronous mode, showing three users simultaneously interacting with the query editor to find a specific set of data. Each user's cursor gets tracked and as a user interacts with a widget, that widget is locked for all other users to prevent conflict.

Experiments were conducted using the SQSynC prototype, to explore how geographers collaboratively query spatial data. Participants, working in pairs, were asked to collaboratively work through a set of tasks using the synchronous mode of the application, and then another set of tasks using the asynchronous mode (half of the participants started with the synchronous mode, the other half started with the asynchronous mode). In the synchronous mode, each user could simultaneously interact with a shared query building interface.

In the asynchronous mode users could communicate with each other and share data on the public tab, but there was no shared query building interface with which they could simultaneously interact, only a private interface (Figure 4). A total of 16 participants took part in the experiment, and they each completed a follow-up survey that asked questions about which mode they preferred, any critical incidents that occurred, features that they liked, and other comments and suggestions. For various given reasons, 56% of participants preferred the asynchronous mode of collaborative query building, while 44% preferred the synchronous mode. The results of this experiment ultimately show that in geocollaboration software there is a place for both private and shared query building tools.

Figure 4: Part of the SQSynC interface when running in asynchronous mode, showing the private and public tabs.

GeoJunction Software Architecture – adapted to support NSF grants data analysis

With support from the National Science Foundation in the form of a contract ($25,000), a second generation GeoJunction interface has been developed. As part of the process, we have begun the process of developing a more robust and extensible system architecture that shares major components with those developed for our SensePlace application (discussed within MDRP #1).

The GeoJunction-2 system provides capacity for querying documents based on themes, places, and time. Figure 5 shows the user interface that facilitates the query. Specifically, the Tag cloud view and Text Query view allow fuzzy querying on multiple thematic terms. The Time view and Map view restrict the results (a list of documents) for specific time periods, with the ability to focus in some geographical regions (e.g., a particular country). The documents are displayed in the Table view, allowing users to scan through and select an interesting one to read in the Document view. Related documents can be found for the selected one. The system architecture is described next.

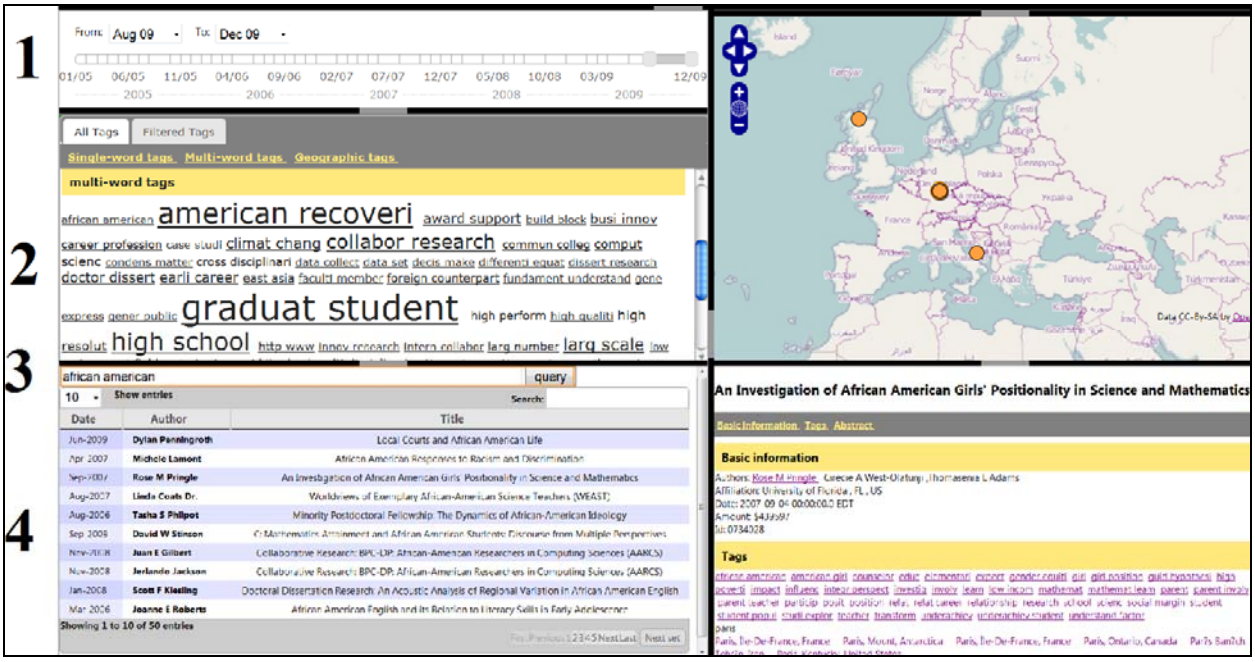


Figure 5: User interface at client-side: (1) Time view, (2) Tag cloud view, (3) Text Query view, (4) Table view, (5) Map view, (6) Document view.

The system was approximately 50% implemented at the close of project year one. It will include the following components: (1) user interface that provides the views; (2) database that store the data that contain text documents and information of geographical location and time); (3) Lucene indexer that facilitate fast query and relevant ranking on the results; (4) Query worker that perform queries on database and Lucene indexer (5) components that provide following services: text processing and extraction, geocoding, and map services; (6) central controller to coordinate client side requests, services, and queries on database and Lucene indexes. The system has been designed, so it can be easily adapted and merged to SensePlace. It can take advantage of the central database system and Lucene indexer of SensePlace, employing the existing services. Assuming the query requests from the users are similar, the main structure of the user interfaces and controller can be reused for SensePlace. More functions can be added to the user interface to meet the requirement of SensePlace.

Key objectives addressed during this initial time period were: (a) a prototype already in development (based on NEVAC research) was completed that implements strategies to support component ‘a’ of the overall environment; (b) produced a video demo (posted online) and a paper (submitted) documenting the visually-enabled approach to place-time-concept query; (c) designed an architecture for generalizing the system to add support for: collaborative information foraging; entity and relation extraction, and contextualization; and situation assessment and monitoring. In addition two student initiated projects were carried out; one focused on collaborative distributed spatial query methods and the other on developing tools for distributed, map-based deliberation.

MDRP 14: Multimedia Visual Analytics for Investigative Analysis

Team

VACCINE:
University of North Carolina at Charlotte: William (Bill) Ribarsky, Jianping Fan, Jing Yang
Purdue University: Edward Delp

Abstract

When many online news sources are analyzed simultaneously, there can be tens of thousands of stories over a given period of time. The ebb and flow of stories on a subject can also have a highly dynamic temporal character. To handle this, VACCINE is developing a fully integrated and comprehensive multimedia visual analytics system that starts with the automated results, and then employs a detailed time-based analysis of related news stories to turn them into "event streams" or "event clusters". Each event stream or cluster begins with a significant motivating event (an occurrence at a particular time and location), which is then followed by a flow of relevant stories and follow-on events. In addition to temporal views of the narrative structures, the team is developing relational views that make the connections among different clusters of stories clear and concept views that go well beyond usual keyword lists to provide meaningful clustering of stories based on content. This group of methods is being applied to online news, blogs, and RSS feeds. The events and concepts appear naturally from the news or other data collections. They are not predetermined; thus an analysis will cover, categorize, and structure all the news.

Technical Approach

This research employs data mining and analytic methods, including modeling, sampling, and statistical analysis. The nature of the research is hybrid basic-applied.

Milestones

Milestones for VACCINE Period 1	**07/01/09 – 03/31/10**			
	Quarterly			
Development of initial system for weakly-tagged images/video			X	
Integrated exploratory event-based tools built on automated analyses			X	
Beginning capabilities for Multi-source event browser				X
Publication of papers in journals and conference proceedings				X

Project Outcomes (Results, Accomplishments, Conclusions)

Publications

1. Y. Gao, J. Peng, H. Luo, D. Keim, **J. Fan**, ``An Interactive Approach for Filtering out Junk Images from Keyword-Based Google Search Results", **IEEE Trans. on Circuits and Systems for Video Technology**, vol. 19, no.10, 2009.

2. D. Luo, J. Yang, M. Krstajic, J. Fan, W. Ribarsky, and D. Keim. EventRiver: Interactive visual exploration of constantly evolving text collections. To be published. **IEEE Trans. On Visualization and Computer Graphics** *(2010).*
3. Fan, J. D. Keim, Y. Gao, H. Luo, Z. Li, "JustClick: Personalized Image Recommendation via Exploratory Search from Large-Scale Flickr Images", IEEE Trans. on Circuits and Systems for Video Technology, vol. 19, no.2, pp.273-288, 2009.
4. Fan, J., Y. Shen, N. Zhou, Y. Gao, ``Harvesting Large-ScaleWeakly-Tagged Image Databases from the Web", **IEEE Conf. on Computer Vision and Pattern Recognition (CVPR'10)**, 2010.
5. Shen, Y., J. Fan, ``Leverage loosely-labeled images and inter-object correlation for classifier training and multi-label image annotation", ACM Multimedia, 2010.
6. Xue, X., H. Luo, **J. Fan**, ``Structured Max-Nargin Learning for Multi-Label Image Annotation", **ACM Conf. on Image and Video Retrieval (CIVR'10)**, 2010.
7. Zhang, Y., L. Cen, X. Xue, N. Zhou, ``Bilingual query translation for multimedia retrieval", ACM Multimedia, 2010.
8. Zhou, N., J. Fan, ``Integrating bilingual query results for junk image filtering", ACM Multimedia, 2010.

Other outcomes/impacts

Click2Annotate and ManyInsights

Click2Annotate is an insight annotation approach that allows semi-automatic insight annotation at the sub-task level, which has higher levels of semantic richness and abstraction than at the action or event level in which the automation of most existing insight externalization approaches are conducted. ManyInsights is a multidimensional visualization system with Click2Annotate and other insight management activities implemented. The semantically rich annotations semi-automatically generated by Click2Annoate enable a rich set of effective insight management activities. Currently, there are scented insight browsing and faceted insight search (see Figure 1) implemented. The team is working on more insight management activities.

Touch2Annotate

This approach is an application of Click2Annoate on multi-touch interfaces. The semi-automated approach greatly reduces the need of typing notes in multi-touch interfaces where typing is effort-intensive. Figure 2 shows Touch2Annotate is used on a multi-touch table.

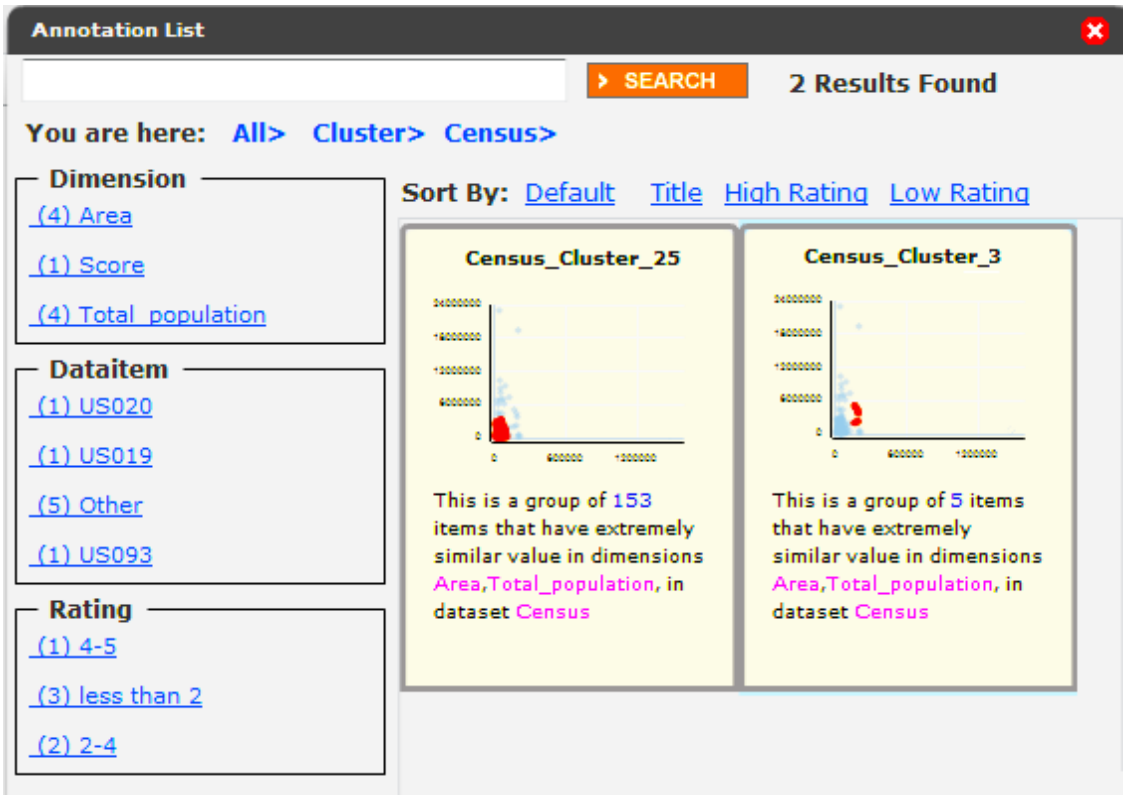


Figure 1: Faceted search of semi-automatically annotated insights in ManyInsights

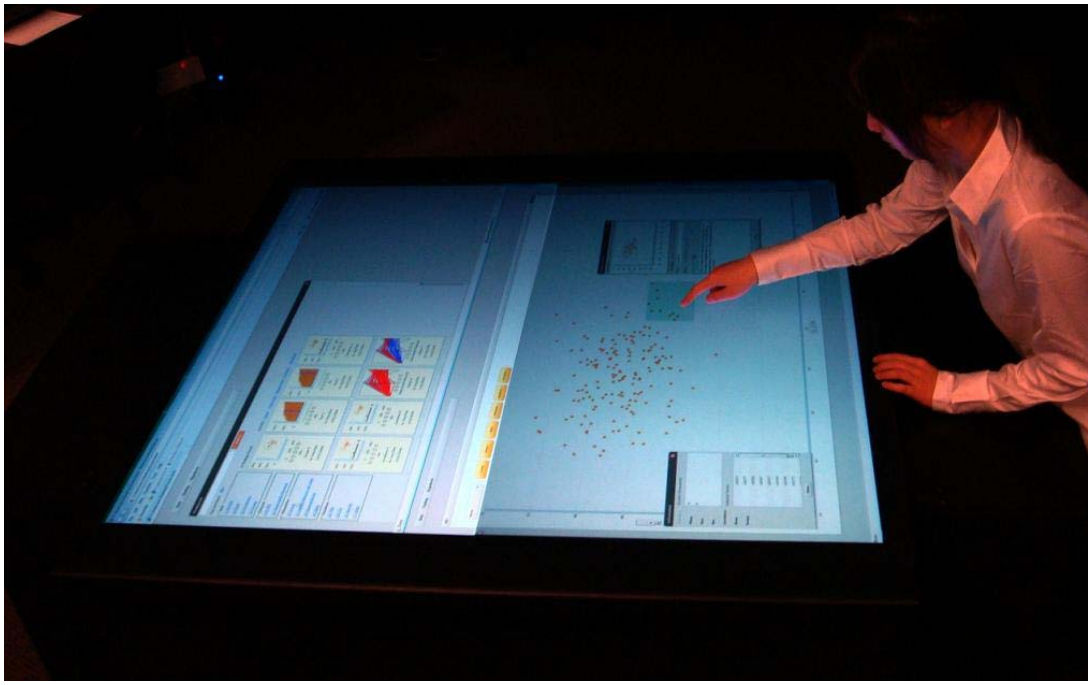


Figure 2: Touch2Annotate

Pixel-Oriented Graph Visualization

The pixel-oriented graph tool relates groups of keywords derived from full text analysis of documents, including multimedia documents, or from other content such as analyzed and tagged sets of images. With the tool (Figure 3), one can simultaneously manipulate and understand the conceptual clusters, and the first, second, or third nearest-neighbor relations among the keywords in any cluster or set of clusters.

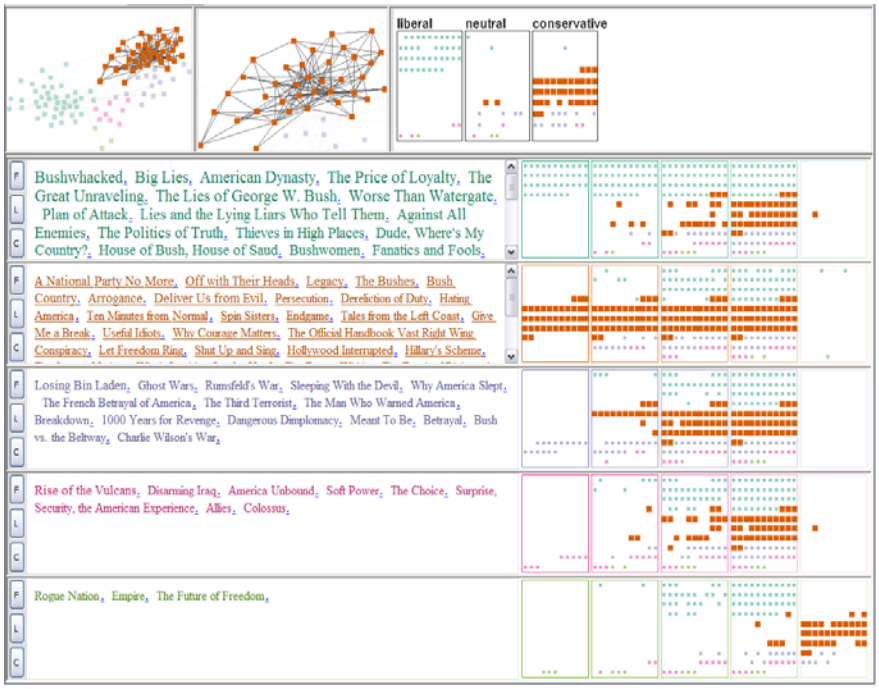


Figure 3: Exploring and understanding relationships among concept clusters

Cross-Media Alignment for Online Multimedia Analysis

As digital images/videos are growing exponentially on the Internet, there is an urgent need to develop new algorithms for web image/video indexing by automatically aligning the web images/videos with their most relevant auxiliary text terms (which are extracted from the associated text documents). Google Images has achieved big success on supporting keyword-based web image retrieval by loosely indexing the web images with their auxiliary text terms. For one given web image, many of its auxiliary text terms are weakly-related or even irrelevant with its semantics because the relevant web page may consist of rich word vocabulary for web content description rather than only the auxiliary text terms for image semantics description. When all these auxiliary text terms are loosely used for web image indexing, Google Images may seriously suffer from low precision rates and result in large amounts of junk images. To enable more effective web image indexing and retrieval, it is very attractive to develop new algorithms for achieving more precise alignment between the web images and their auxiliary text terms.

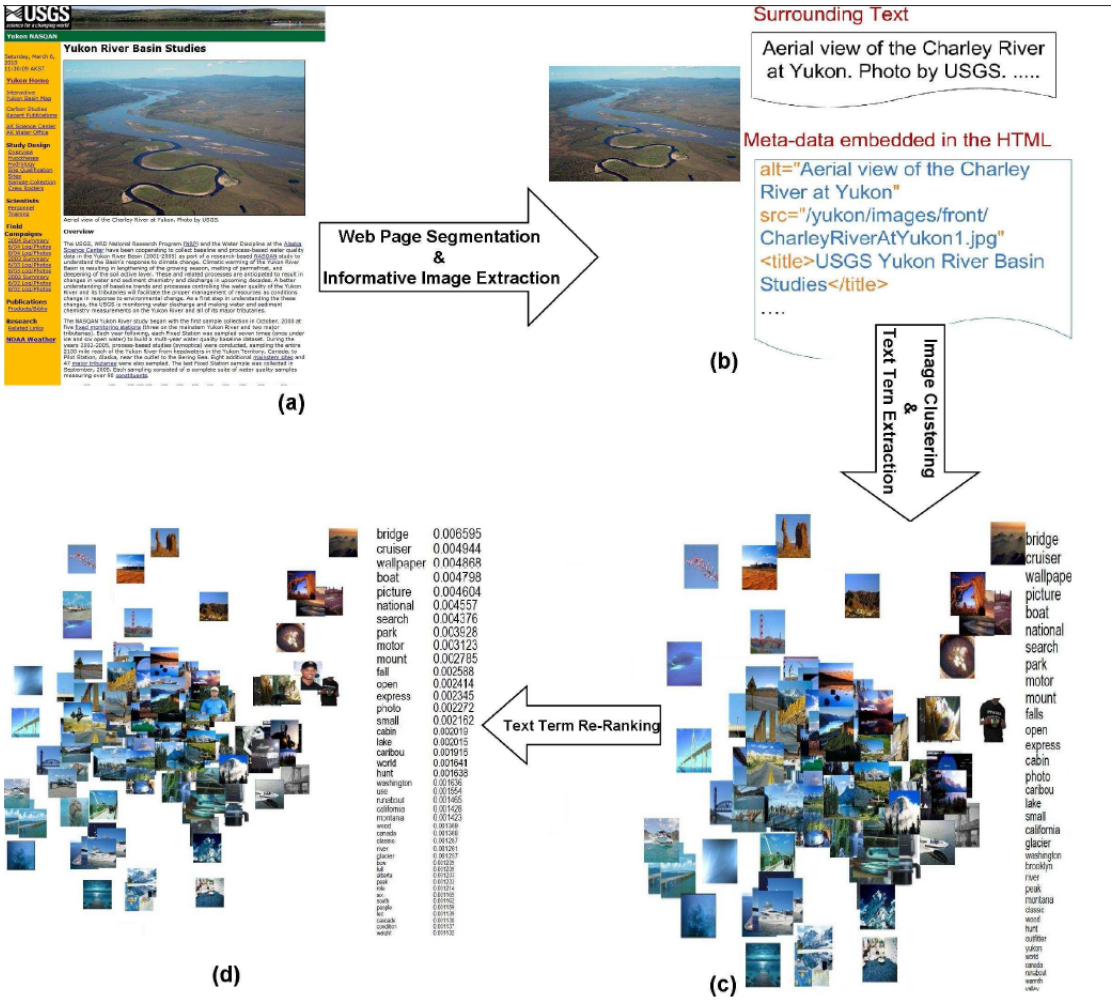


Figure 4: The illustration of the key components of our image-text alignment scheme: (a) web page crawling; (b) image-block pair generation; (c) image cluster and ranked auxiliary text terms; (d) image cluster and re-ranked auxiliary text terms.

In this project, a novel algorithm was developed for image annotation by automatically aligning the web images with their most relevant auxiliary text terms as shown in Figure 4. First, large-scale web pages (i.e., web images and their auxiliary text documents) were crawled and automatic web page segmentation was performed to extract the informative images and their most relevant auxiliary text blocks. Second, automatic image clustering was performed to partition the web images into a set of image clusters according to their visual similarity contexts. Such image clustering process, which could group the web images according to their common properties, could significantly reduce the uncertainty on the relatedness between the web images and their auxiliary text terms. Thus the semantics for the visually-similar web images in the same cluster could be described effectively by a same set of auxiliary text terms which may have co-occurred frequently in their text blocks. Finally, a relevance re-ranking algorithm was developed for aligning the web images with their most relevant auxiliary text terms automatically. The

team's experiments on large-scale web pages (as shown in Figure 5 and Figure 6) have provided very positive results.

Figure 5: Image clustering results and their auxiliary text terms.

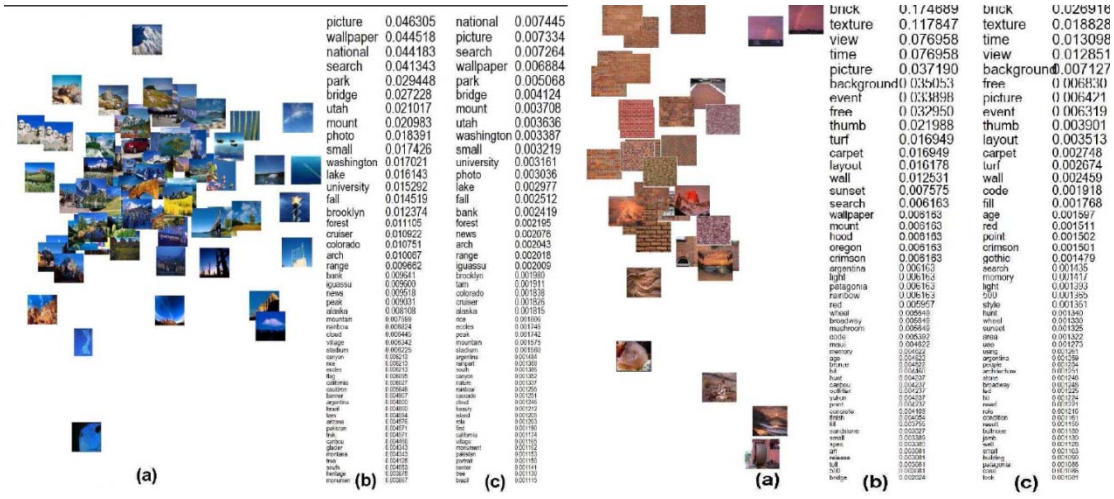


Figure 6: Image-text alignment: (a) image cluster; (b) ranked text terms before performing random walk; (c) re-ranked text terms after performing random walk.

MDRP 16: Context-aware Mobile Visual Analytics for Emergency Response

Team

VACCINE:
Purdue University: David Ebert, Ed Delp, Tim Collins
University of Stuttgart: Tom Ertl, Daniel Weiskopf

Abstract

VALET, the Visual Analytics Law Enforcement Toolkit, is being developed to analyze spatiotemporal law enforcement data. VALET provides users with a suite of analytical tools coupled with an interactive visual interface for data exploration and analysis. This system includes linked views and interactive displays that spatiotemporally model criminal, traffic and civil (CTC) incidents and allow officials to observe patterns and quickly identify regions with higher probabilities of activity. Our toolkit provides analysts with the ability to visualize different types of data sets (census data, daily weather reports, zoning tracts, prominent calendar dates, etc.) that provide an insight into correlations among CTC incidents and spatial demographics. In the spatial domain, we have implemented a kernel density estimation mapping technique that creates a color map of spatially distributed CTC events that allow analysts to quickly find and identify areas with unusually large activity levels. In the temporal domain, reports can be aggregated by day, week, month or year, allowing the analysts to visualize the CTC activities spatially over a period of time. Coupled with the spatial mapping view, our system provides analysts with a time series view of the data. Furthermore, we have incorporated spatiotemporal prediction algorithms to forecast future CTC incident levels within a 95% confidence interval. Such predictions aid law enforcement officials in understanding how hotspots may grow in the future in order to judiciously allocate resources and take preventive measures. Our system has been developed using actual law enforcement data and is currently being evaluated and refined by a consortium of law enforcement agencies.

Technical Approach

This research employs compiling and sorting databases, data mining and expert consultation, including modeling and statistical analysis. The nature of the research is hybrid applied-consultation.

Milestones

Milestones for VACCINE Period 1	**07/01/09 – 03/31/10**			
	Quarterly			
Completion of task and scenario generation		X		
Creation of Desktop software		X	X	
MOU development			X	
Agency signing of MOU's				X
Demonstration of desktop system and mobile software components			X	X

Specification document for next phase of project based				X
End-user evaluation, evaluation summary report				X

Project Outcomes (Results, Accomplishments, Conclusions)

Publications

1. Malik, A., Maciejewski, R., Collins, T., Ebert, D., Visual Analytics Law Enforcement Toolkit, IEEE Conference on Technologies for Homeland Security, 2010 (submitted)
2. Liu, Jianfei, Kyle Lyons, Kalpathi Subramanian, William Ribarsky. Semi-Automated Processing and Routing Within Indoor Structures For Emergency Response Applications", Proceedings of SPIE Defense, Security+Sensing, 2010.

Student Theses

1. Ross Maciejewski, PhD, Purdue, Electrical and Computer Engineering, ECE

Other outcomes/impacts

1. A Public Safety Consortium was organized to aid the development of VALET, and future research projects. The VACCINE public safety consortium is a VACCINE resource intended to be utilized by all VACCINE partner schools and the DHS CCI division of S&T. MOU's have been signed between each of the members and the VACCINE Center. VACCINE faculty and student(s) have undergone background checks, and VACCINE has an assigned desk in the Tippecanoe County Sheriff's Dept. within the detective bureau. We are working directly with the officers to build VALET according to their requirements.

   Current Members:
   Tippecanoe County Sheriff's Department
   Lafayette Police Department
   West Lafayette Police Department
   Purdue Police Department
   Tippecanoe County Emergency Management Agency

2. VALET will promote evidence-based policing, prospective policing, and results-oriented policing. In addition, it will help coordinate and improve emergency management's planning, response, mitigation, and recovery efforts for any emergency, manmade or natural.

3. The implementation of spatial-temporal analytics using law enforcement data. This concept is not currently available in the commercial market.

4. Implementing additional data layers to allow advanced analytics. Layers include, weather, GIS locations of abandoned buildings, census, etc. These layers provide a previously unavailable analysis and correlation of law enforcement activity.

5. Public Safety Consortium has given us access to over 10 years of criminal, civil and traffic data.

6. Embedded a graduate student within the law enforcement community.

MDRP 17: START Center Visual Analytics

Team

VACCINE:
Purdue University: Niklas Elmqvist, David Ebert

START:
University of Maryland: Kathy Smarick

CCICADA:
Rutgers University: Fred Roberts, William (Bill) Pottenger

Abstract

The START Center of Excellence at the University of Maryland has collected large amounts of data on terrorism and terrorism response, such as incident event timelines, social networks of terrorist organizations, and studies on international terrorism and preparedness, but these datasets can be difficult to overview and analyze effectively. This complexity and scale pose a barrier when exposing these datasets to the intelligence and defense communities, as well as the general public. In this project, the VACCINE team collaborated with START members in applying visual analytics techniques to their data in a user-centered design methodology intended to elicit their particular needs and requirements for visualizing this data.

Particular objectives included entity resolution for detecting duplicates in the Global Terrorism Database (GTD), as well as spatiotemporal visualization for the Minorities at Risk database.

Additional research missions remain to be identified.

Technical Approach

Data collection methods used included bench tests, compiling & sorting DB, data mining, and expert consultation. Analytic methods used were case studies, modeling, and statistical analysis. The nature of research was hybrid basic-applied.

Supporting documentation can be found at:

- https://engineering.purdue.edu/~elm/projects/timematrix/timematrix.pdf
- http://engineering.purdue.edu/~elm/projects/graphdice/graphdice.pdf

Milestones

Milestones for VACCINE Period 1	**07/01/09 – 03/31/10**			
	Quarterly			
Identification of START visual analytics needs		X		
Collection of requirements for individual visualizations		X	X	
Spatiotemporal visualization of terrorist incidents			X	X
Multivariate and temporal social networks			X	X

Demonstration of prototype visualizations				X
Specification document for next phase of project				X
End-user evaluation and summary report				X

Project Outcomes (Results, Accomplishments, Conclusions)

Publications

1. Yi, J. S., Elmqvist, N., Lee, S. TimeMatrix: Visualizing Temporal Social Networks Using Interactive Matrix-Based Visualizations. *International Journal of Human-Computer Interaction*, to appear, 2010.
2. Bezerianos, A., Chevalier, F., Dragicevic, P., Elmqvist, N., Fekete, J.-D. GraphDice: A System for Exploring Multivariate Social Networks. *Computer Graphics Forum* (IEEE EuroVis 2010 proceedings), to appear, 2010.

Other outcomes/impacts

The START Center researchers were in need of participatory design for their visualization needs. Earlier attempts at using visual analytics for their large-scale datasets have failed due to requirements that do not match their needs.

The GraphDice tool for multivariate social network visualization was developed in collaboration INRIA in France and has been conditionally accepted for publication at IEEE EuroVis 2009. The TimeMatrix tool was developed in collaboration with Ji Soo Yi and Seungyoon Lee from Purdue and the article has been accepted to IJHCI (International Journal of Human-Computer Interaction).

MDRP 18: Visual Analytics for the DHS Centers of Excellence

Team

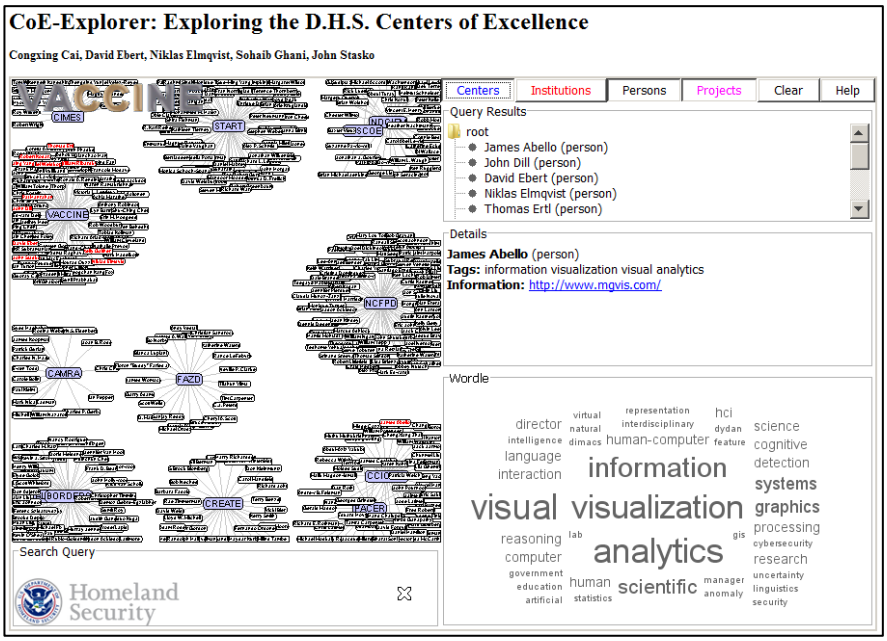
VACCINE:
Purdue University: Niklas Elmqvist, David Ebert
Georgia Institute of Technology: John Stasko
Simon Fraser University: John Dill

CCICADA:
Rutgers University: Fred Roberts
University of Southern California: Ed Hovy
University of Illinois at Urbana-Champaign: Dan Roth

Abstract

The DHS Centers of Excellence has produced a wealth of research, education materials, and projects over the years, and this material will continue to grow as the activities of the various centers expand. The ability for reflection, analysis, and investigation of this material is vital, and may yield both new and old connections, ideas, and thoughts. Furthermore, exposing the material to the scientific community and other government agencies, as well as the general public, is an important goal of our Center. The VACCINE team will leverage visual analytics methods for interacting and exploring this large knowledge resource in an interactive visual form that is easy to access for quick browsing, yet powerful enough to support advanced analysis and review.

This project will develop table-based, graph-based, and timeline-based visualizations of the DHS Centers of Excellence program in an easily accessible format, but which also supports faceted browsing and querying on the multiple dimensions associated with the data. These visualizations will be easy to expand with future data without requiring redesign or rebuilding.



Technical Approach

Data collection methods used were compiling & sorting DB and data mining. Analytic method used was statistical analysis. The nature of research was hybrid basic-applied.

Supporting documentation can be found at https://engineering.purdue.edu/~elm/projects/coe-explorer/CoE-Explorer.pdf.

Milestones

Milestones for VACCINE Period 1	**07/01/09 – 03/31/10**			
	Quarterly			
Data collection and curation		X	X	
Prototype visualizations using existing techniques		X		
Demonstration and formative evaluation		X		
Visual analytic framework for CoE data			X	

Project Outcomes (Results, Accomplishments, Conclusions)

Technology Transitions

1. Provisional patent disclosure in progress (Inventors: Biddanda, Elmqvist, Boutin, Ebert)

Other outcomes/impacts

Additional information can be found at https://engineering.purdue.edu/~elm/projects/coe-explorer/.

SP 1: Improving Query Performance in Large Scale Analytic Databases

Team

VACCINE:
Stanford University: Pat Hanrahan

Abstract

United States Customs and Border Patrol need to cope with an ever-increasing volume of cargo data. This data was derived from sources such as shipping manifests and United States Coast Guard Notice of Arrival documents, which contain details on the *who, what, when*, and *where* of shipments and are stored in a large Oracle database. This database was and is used by intelligence analysts to detect suspicious cargoes. As the size of the database continues to grow rapidly, the queries which are used to extract the data have slowed down. Since the analysts need timely responses in order to be effective, our project's goal is to improve this query performance by applying visual analytic techniques to illuminate its causes and then developing remedies.

Technical Approach

Data collection methods used were compiling & sorting DB and expert consultation. The analytic method used was case studies. The nature of research was hybrid applied-consultation.

Milestones

The deliverables are all in the form of technical reports delivered to customer.

Project Outcomes (Results, Accomplishments, Conclusions)

Other outcomes/impacts

The VACCINE team conducted an initial survey of the sample data and database descriptions provided to them in April 2010. So far they have identified several structural areas, which appear to be causing problems and need to be investigated in detail to determine their impact on performance.

Seed Project Competition

VACCINE has dedicated approximately 9% of our annual funds to competitively award peer-reviewed seed project proposals to foster the development of new MDRPs, including education-oriented MDRPs. Successful seed projects will transition into our MDRP portfolio and help evolve research thrusts within each pillar as well. The VACCINE 2010 Seed Project competition identified innovative research and educational needs in the following visual analytic areas of homeland security:

- Video Surveillance Visual Analytics
- Introducing Sustainable Visual Analytics into Command Center Environments
- Mobile Imaging and Light-Weight Visual Analytics for In-Field Analytics
- Context-aware Mobile Visual Analytics for Emergency Response
- MSI Education and Programs of Study
- Visual Analytics for the DHS Centers of Excellence
- Cognitive Science Applied to Visual Analytics

Four seed projects were selected from an applicant pool of 25 proposals. The criteria used to evaluate these proposals were based on:

1. Responsiveness, Technical Merit and Quality
2. VACCINE Mission-Related Significance
3. Qualifications of Investigators
4. Budget and Productive Use of Federal Resources
5. Minority Serving Institution Engagement
6. Results Transition (10%)

The Seed project competition was held in year 1. However, funds were distributed to begin the seed projects in year 2.

SEED 1: Investigation of the Impacts of Visualization Literacy on the Performances of Visual Analysis

Team

VACCINE:

Purdue University: Ji Soo Yi, Sung-Hee Kim

Simon Fraser University: Brian Fisher

University of British Columbia: Linda Kaastra

Abstract

The rapid evolution of visual analytics (VA) has created many innovative visualization techniques. These novel visualizations provide some visual analysts with additional capabilities and insights, but others perceive them unfamiliar and difficult to use due to their novelty. One factor that could explain individual differences would be the level of prior knowledge about visualizations, which we called "visualization literacy." For example, one who knows parallel coordinates may easily find an interesting multidimensional pattern from a parallel coordinate view, but others without the knowledge will fail to gain the same insight. In spite of this obviousness, visualization literacy remains as one of uncharted areas in visual analytics research. We have not clearly understood how low or high visualization literacy impact on visualization performance, in what situations low visualization literacy becomes a more limiting factor than other innate individual differences (e.g., cognitive style and personal traits), and how to measure visualization literacy. This lack of understanding of visualization literacy increases the unidentifiable variances in evaluation studies and slows down development of systematic methods to educate visual analytics users.

Thus, the aim of this proposed study is to better understand the influences of visualization literacy on the performance of visual analysts (visualization performance, henceforth) and develop initial cognitive model to predict the performance of visual analysts. Eventually, the results of the current project could be incorporated into a comprehensive mathematical/computational cognitive model and a measurement tool of visualization literacy. This is a collaborative project between two institutes: Purdue University and Simon Fraser University. The former will contribute the experiences in creating taxonomy of visualizations, development workforce for interactive visualization systems, and infrastructure to conduct a controlled lab experiment. The latter will contribute its expertise in cognitive science and experiences in conducting field studies in a more realistic environment.

Milestones

- ✓ Comprehensive literature review.
- ✓ IRB approval.
- ✓ A field study to understand the impacts of visualization literacy.
- ✓ Develop a cognitive model to predict the visualization performances.
- ✓ Final report.

SEED 2: Gang Graffiti Recognition and Analysis Using a Mobile Telephone

Team

VACCINE:

Purdue University: Mireille Boutin, Ed Delp

Abstract

The United States Department of Justice estimates there are approximately 30,000 gangs, with 800,000 members, impacting 2,500 communities across the United States. Our goal in this project is to provide the police and other agencies with a tool to allow them to monitor gang activity by collecting images of gang graffiti for real-time interpretation and archiving.

Graffiti is a form of expression used by many subcultures and groups in the US. There are different types of graffitis which are associated with different groups or cultures. For example, "tagging" is a graffiti type consisting of a stylized signature and is associated with youths. It is an important component of the Hip Hop culture and is very widespread today, so much so that it is considered an epidemic in many urban areas. "Gang Graffiti" is another type of graffiti. A typical gang graffiti image is shown below.

Graffiti is often perceived as a form of vandalism. However, for law enforcement agencies, it is a valuable source of information. Gangs, for example, use graffiti as part of their fight for dominance. They use it to mark their territory as well as challenge rival gang's territories. Thus knowing the location of the graffiti of the different gang present in a city enables one to estimate the extent of their respective territories. Gangs also use graffiti to express their individualities, associations, and enemies. Thus tracking the evolution of gang graffitis over time can be used to understand gang's dynamics and social changes. Graffiti is also an important means of communication used by gangs, for example to make threats against the state or some other group; understanding this communication can be vital in some cases.

However, the symbols used in graffiti are difficult to interpret by outsiders. For example, a chicken and a slice of bread is a sign of disrespect used to challenge another gang. To assist law enforcement officers in making sense of the different graffitis they encounter, taxonomies have been developed. For example, a type of letters called "placas" is known to be associated with Hispanic gangs while the pitbull is a symbol associated with blood gangs. Unfortunately,

because of the changing nature of the symbolism used in this media as well as its high complexity, taxonomies often do not convey the full extent of a graffiti's message. Moreover, browsing a graffiti's taxonomy is made difficult by the graphical nature of its elements.

The long term goal of this project is to develop and implement a method for recognizing and analyzing gang graffitis using a commercially available mobile telephone. This will be accomplished by taking a color picture of the graffiti using the camera embedded in the device. The user will then be able to select either the full graffiti or a specific element of a graffiti and search for similar graffitis in a database. The most similar matches will be displayed and the user will be able to browse the information relating to each match using a simple graphical user interface. The system will also propose categories for the given graffiti/element along with the meaning of these categories. For example, identifying a graffiti as a "tagging" will enable a law enforcement officer interested in gang activities to be able to dismiss the graffiti as irrelevant. When appropriate (e.g., for the spatial distribution of a given gang graffiti), the information will be provided graphically and the navigation will be driven by touch and gestures.

We use as a starting point our experience in developing a language translation system that uses mobile telephones for real-time translational of signs from various languages (e.g. Spanish and Arabic) into English. This system has been deployed on an Apple iPhone. In this one year effort, we will develop a proof of concept for this system, which will consist of a small graffiti dictionary along with a first version of our proposed recognition and analysis software.

Milestones

- ✓ Develop methods for graffiti shape recognition.
- ✓ Develop graffiti dictionary.
- ✓ Initial hardware deployment.
- ✓ Delivery of working prototype.

SEED 3: Uncertainty Visualization in Video Visual Analytics

Team

VACCINE:

Swansea University: Min Chen

Purdue University: David Ebert

Abstract

This seed project addresses the problem of uncertainty visualization in *Video Visual Analysis* (VVA). A VVA pipeline differs from a *traditional automated video analysis* (AVA) pipeline in several ways: (i) VVA focuses on the depiction of important information in the videos, whilst AVA focuses on making analytical decisions using machine intelligence; (ii) VVA provides an integrated analysis and visualization pipeline with feedback loops; whilst AVA usually does not rely on visualization, or provides only primitive visual annotation of videos. In both VVA and AVA pipelines, there are a large number of video processing components, such as filters for morphological operations, convolution, segmentation and recognition. In addition to typical imagery, geometrical and annotative results, these filters also result in a huge amount of information about the uncertainty levels (and processing errors). While some of such information may be part of the decision mechanism (typically through coarse quantization), most information would be discarded at various stages of the pipeline. Without such information, adjusting various parameters for video processing components is almost a black art.

Raw uncertainty information can be pixel-based, object based and frame-based, and typically of a time-varying nature. It is thus desirable to make the uncertainty information available to the viewers visually. However, video visualization components in a VVA pipeline usually lack in both information bandwidth and distinguishable visual channels. The former refers to the limited screen resolution to handle a large amount of temporal data. The latter refers to the factor that most effective visual channels usually have already been used by the traditional video visualization without uncertainty information. For example, the *VideoPerpetuoGram* (VPG) proposed by Botchen et al. [TVCG, 2008] uses the texture channel for keyframes, and streamlines for object tracking, making uncertainty visualization a challenging issue. In this seed project, we will build upon the expertise in our partner universities in video visual analytics, non-photorealistic and illustrative visualization, and perception in visualization. We will investigate into the following aspects of uncertainty visualization in video visual analytics:

(a) Efficient uncertainty representation, computation, and storage in the VVA pipeline.
(b) Visual channels for integrated VVA and uncertainty visualization.
(c) Bandwidth for multiple-view VVA and uncertainty visualization.
(d) The role of uncertainty visualization in the feedback loop of the VVA pipeline.
(e) Evaluation of the usefulness and effectiveness of uncertainty visualization in VVA.

Milestones

- ✓ Literature review and mapping out research scope (a technical report).
- ✓ Design specification of a demo system.
- ✓ Development of demo system.
- ✓ Evaluation of demo system.
- ✓ Preparing the final report and publication.
- ✓ Exchanging visits between Swansea and Purdue researchers.

SEED 4: An Advanced Data Integration and Visual Analytics Framework for Enhancing Emergency Response Situation Reports with Multi-agency, Multi-partner Multimedia Data

Team

VACCINE:

Florida International University: Shu-Ching Chen, Tao Li, Steve Luis

Abstract

Responders in the field (both public and private participants) are capturing footage, pictures and video, of the disaster area with mobile phones. This data is being collected, but is not integrated in the incidence command systems where situation reports, incidence action plans, etc. are being held. When these materials are entered into such systems, it's done so in a manual way which is time consuming and requires careful human review and management.

The team proposes to use advance data integration and visual analytics techniques to analyze (1) situation reports, incident action plans, and other types of responder reports and (2) pictures and video clips captured in the field and automatically augment/link the reports directly to relevant multimedia content obtained in the field. Similarly, such multimedia content is not linked to situation reports which can help to better communicate the interpretation of the multimedia content in the context of the latest situation reports. Data mining techniques will be used to identify specific disaster keywords and will be used to identify the documents.

Milestones

- ✓ Development of novel techniques for integrating clustering and dimensionality reduction for visualization.
- ✓ Development of outlier removal and principal component subspace projections algorithms.
- ✓ Development of Multiple Kernel Learning (MKL) methods for multi-source data fusion.
- ✓ Automatic selection of representative principal components.
- ✓ Development of a prototype system with applications to emergency response.
- ✓ Development of techniques for incorporating domain knowledge and user interaction.
- ✓ Development of disaster-situation detection classifiers.
- ✓ Coupling with Translated Learning.

VI. Education Scope

During year one, VACCINE's educational initiatives have spanned the entire career development pipeline ranging from K-12 programs through undergraduate and graduate level work, to professional education and training programs. The mission is to educate current Homeland Security stakeholders and the next generation of talent and enable them to make effective decisions from the mass of multisource, multimedia data they will face in their careers.

VACCINE educational efforts have targeted four main areas: K-12 projects, undergraduate and graduate programs, professional training and minority serving institution (MSI) collaboration. In addition to the established programs, VACCINE has focused on many outreach activities such as attendance at conferences and presentations to groups in order to promote VACCINE efforts and get support and participation for VACCINE year two and beyond. The education group has also submitted several proposals to further increase funding opportunities and future program development. Throughout all of these efforts, VACCINE and CCICADA have maintained a collaborative relationship with monthly conference calls, joint attendance and presentations at conferences and shared development of material and content.

K-12 Programs

Middle School/High School Teacher Workshops

Teacher workshop development and planning has taken place during VACCINE year one. These workshops will be held in June 2010. The workshops will focus on introducing visualization and data analysis to middle school and high school teachers through homeland security examples, curriculum development and hands-on learning modules.

The goals of the workshop include:

1. Define visual analytics (VA) and teach participants how they can incorporate VA into their curriculum to enhance learning.
2. Explain how to identify misinformation within graphics, good & bad visualizations and good & bad data.
3. Show participants how to find online resources for further explorations.

High School Elective

A three class period module on visualization and analytics has been developed for high school juniors and seniors. The module was taught in September 2009 and again in February 2010 as part of an emergency management/homeland security elective course at Triton Central High School in Fairland, IN. The high school module is being modified for use during summer 2010 teacher workshops and then as part of an online teacher tutorial for middle school and high school teachers. This teacher tutorial will be available during the summer 2010. University of Houston Downtown and CCICADA partners are in line to use the materials once they are available.

Middle School Modules

Two middle school teachers from Granger, IN worked with VACCINE on the teacher workshop materials and presented tools they have used in their classrooms. VACCINE has also worked on the development of mini classroom modules that focus on use of visual analytics within existing curriculum. The most recent module uses visual analytics tools within a middle school classroom to study the spread of religion. It ties in the current media attitudes and religious perceptions that relate to terrorism and homeland security. Downloadable modules can be found under the K-12 education tab on the VACCINE website. www.visualanalytics-cci.org

Undergraduate/Graduate Studies

VACCINE is developing a course catalog for our partner schools and will expand upon this diverse set of visual analytic curricula, refining and integrating our existing efforts and identifying niche opportunities where future visual analytic and master's degree courses can be developed and offered. Planning is also underway for a one-day workshop on visual analytic courses and curricula in conjunction with the NVAC Consortium meeting on August 30, 2010.

HS-STEM Career Development

With Dr. David Ebert as the PI, a competitive Homeland Security - Science, Technology, Engineering, and Mathematics (HS-STEM) Career Development program has been established at Purdue to support students interested in pursuing homeland security related careers. Eleven students have been funded through the HS-STEM Career Development program. These students meet monthly for networking lunches where students share their research. HS-STEM students have also had the opportunity to present research topics in high school classrooms. Five high school classroom presentations have been done during the 2009/2010 school year. Three HS-STEM students have graduated during the 2009/2010 school year with PhDs and gone on to continue work in homeland security fields. VACCINE hopes to expand this program by obtaining an HS-STEM 2010 award.

Visual Analytic Summer Program (VASP)

VASP is a hands-on, 10-week summer learning program that introduces students to visual and data analytics projects. The VASP program is part of the long time established Summer Undergraduate Research Fellowship (SURF) program at Purdue. VASP students participate along with SURF students in a variety of activities that include seminars and workshops on graduate school, current events, and social activities. All students present a research poster at the end of the summer with awards given to the best presentations. Four students participated in VASP during summer 2009. Two of these students had their posters finish in the top 10 out of all 115 SURF poster presentations. VACCINE welcomes undergraduate students from all universities who are interested in visual analytics topics to apply for this program. Eleven students will be chosen to work with VACCINE during the summer of 2010.

Professional Training

VACCINE and CCICADA will co-host RECONNECT faculty workshops during the summer 2010 at University of Southern California. These workshops offer the opportunity for junior faculty as well as mid-level and senior faculty, and Homeland Security Professionals to consider research questions in a new area of the mathematical/computer sciences. The topic for the USC will focus on 'Text Extraction.

Planning is underway for a September workshop on visual analytics for public safety professionals. The workshop will be held at the Justice Institute of British Columbia. Participants will include first responders from Canada and the US as well as VACCINE partners. The purpose of the workshop is to define the requirements that will lead to developing academic programs and relevant technology utilizing visual analytics in the public safety environment.

VACCINE has met with FLETC, the Indiana Law Enforcement Academy, and the Justice Institute of British Columbia to discuss training needs and ways that VACCINE and the practitioners can work together. Some initial discussions include: faculty partnering with VACCINE faculty for research, developing a basic course on visual analytics and how it can be an essential tool for law enforcement, and continuing development of mobile applications.

Visual Analytics Digital Library (VADL)

Georgia Tech has been maintaining VADL, an online resource center and leading international repository for visual analytics educational and training material, as well as practical lessons-learned and best practices for Command, Control and Interoperability (CCI) material. VACCINE would like to expand these materials and have them available for all VACCINE partner faculty and students as well as those interested in visual analytic studies. **http://vadl.cc.gatech.edu**

MSI Collaboration

VACCINE began to raise awareness at Minority Serving Institutions (MSI) about its mission by offering research talks at conferences such as Society for Advancement of Chicanos and Native Americans in Science (SACNAS), Grace Hopper Women in Computing, Hispanic Engineer National Achievement Awards Conference (HENAAC), TeraGrid, Students and Technology in Academia, Research and Service (STARS Alliance), Association of Computer/Information Sciences and Engineering Departments at Minority Institutions (ADMI) /African-American Researchers in Computing Sciences (AARCS). In addition, VACCINE and CCICADA co-facilitated a workshop on MSI and Centers of Excellence (COE) relationships during the Education Day at the DHS University Summit in March 2010. COE education reps learned basic facts about MSIs as well as ways to successfully work with and integrate educational programs.

VACCINE has collaborated with Jackson State University to develop an affinity research pod. These research pods foster an innovative educational and research partnership between Research 1(R1) Universities and the minority-serving institutions and target visual analytics and homeland

security applications. Dr. David Ebert traveled to JSU in January to meet with Drs. Loretta Moore and Jacqueline Jackson to kick off the research pod. Two undergraduate students and one master's student will live at Purdue University over the summer 2010 to further develop the research. They will then continue the project at JSU during the 2010-2011 school year. University of Houston Downtown and Penn State will begin the second research pod during year two of VACCINE.

Dr. Erin Hodgess from the University of Houston Downtown has been awarded a grant for the DHS MSI Summer Research Program. Professor Hodgess and one of her students will spend the summer with VACCINE faculty and students at Purdue University doing research related to healthcare data analysis.

Additional MSI collaboration during VACCINE year one have included, the offering a Special Topics Visual Analytics Course at JSU during the spring semester 2010 at Jackson State University and meetings with Navajo Technical College to include them as a funded VACCINE partner.

Additional Education Engagement Activities

Presentations to Groups and Conferences

Dept of Homeland Security Indiana Higher Ed Consortium Annual Conference – October 2010
Regenstrief for Healthcare Engineering (RCHE) Monthly Stakeholder meeting – March 2010
Purdue Discovery Park Monthly Engagement Meeting – February 2010
Hoosier Association of Science Teachers International (HASTI) – February 2010
Capital Forum Indiana – January 2010

Proposal Submission

HS-STEM Career Development Grant 2010
Innovative Educational Opportunities BAA 09-07- The Learning Hub
CAL State Dominquez Hills – submitted two VACCINE seed grant proposals

VII. Education Programs and Outcomes

EP 1: Leadership and Coordination

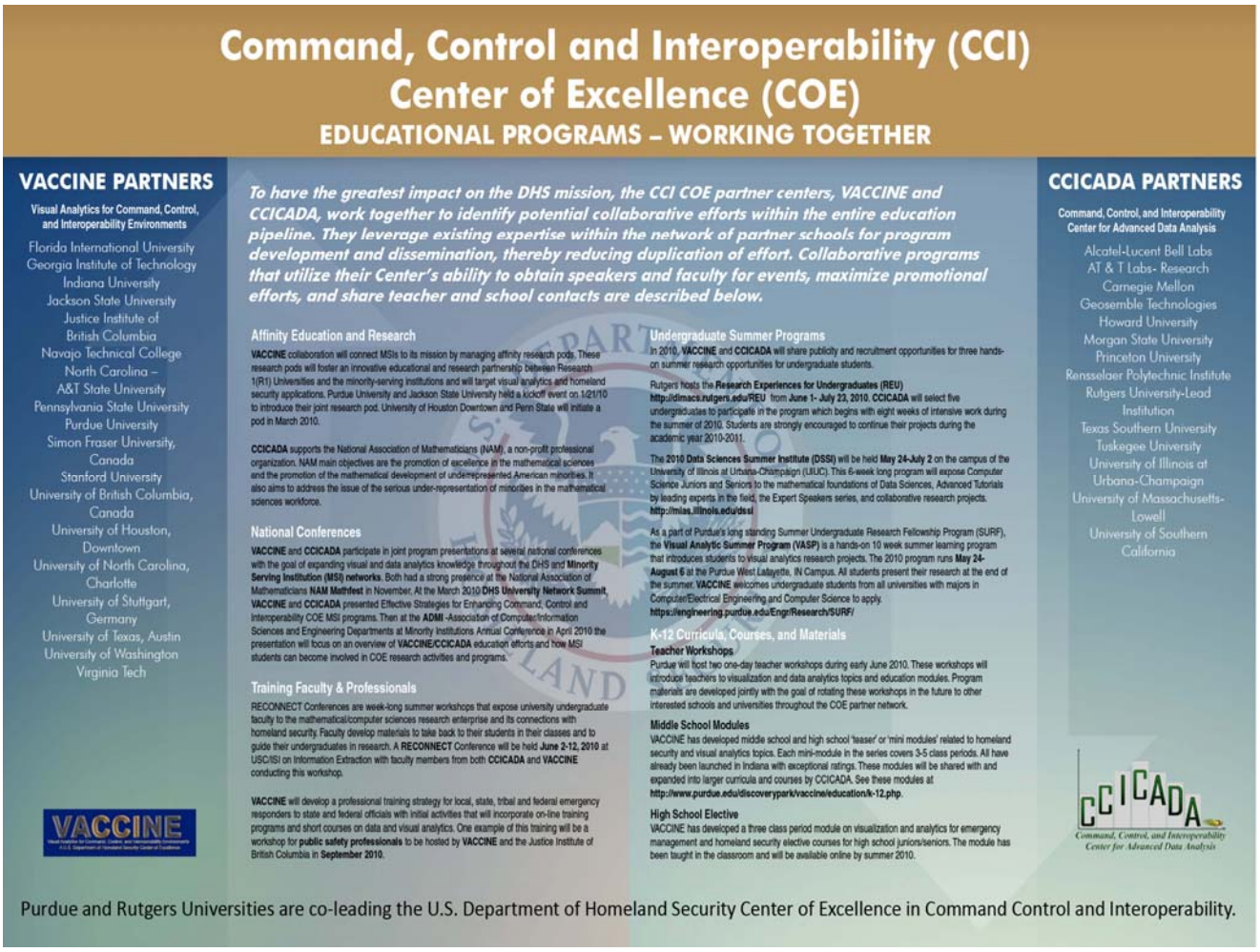
Team

VACCINE:
Purdue University: Marti Burns
Georgia Institute of Technology: John Stasko
University of Houston-Downtown: Richard Alo, Ping Chen, Erin Hodgess
Indiana University: Maureen Biggers
Jackson State University: Loretta Moore, Jacqueline Jackson

CCICADA:
Rutgers University: Midge Cozzens

Abstract

The main objective during the initial period was to develop an innovative and coherent education plan that spans across the individual members' schools of the VACCINE center that also work in coordination with the Rutgers-led Data Analysis Center. The team sought to develop curricula and courses for visual analytics that could be used as examples and templates at different schools. Working together with colleagues at Purdue, Dr. Stasko oversaw the strategic mission and efforts in the educational programs of the VACCINE Center. He was involved with all of the educational programs and missions, and represents the Center at various activities. His particular responsibilities during the period included participating and teaching in the RECONNECT summer school session led by the CICCADA Center, oversaw the Affinity Education Research Groups, maintained the Visual Analytics Digital Library, and participated at the education meeting of the DHS University Summit.



Milestones

Milestones for VACCINE Period 1- Leadership and Coordination	07/01/09 – 03/31/10 Quarterly			
Teaching in RECONNECT seminar		X		
Monthly conference calls		X	X	X
Visit Purdue to work on education plan		X		
Develop education brochure		X		
Development of education program document			X	
Plan for AERG research pod initiation				X
Presented at DHS Summit				X
Helped develop HUB zero concept paper		X		
Helped develop/submit HUB zero proposal				X
Developed / submitted Scholarship Proposal				X
Collaboration calls with Rutgers		X	X	X

Program Outcomes (Results, Accomplishments, Conclusions)

Other outcomes/impacts

Dr. Stasko led the strategic educational initiatives of VACCINE working closely with Marti Burns of Purdue, the Assistant Director for Engagement and Education. They produced overview materials for the educational aspects of the center such as brochures and posters.

In August, the Rutgers component of the CCI Center held a weeklong RECONNECT summer school session for professors and researchers, who may be from other areas but that want to learn more about data analysis and visual analytics. Dr. Stasko represented the VACCINE Center at the session and taught two days about visual analytics and information visualization. He also introduced and described the Jigsaw system and gave each of the attendees a copy of the system to use in an example analytic exercise.

Dr. Stasko also coordinated educational efforts for the CCI Center with colleagues in the Rutgers-led data analysis component, specifically Dr. Midge Cozzens of Rutgers. The two groups held monthly conference calls to discuss plans and joint activities.

Dr. Stasko worked with Maureen Biggers and Richard Aló to develop educational activities in coordination with minority-serving schools in our component of the center. They created two instances of the Affinity Education Research Pod concept, one jointly between Jackson State University and Purdue University, and one jointly between University of Houston-Downtown and Penn State University.

In March, the VACCINE and CICCADA groups collaborated to jointly plan and run the morning sessions of the Education meeting at the DHS University Summit. The focus of the meeting was on MSI partnership and collaboration within the Centers. The VACCINE CCI group conducted a series of panels and activities to heighten awareness and knowledge about MSI universities and programs. Dr. Stasko spoke on one of the panels.

Additionally, the Center for Computational Science and Advanced Distributed Simulation (CCSDS) at the UHD initiated VACCINE research during the past year with faculty members Richard A. Aló (lead), Ping Chen and Erin Hodgess (all of UHD) and Mohsen Beheshti (Chair of Computer Science at California State University Dominguez Hills). The team created four Affinity Research Groups at UHD and one at CSUDH each consisted of 3 students and one faculty lead chosen from the faculty listed here. Students and faculty submitted their joint research papers and posters to international and national conferences as listed. International acceptance is in Beijing, China (2 papers), Uppsala, Sweden (1 paper) and Cambridge England (1 paper) and National SIGSCE Conference in Milwaukee, Wisconsin.

EP 2: MSI Collaboration

Team

VACCINE:
Purdue University: David Ebert, Marti Burns
Georgia Institute of Technology: John Stasko
University of Houston-Downtown: Richard Alo
Florida International University: Shu-Ching Chen
Jackson State University: Loretta Moore, Jacqueline Jackson

CCICADA:
Rutgers University: Midge Cozzens

Abstract

The Center for Computational Science and Advanced Distributed Simulation (CCSDS) at UHD, Richard A. Aló, lead, began MSI outreach to the following: California State University Dominguez Hills, West Houston Center for Science and Engineering of Houston Community College, Navajo Technical College, American Indian Higher Education Consortium, California State University Long Beach and Long Beach Community college. In addition the team recommended Katherine Mitchell to the VACCINE Advisory Board, as the AIHEC representative. Also the team has assisted in the development and submission of Scholarship Proposal for WHC of HCC, HubZero proposal for UHD and other MSI participations, and UHD scholarship program. The team provided two minority faculty candidates and students for the DHS faculty /student summer development program.

Milestones

Milestones for VACCINE Period 1 – MSI Collaboration	**07/01/09 – 03/31/10**			
	Quarterly			
SACNAS conference VACCINE poster/brochures		X		
Affinity pod training and monthly teleconference coaching sessions		X	X	X
Distinguished lecture on new MSI campus				X
SC 09 conference - November 2009		X		X
Grace Hopper Women in Computing Oct 2009		X		
HENAAC October 2009		X		
TerraGrid 10 June 2010				X
STARS conference VACCINE poster		X		
Affinity pod training and monthly teleconference coaching sessions		X	X	X
Distinguished lecture on new HBCU campus				X
ADMI/AARCS conference - April 2010				X
Meet with R1 Research Partner to discuss research project goals and objectives		X	X	
Selection of student participants		X		
Affinity pod training and project kickoff event		X	X	

Weekly pod meetings		X	X	X
Monthly teleconference meetings		X	X	X
ADMI/AARCS conference				X

Program Outcomes (Results, Accomplishments, Conclusions)

List of all Publications

Conference Publications

1. Aló*, Richard A., Diane Baxter, Karl Barnes, Al Kuslikis, Geoffrey Fox, Alex Ramirez; Advancing Computational Science, Visualization and Homeland Security Research/ Education at Minority Serving Institutions National Model Promoted/ Implemented by MSI-CIEC (Minority Serving Institutions-CyberInfrastructure Empowerment Coalition); International Conference on Computational Science, ICCS, 2010. Amsterdam, June, 2010.
2. Aló*, Richard A., Diane Baxter, Karl Barnes, Al Kuslikis, Geoffrey Fox, Alex Ramirez; A Model for LACCEI: Minority Serving Institutions and CyberInfrastructure Research/ Education Minority Serving Institutions-CyberInfrastructure Empowerment Coalition-MSI-CIEC; Latin American and Caribbean Consortium for Engineering Institutions, International Conference, LACCEI 2010, Arequipa, Peru, June, 2010.
3. A. Tran, C. Bowes, D. Brown, P. Chen, M. Choly, W. Ding, 'TreeMatch: A Fully Unsupervised WSD System Using Dependency Knowledge on a Specific Domain', SemEval 2010 Workshop with the 48th Annual Meeting of the Association for Computational Linguistics (ACL), July, 2010. Uppsala, Sweden.
4. P. Chen, A. Barrera, C. Rhodes, "Semantic Analysis of Free Text and its Application on Automatically Assigning ICD-9-CM Codes to Patient Records", The 9th IEEE International Conference on Cognitive Informatics, July 7-9, 2010, Beijing, China. (Acceptance rate: 29%).
5. P. Chen, R. Alo', J. Rundell, "From Language to Vision: A Case Study of Text Animation", The 9th International Conference on Artificial Intelligence, Knowledge Engineering and Databases, Cambridge, UK, Feb., 2010.

Other reports

1. P. Chen, N. Ozoka, R. Ortiz, A. Tran, D. Brown, "Word Sense Distribution in a Web Corpus," The 9th IEEE International Conference on Cognitive Informatics, July 7-9, 2010, Beijing, China. (Poster, acceptance rate: 32%).
2. P. Chen, W. Garcia, "Hypothesis Generation and Data Quality Assessment through Association Mining", The 9th IEEE International Conference on Cognitive Informatics, July 7-9, 2010, Beijing, China. (Acceptance rate: 29%).
3. P. Chen, I. Chen, R. Verma, A. Tran*, "An Undergraduate Data Mining Course Integrated with Research and Industry Projects", SIGCSE 2010, Poster session, March, 2010.

Internship Programs

Summer Internships: All three Jackson State University students (Chicora Chandler, Jotham Greer, and Ashley Rhodes) are participating in summer internships at Purdue and will be

integrated into the newly established Visual Analytic Summer Program based on the Summer Undergraduate Research Foundation framework.

Other outcomes/impacts

- In August, 2010, we announced a new position, Assistant Director of MSI Education, Richard Alo, from the University of Houston-Downtown. In this unique position, he has strategic oversight for development and implementation of VACCINE outreach initiatives involving minority serving institutions.
- The JSU VACCINE project Kickoff was held January 21, 2010. Dr. David Ebert visited JSU and provided three lectures. The first lecture was an open forum attended by undergraduate students, graduate students, and faculty from a number of disciplines. The second lecture was a presentation and round table discussion with the JSU Pod members (Dr. Loretta Moore, Dr. Jacqueline Jackson, Chicora Chandler, Jotham Greer, and Ashley Rhodes). The third lecture was given to the Visual Analytics class and guests. Dr. Ebert also visited and held discussions with other DHS Center Directors and Project PIs. During the Project Kickoff, each student identified a specific research focus.
- 2010 ADMI/A4RC Conference: The 2010 ADMI/A4RC conference was held on April 8-11, 2010 in Jackson, MS. VACCINE was spotlighted in two conference sessions and the evening Social Networking Event. Undergraduate students Chicora Chandler and Jotham Greer participated in the conference session that spotlighted VACCINE's research opportunities for students.
- Maureen Biggers, from Indiana University, was involved in the SACNAS conference poster/brochures, affinity pod training and the monthly teleconference coaching sessions, STARS conference poster, ADMI/AARCS conference in April 2010, and the selection of student participants. She flew to Jackson State University and met face to face with each student in the pod to talk about how thing were evolving as they got ready to move to Purdue for a summer of research.
- The Pennsylvania State University, Jackson State University, Indiana University, Purdue University, Oak Ridge, Northeastern University, and Rutgers teamed to establish Affinity Research Pods/Groups and to provide faculty candidates for DHS summer faculty /student development.
- Jackson State University conducted a special topics visual analytics course. The course was offered during the Spring semester of 2010 for graduate special topics in Visual Analytics. Eight Computer Science students registered for the course. Dr. Ebert taught the first lecture and gave an introduction to Visual Analytics. The book used for the course was "Illuminating the Path: The Research and Development Agenda for Visual Analytics". Students reviewed papers and gave oral presentations on various topics from the book. For the term project, students downloaded/evaluated a visual analytics application, presented their findings and created a technical report of the work.

- VACCINE has been working to add the Navajo Technical College as a funded partner beginning in Year 2. The strategy is to help NTC transition to a 4-year degree program with an emphasis in visual analytics. The following objectives will aid this effort:

    - Promote 'best practices' to assist in NTC retention, recruitment and motivational activities for their new four year computing degree programs.
    - Assist in further development of Cyber Infrastructure-enabled distributed research and education network providing visualization education and research opportunities to NTC faculty and students.
    - With above, proceed to develop an Affinity Research Pod (ARP) to join the research of the ARPs within the VACCINE group at Center for Computational Science and Advanced Distributed Simulation (*CCSDS*/ University of Houston Downtown UHD.
    - Assist to develop and implement curricula and courses for visual analytics that are available through the VACCINE network such as the Visual Analytics Digital Library (VADL).

EP 3: K-12 Programs

Team

VACCINE:
Purdue University: David Ebert, Marti Burns
Discovery Middle School: James Howard, Kent Mikel
Triton Central High School: Steve Elder

CCICADA:
Rutgers University: Midge Cozzens

Abstract

VACCINE's mission is to educate the next generation of talent through the VACCINE K-12 programs which include; the development of middle school and high school teacher workshops, a high school teacher tutorial for the high school visual analytic module, and the development of middle school modules. In addition, outreach for the K-12 programs included; the attendance at conferences and events, as well as classroom and organization presentations.

Milestones

Milestones for K-12 Teacher Workshops	07/01/09 – 03/31/10			
	Quarterly			
Meet with Director of Discovery Park K-12 Education to discuss DP collaboration – Workshop coordination		X	X	X
Quarterly Planning Meetings			X	X
Discuss Collaboration with Rutgers – monthly conference calls		X	X	X
Workshops offered (2) June 3 & 4				
Milestones for High School Elective				
Meet with Steve Elder – planning Triton Central High School		X	X	X
Regular Planning Meetings / conference calls		X	X	
Present module in elective course classroom – two times (Triton Central High School on 9/21-9/23and 2/22-2/24)		X		X
HS-STEM Students present research in High School Classrooms 11/24/09, 10/29/10 and Capital Forum 1/25/10		X		X
Evaluate, enhance, expand high school module through online teacher tutorial				X
Discussions on distribution to Houston/ other schools				X
Milestones for Middle School Modules				
Meet with teachers to discuss existing modules for updates and visualization additions		X	X	X
Develop new visualization modules				X
Planning discussions with Rutgers for new school opportunities / new modules		X	X	X
Booth at HASTI Convention – Indianapolis, Feb 4&5				X
Incorporate new visualization (2) modules into curriculum				X

Program Outcomes (Results, Accomplishments, Conclusions)

Other outcomes/impacts

Two one-day middle school/high school teacher workshops on visualization and data analysis will be held on June 3 & 4 at Discovery Park on Purdue University's main campus. Planning and development of content for these workshops took place during this first period of VACCINE.

A three class period high school module on visual analytics has been developed and taught at Triton Central High School during both the Fall Semester 2009 (9/21-9/24) and the Spring Semester 2010 (2/22-2/24). The module was taught to a class of juniors and seniors enrolled in a Homeland Security /Emergency Management elective course developed by Steve Elder. Goals of the module were to introduce the topic of visual analytics and tie it to examples relevant to their course work, such as use of VA to improve decision making for the emergency management community. Work assignments were tied to their class tabletop exercise scenario that dealt with an Anhydrous Ammonia explosion at a local food cold storage plant. Materials

from this high school module were converted into a teacher tutorial that were used during the June workshops and then modified and placed online to be initially shared with Houston Prep.

A big piece of the K-12 Education Program includes outreach to various organizations to build relationships and collaboration for future program development. Marti Burns attended several conferences and offered presentations to groups to introduce VACCINE and distribute materials regarding VACCINE education programs.

- Capital Forum Indiana – January 31, 2010 Indianapolis. Capital Forum Indiana is a civic engagement and international education program that focuses on current global issues and U.S Foreign Policy. This program is intended for high school teachers and students across Indiana, and is part of a larger, multi-state initiative of *The Choices Program* at Brown University.
- Hoosier Science Teachers Association Conference – February 3&4, 2010, Indianapolis. VACCINE participated in the Discovery Park booth at the conference.
- Indiana Department of Homeland Security Higher Education Consortium Annual Conference – October 14, 2010, Indianapolis. Tim Collins, Marti Burns. General VACCINE and Education presentation.
- Regenstrief Center for Healthcare Engineering – Marti Burns, Purdue University. General VACCINE and Education presentation, March 1, 2010
- Discovery Park Engagement Group – Marti Burns, Purdue University. General VACCINE and Education presentation, February 2, 2010.
- Triton Central High School: Marti Burns, Shawn McKay, Karla Combs October 19, 2010. General VACCINE presentation and HS STEM research.
- West Lafayette Junior High School: Marti Burns, Bryan Sims November 24, 2010. General VACCINE presentation and HS STEM research.

EP 4: Undergraduate Visual Analytics Summer Program (VASP)

Team

VACCINE:
Purdue University: Marti Burns, Ross Maciejewski
Jackson State University: Jacqueline Jackson

Abstract

As a part of Purdue's long standing Summer Undergraduate Research Fellowship Program (SURF), VASP is a hands-on 10 week summer learning program that introduces students to visual analytics research projects. All students present their research at the end of the summer. VASP 2009 hosted 4 students.

Technical Approach

PURDUE UNIVERSITY School of Electrical and Computer Engineering PURVAC

SURF

Health Ecotopes through K-Means Clustering

Michael A. Mitchell
Ross Maciejewski
David Ebert, PhD

Introduction

In this project an attempt to stratify the state of Indiana into health ecotopes, or regions of similar disease symptom rates, was undertaken. This method uses kernel density estimation to take a number of discrete symptom reports and turn them into a continuous probability density function for each symptom. These symptom PDFs are then placed orthogonally into a data-space and clustered into groups with similar symptom rates using K-Means clustering.

Kernel Density Estimation

Because of the geographically discrete nature of disease symptom reports a way to extrapolate the data to the entirety of the state is needed. Kernel density estimation provides a way by estimating the probability density function from a sample of the population. It works by summing an instance of a weighting function kernel, in this case a standard Gaussian distribution, over every discrete sample.

PURDUE ENGINEERING intel Education

K-Means Clustering

K-Means Algorithm

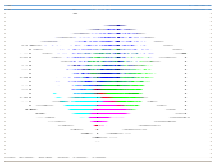
1) Randomly pick k cluster centers

2) Assign each point to the nearest cluster center

3) Recalculate cluster centers as the mean of all data members

4) Repeat steps 2 and 3 until no points switch cluster assignment

K-Means clustering is a greedy algorithm that iterates through data points placing them into k number of groups according to an "is like" classification system.

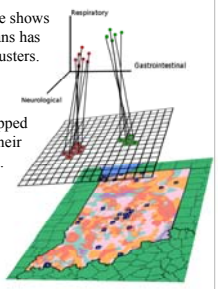


Mapping clusters onto geographic space

This top portion of the figure shows the data-space where K-means has grouped the data into two clusters.

The data points are then mapped onto a matrix according to their original geographic location.

Then the cells of the matrix are colored in according to the cluster they belong to and ecotopes form.



Results

The results show that geographic health ecotopes do emerge when clustering over non-geographic symptom data. This suggests a link between symptom rates and geographic locations and shows that the presented method is a viable way to find and delineate these links into health ecotopes.

Conclusion and Future Work

While using K-Means clustering is a viable way to automatically find and delineate Health Ecotopes there are a few problems with the algorithm that must be addressed. Most notable is the fact that the numbers of clusters must be given to the algorithm a priori, a difficult task if one has not seen the data before nor has a good way to intuit the correct number of clusters.

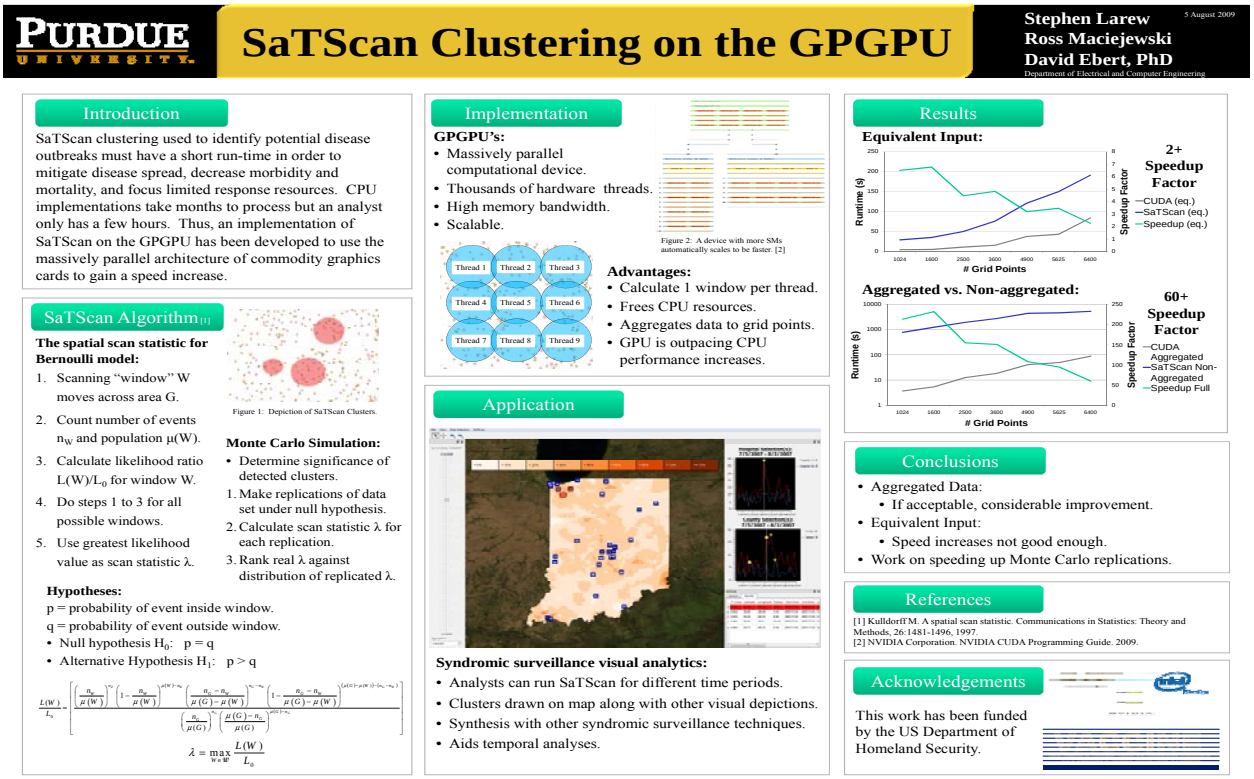
References

M. L. D. E. J. Carol L. Williams, William W. Hargrove. Agroecoregionalization of iowa using multivariate geographical clustering.
W. W. Hargrove and F. M. Hoffman. Potential of multivariate quantitative methods for delineation and visualization of ecoregions. Environmental Management, 34(?):S39–S60, ? 2005.
I. K. T. K. P. F. Ville Hautamaki, Svetlana Cherednichenko. Improving k-means by outlier removal.

Acknowledgements

This work has been funded by the US department of Homeland Security.

VACCINE



Milestones

Milestones for VACCINE Period 1 – Undergraduate Visual Analytics Summer Program	07/01/09 – 03/31/10 Quarterly			
Summer 2009 Pilot		X		
Evaluate 2009 program, suggestions/corrections for 2010		X	X	
Outreach to partnering schools for applications			X	X
2010 Applications & selection March, April 2010			X	
2010 Program begins – May 24, 2010				

Program Outcomes (Results, Accomplishments, Conclusions)

Other outcomes/impacts

Four students participated in VASP as a part of 2009. All 115 SURF students presented their research at the end of the summer. Ten awards were given for the best posters and two of those came from the VACCINE (VASP) students.

Selections have been made for the VASP 2010 program and eleven students will participate with VACCINE summer research projects. VACCINE and CCICADA have collaborated with their partner schools for the advertising of the program. Purdue University will have two students from Jackson State University participate in VASP 2010.

EP 5: Graduate Programs

Team

VACCINE:
Georgia Institute of Technology: John Stasko

Abstract

This component of the education mission involves graduate programs and education at the VACCINE Center. One specific goal of this project was to promote and coordinate graduate courses and programs at the VACCINE member universities and beyond. The Center gathered and maintained a list of related courses already being taught at our universities. Also, together with colleagues from the FODAVA Center and NVAC, the team created and publicized a "Body of Knowledge" for visual analytics.

Milestones

Milestones for VACCINE Period 1 – Graduate Programs	**07/01/09 – 03/31/10**			
	Quarterly			
Develop directory of visual analytics courses and education programs		X	X	X
Begin work on developing for a VA course and Masters degree		X	X	X
Develop body-of-knowledge for visual analytics		X	X	X
Plan an education workshop about visual analytics courses		X	X	X

Program Outcomes (Results, Accomplishments, Conclusions)

Other outcomes/impacts
The team gathered an initial list of courses and education programs at our member universities, but they have not made this into web pages for the VACCINE website yet. In the current educational climate of limited budgets and expansion, the team decided not to spend time developing a new visual analytics course or Masters degree.

We created the body-of-knowledge and posted it on a website at Georgia Institute of Technology for members of the visual analytics community to comment on. It is unclear what next steps this project will take.

We are planning to hold a one-day workshop on visual analytics courses and curricula in conjunction with the NVAC Consortium Meeting on August 30, 2010.

EP 6: HS-STEM Career Development Program

Team

VACCINE:
Purdue University: David Ebert, Marti Burns

Abstract

The HS-STEM Career Development program is a competitive program funded through DHS that was established at Purdue in 2007 under David Ebert. Purdue has offered fellowships to 7 graduate students in 2007 and 4 graduates and 1 undergraduate student in 2008. VACCINE has applied for a continuation of funding through a proposal for the HS-STEM 2010 program.

Purdue University
HS-STEM Career Development Program Activities
Annual Summary for 2009

- First Group of HS-STEM Students Graduate:
  - **Shawn McKay** graduated in December 2009 with a PhD in Mechanical Engineering. Shawn is currently employed as an engineering associate with RAND Corporation in California.
  - **Ross Maciejewski** graduated in December 2009 with a PhD in Electrical and Computer Engineering. Ross continues his research in visual analytics while seeking a faculty position in computer science or computer engineering at a university in the United States.
  - **Karla Combs** will complete her PhD in Biology in May 2010. While finishing her degree, Karla is working for the Illinois State Department of Health in the Emergency Preparedness section. She will be working on H1N1 planning for the State of Illinois and will be assisting in planning and executing full-scale exercises to test these plans later on in the year.
- Monthly HS-STEM Student Lunches:
  - Networking for students and advisors.
  - One student presents their research and activities each month.
  - Information sharing: poster/presentation opportunities, campus opportunities, DHS opportunities, etc.
- Presentations / Posters:
  - Students are encouraged to do presentations and present posters during campus activities as well as off site conferences (see individual summaries for specifics).
- DHS UP Conference in October 2009
  - Required for all recipients to attend. All four HS-STEM II recipients attended the 2009 October conference in Washington, DC – summary attached.
- Reporting:
  - Student and advisor summary reports due twice each year, May and January.

- Annual Report to DHS University Programs – January 2010.

- Goals for 2010:
  - Submit proposal for HS-STEM Career Development Program 2010 - due 2/16/2010.
  - Obtain funding for the HS-STEM 2010 Program and integrate new students with current students and activities.
  - Encourage collaboration among HS-STEM I and HS-STEM II students and advisors.
  - Increase knowledge of internships opportunities for undergraduates and job placement for both undergraduates and graduates.
  - Expand relationships with other Centers of Excellence and their HS-STEM programs.
  - Continue HS-STEM Career Development Program integration with K-12 efforts by having HS-STEM students present in three middle/ high schools to present their research. Five presentations were done in 2009 and students have been invited to return for the spring 2010 semester. Post these presentations on the VACCINE website for access by other teachers.
  - HS-STEM students plan to sponsor a speaker of campus wide interest during 2010.

Milestones

Milestones for VACCINE Period 1 – HS-STEM Career Development Program	**07/01/09 – 03/31/10**			
	Quarterly			
Present in High School Classrooms			X	X
Attend monthly UP conference Calls		X	X	X
Monthly lunches (highlight one student/month)		X	X	X
Four students attend OUP conference in DC			X	
Job recommendations / placements for students		X	X	X

Program Outcomes (Results, Accomplishments, Conclusions)

Other outcomes / impacts

Eleven students at Purdue University have been funded through the HS-STEM Career Development program. The program is by Dr. David Ebert, Director of VACCINE and managed by Marti Burns, Asst. Director of Engagement and Education. Three students have graduated with PhDs and have gone on to continue work in homeland security fields.

This program has been a great success for the students, the University and DHS. The first three graduates of the program have gone on to pursue careers in homeland security fields as is the goal of the program. Purdue hopes to receive funding for HS-STEM 2010 to continue the successful continuity of the program.

EP 7: Professional Training

Team

VACCINE:
Purdue University: Tim Collins, Marti Burns
Georgia Institute of Technology: John Stasko

CCICADA:
Rutgers University: Midge Cozzens

Abstract
VACCINE developed and continues to work on a professional training strategy for local, state, tribal and federal emergency responders to state and federal officials with initial activities that incorporate on-line training programs, workshops and short courses on data and visual analytics.

RECONNECT Workshop 2010

Extracting and Visualizing Information from Natural Language Text

Where: University of Southern California

When: June 6-12, 2010

About Reconnect: This CCICADA/VACCINE Summer Reconnect Workshop exposes faculty teaching undergraduates to the role of the mathematical sciences in homeland security and provides an opportunity to researchers in government or industry to learn about recent techniques in data analytics. Topics are presented in a weeklong series of lectures and activities; participants are involved in both research activities and in writing materials useful in the classroom or to share with their colleagues. Participants may develop materials for publication in either the CCICADA Technical Reports or the Educational Modules Series published by the DIMACS Center at Rutgers University.

Topic: Automatic identification and extraction of desired information from natural language text is increasingly used as a way to improve general purpose search and has a range of applications in medical informatics, business applications, and for the intelligence community. The input is

one or more texts in the domain in question, and the output is a database containing just the desired fields of information, extracted from the source material and formatted appropriately. Information extraction (IE) techniques have been developed since the early 1980s, and include finite state technology, pattern-based extraction, and appropriate machine learning methods. This week-long workshop will take participants from the early, simpler, methods through the modern ones, and will include theoretical and practical topics as well as hands-on exercises using software packages. The material is very relevant to the undergraduate classroom and to many applications. The lecturers are renowned experts in the various aspects of IE and its visualization, and have a long history of giving informative, engaging, and fun lectures.

Organizers: Eduard Hovy, USC/ISI; Zornitsa Kozareva, USC/ISI; Midge Cozzens, CCICADA.

Speakers: Eduard Hovy, USC/ISI; Zornitsa Kozareva, USC/ISI; Dan Roth, UIUC; and others TBA.

Registration fees, lodging, meals and travel: *Academic participants*: registration, lodging and meals will be provided through DHS funding. *Government participants*: $350. *For-Profit Corporation participants*: $500 (includes all meals from Sunday dinner to Saturday lunch). Limited funds are expected to be available to waive part or the entire registration fee and/or provide partial support for travel.

Deadline for Applications is April 5, 2010 or until all slots are filled. Applications are to be submitted online at http://ccicada.org/Reconnect/2010/ApplicationChecklistPage.html and will be reviewed as they are received.

For more information: Christine Spassione (spassion@dimacs.rutgers.edu) or visit the Reconnect web page http://ccicada.org/Reconnect/2010/.
(This Reconnect Conference takes place in Los Angeles the week immediately after the Human Language Technologies (HLT-NAACL) conference of the North American Association for Computational Linguistics (June 1-6, 2010). HLT-NAACL is one of the primary conferences in this field, and participants are encouraged to consider attending it as well. See http://naaclhlt2010.isi.edu/).

Deliverables (Milestones)

Milestones for VACCINE Period 1 – Professional Training	**07/01/09 – 03/31/10**			
	Quarterly			
Discuss Professional Training opportunities with Rutgers / University of Houston DT /partner schools		X	X	X
John Stasko teacher in Rutgers RECONNECT summer program		X		
Discuss / Develop relationship with Ivy Tech/ Public Safety Academy of Northeast Indiana for continuing ed opportunities				X

Attend IN District 4 first responders meetings for opportunities/ needs		X	X	
Design program for VACCINE period 2				X

Program Outcomes (Results, Accomplishments, Conclusions)

Other outcomes/impacts

CCICADA/VACCINE Summer Reconnect Workshops exposed faculty teaching undergraduates to the role of the mathematical sciences in homeland security and provided an opportunity to researchers in government or industry to learn about recent techniques in data analytics.

John Stasko participated in teaching RECONNECT 2009 that was held at Rutgers University. RECONNECT 2010 was held at USC. Midge Cozzens and John Stasko taught at RECONNECT during the summer 2010. Planning is currently underway for a workshop with public safety professionals to be hosted by VACCINE and the Justice Institute of British Columbia in September 2010.

Planning is underway for a September workshop on visual analytics for public safety professionals. The workshop will be held at the Justice Institute of British Columbia. Participants will include first responders from Canada and the US as well as VACCINE partners. The purpose of the workshop is to define the requirements that will lead to developing academic programs and relevant technology utilizing visual analytics in the public safety environment.

VACCINE has had several meetings with FLETC to discuss training needs and ways that FLETC and VACCINE can work together. Some initial discussions include: FLETC faculty partnering with VACCINE faculty for research, developing a basic course on visual analytics and how it can be an essential tool for law enforcement, and continuing development of mobile applications.

EP 8: Educational Materials

Team

VACCINE:
Georgia Institute of Technology: John Stasko

Abstract

Georgia Institute of Technology presently maintains the Visual Analytics Digital Library (VADL) for storing and presenting educational materials in visual analytics. The VADL includes a wide variety of materials including lectures, lecture notes, lecture slides, related papers, exams, exercises, course outlines, etc. In this project, the team will continue to maintain and enhance the VADL and the materials contained within. http://vadl.cc.gatech.edu

Milestones

Milestones for VACCINE Period 1 – Educational Materials	**07/01/09 – 03/31/10**			
	Quarterly			
Maintain and enhance the VADL		X	X	X

Program Outcomes (Results, Accomplishments, Conclusions)

Other outcomes/impacts

The materials for the VADL are still in place, but the team did not actively work on and enhance the digital library. No funds were included in the first year budget to support a student assistant to work on the VADL.

VIII. Partnerships and Outreach

A. List of Partnerships and Major Outcomes

VACCINE has been very active in both its domestic and international outreach efforts. We have built upon our existing relationships and established many new collaborative partners.

(a) VACCINE Public Safety Partnerships

1. Indiana Public Safety Consortium a. Tippecanoe County Sheriff's Office b. Tippecanoe County Emergency Management Agency c. Lafayette Police Department d. West Lafayette Police Department e. Purdue University Police Department	**2. Florida Division of Emergency Management's Region Seven** a. Florida International University b. Miami-Dade County c. Broward County d. Palm Beach County e. Monroe County
3. UNCC Consortium a. Charlotte Mecklenburg Police Department	**4. Washington State Consortium** a. Port of Puget Sound b. U.S. Coast Guard, Seattle Sector
5. Penn State Law Enforcement Consortium a. Harrisburg Police Department	

In 2010, VACCINE created the Indiana Public Safety Consortium which is comprised of local Indiana law enforcement and emergency management agencies. This Consortium was established with the goal of connecting researchers with first responders. Memorandums of Understanding have been established with all agencies which will permit the sharing of data, information and materials needed to conduct projects with the goals to coordinate and to improve public safety's analysis of data to promote improved operations in the law enforcement, fire, and emergency management and EMS fields. The group meets monthly to discuss tools that are available, as well as to brainstorm tools that are needed. VACCINE envisions the Consortium as a resource that can be utilized by not only VACCINE but also DHS S&T.

(b) VACCINE's State and Federal Partnerships

VACCINE works closely with state and federal agencies to receive data relevant to our projects. VACCINE has a well-established process for entering into Memorandums of Understanding to ensure protection of the parties and data involved.

- Army Research Lab CERDEC
- U.S. Army Corps of Engineers
- Defence Research & Development Canada
- National Institute of Justice
- Army Research Office
- National Institutes of Health

- Department of Health and Human Services
- U.S. Coast Guard
- Oak Ridge National Laboratories
- U.S. Department of State, Office of the Geographer
- Pacific Northwest National Laboratory
- Foreign Broadcast Information Service
- Department of Defense
- National Science Foundation
- Argonne National Laboratory
- National Maritime Intelligence Center
- Indiana Department of Homeland Security (IDHS)
- Indiana State Department of Health (ISDH)
- Indiana Board of Animal Health (IBOAH)
- Georgia Department of Health (GaDH)
- Renaissance Computing Initiative (RENCI)

(c) VACCINE Corporate Involvement

Purdue is in the process of establishing of a unique VACCINE Corporate Affiliate Program (VCAP) to promote VACCINE's goals. In broad outline, Purdue intends for this program to provide a framework to enable corporations to make tax deductible gifts to support research, education, and training activities for all VACCINE member institutions. The basic structure of VCAP would be through the establishment of a non-profit Limited Liability Corporation to be managed by the Purdue Research Foundation and overseen by VACCINE's Leadership Board. The VACCINE Leadership Board would advise the LLC on how to distribute donated funds to support VACCINE's goals. The corporate donations would be true gifts, with no strings attached. However, there would be some practical benefits to donors, including:

- Membership in the VCAP would be by invitation only. Member companies would be complementary as opposed to competitive in nature.
- Corporate affiliates would be in a position to enter into separate research agreements directly with member institutions to develop proprietary technologies, contractually obligated deliverables, etc.
- Corporate affiliates would be in a position to gain early information about ongoing basic research projects at member institutions before publication.
- Corporate affiliates would have opportunities to identify and recruit students at member institutions.
- The Corporate Affiliates program would provide a forum for donors to explore research collaborations and sponsorships.
- Corporate donors would have opportunities for employed scientists to take sabbaticals to participate in relevant research projects at member institutions.

The VCAP Program is anticipated to be in place by the fall semester of 2010. Until then, VACCINE's team has been engaged with a number of companies including:

- Harris Corporation
- The Boeing Company
- VIN
- Raytheon
- Oculus Info, Inc.
- ArgonST
- General Dynamics
- Hallmark
- Charles F. Day & Associates
- NextWave Systems, LLC
- BanField, the Pet Hospital
- MacDonald Dettwiler and Associates
- Kx Systems
- Motorola
- Kimberly Clark
- NVIDIA

(d) In addition to the 19 schools comprising VACCINE, other academic partners include:

- Regenstrief Institute
- Mississippi Valley State University
- Center for Infectious Disease Dynamics, Penn State
- San Diego Supercomputer Center
- West Houston Center for Science and Engineering
- Central Washington University
- SUNY Albany
- START
- DIEM
- Purdue Veterinary Hospital
- University of Colorado
- Penn State Institute for Cyberscience
- Houston Community College
- California State University Dominguez Hills
- University of Maryland
- FAZD
- NCFPD
- CCICADA

VACCINE's first seed project request for proposals generated twenty-five exciting project submissions from all over the world and gave us the opportunity to confer with researchers outside of our existing partners. We were able to provide funding to Swansea University, Florida International University and Navajo Technical College.

VACCINE partner universities have generated over twenty-one million dollars in follow-on funding from additional government sources.

(e) VACCINE International Partnerships:

VACCINE's international outreach efforts have resulted in a strong relationship with the Justice Institute of British Columbia (JIBC). Utilizing additional funding by DHS and DRDC, VACCINE and JIBC have planned an invitational workshop entitled, "Visual Analytics for Public Safety Professionals". The workshop will take place in Canada on

September 20-21, 2010. This corroboration will join public safety professionals from the U.S. and Canada to provide opportunities to network and discover the value that visual analytics can bring to their decision-making and daily lives.

- Simon Fraser University, Canada
- University of British Columbia
- Justice Institute of British Columbia
- Swansea University
- University of British Columbia
- University of Stuttgart, Germany
- University of Konstanz
- University of Groningen, Netherlands
- Linkoping University
- Linnaeus University
- University of Gävle
- Chalmers University, Sweden
- National Institute for Research in Computer Science and Control

B. Workshops and Events

VACCINE Kick-off and Technology Showcase

The teams of VACCINE gathered at Purdue University on November 9, 2009 to celebrate their official kick-off. Indiana Representatives Stephen Buyer and Mark Souder lent their voices in support of VACCINE and its mission. DHS officials Starnes Walker, Matthew Clark and Joe Kielman spoke at the event emphasizing VACCINE's research to reality goals. Law enforcement officials from the state, county and local levels were also in attendance. VACCINE is working closely with the first responder community to evaluate their needs. The celebration concluded with a technology showcase that engaged VACCINE's stakeholders with hands-on demonstrations of varied projects.

C. Presentations by VACCINE

It has been a busy year for members of VACCINE. As part of our outreach efforts, researchers are encouraged to present their research at professional conferences and events. VACCINE's list of presentations is detailed below. Copies of the presentations are available upon request.

Georgia Institute of Technology

1. Stasko, J., "VACCINE Education Program Overview," NVAC Consortium Meeting, Richland, WA, August 2009.
2. Stasko, J., "Visual Analytics for Investigative Analysis and Exploration of Document Collections," M.I.T., Cambridge, MA, September 2009.
3. Stasko, J. Görg, C., and Kang, Y., "Evaluating Visual Analytics Systems for Investigative Analysis: Deriving Design Principles from a Case Study," IEEE VAST 2009 Symposium, Atlantic City, NJ, October 2009.

4. Stasko, J., "Information Exploration and Analysis through Interactive Visualization," Emory University, Atlanta, GA, November 2009.
5. Stasko, J. and Görg, C., "Developing Visual Analytics Applications: Lessons Learned from the Trenches," NSF/DHS FODAVA Annual Review Meeting, Atlanta, GA, December 2009.
6. Stasko, J., "VACCINE Education Program Overview," DHS UP Summit, Washington, D.C., March 10, 2010.

Indiana University

1. Biggers, M., "VACCINE," ADMI/A4RC Conference, Jackson, Mississippi, April 8-11, 2010. Jackson State University.

Pennsylvania State University

1. Fekete, J.D., "Stack Zooming for Multi-Focus Interaction in Time-Series Data Visualization," IEEE Pacific Visualization Symposium, March 2010.
2. Hardisty, F., "The GeoViz Toolkit: An Easy-to-Use Approach to ESDA," URISA Public Health, Providence, RI, 2009.
3. Hardisty, F., "GeoViz Toolkit Tutorial," Workshop at URISA Public Health, Providence, RI, 2009.
4. Hardisty, F., "Research to Reality: Supporting Public Health Research, Surveillance, and Practice with Geovisual Analytics," The Third Annual DHS University Network Summit, Washington, D.C., 2009.
5. Hardisty, F., "Syndromic Surveillance of Influenza Rates using the GeoViz Toolkit," Invited Presentation to the Analysis, Visualization and Reporting (AVR) Webinar for the CDC AVR Series.
6. Hardisty, F. and Klippel, A. "Analyzing Spatio-Temporal Autocorrelation with LISTA-Viz. GeoVA(t) - Geospatial Visual Analytics: Focus on Time,"ICA Commission on GeoVisualization, Guimarães, Portugal, 2010.
7. MacEachren, A., "Geographic Visualization & Analysis for Public Health," Keynote presentation at URISA GIS in Public Health Conference, Providence, RI, June 5-8, 2009.
8. MacEachren, A., "Geovisualization →Geovisual analytics for public health," Invited Presentation to the CDC Analysis-Visualization-Reporting Webinar, July 17, 2009.
9. MacEachren, A., Hardisty, F., and Stryker, M., "GeoVISTA Center Exploratory Geovisualization & Visual Analytics Research & Development," Invited Presentation at NATO IST-085 Workshop on Visualizing Networks, University Park, Pennsylvania, Oct. 7, 2009.
10. MacEachren, A. "SensePlace Demonstration," NATO IST-085 Workshop on Visualizing Networks, University Park, Pennsylvania, October 7, 2009.
11. MacEachren, A. "Grounding Geovisualization & ESDA Tool Design in Cartographic Theory, Practice, and Empirical Research," Joint Statistical Meeting, Washington, D.C., August 4-9, 2009.
12. MacEachren, A. "SensePlace Demonstration," 2010 Technologies for Critical Incident Preparedness Conference and Exposition, Philadelphia, Pennsylvania, 2010.

13. Mitra, P., “Information Extraction for Geospatial Visual Analytics,” University of Houston Downtown, Computer Science Seminar, Houston, 2010.
14. Roth, R., “CrimeViz Demonstration,” 2010 Technologies for Critical Incident Preparedness Conference and Exposition, Philadelphia, PA, ,2010.
15. Yu, B. and Cai, G., “Facilitating Participatory Decision-Making in Local Communities Through Map-Based Online Discussion,” The Fourth International Conference on Communities and Technologies (C&T 2009). J. M. Carroll. University Park, Pennsylvania, ACM, June 25-27, 2009.

Purdue University

1. Bathur, “CERIAS Information Security Symposium,” West Lafayette, Indiana, March 24, 2010.
2. Burns, M. and Collins, T., “VACCINE and Education,” Indianapolis, Indiana, October 14, 2009.
3. Burns, M., Combs, K., and McKay, S., “General VACCINE and HS STEM Research,” Triton Central High School, Bourbon, Indiana, October 19, 2009.
4. Burns, M. and Sims, B., “General VACCINE and HS STEM Research,” West Lafayette Junior High School, West Lafayette, Indiana, November 24, 2009.
5. Burns, M., “VACCINE Overview,” Capital Forum Indiana, Indianapolis, Indiana, January 31, 2010.
6. Burns, M., “VACCINE and Education,” Discovery Park Engagement Group, West Lafayette, Indiana, February 2, 2010.
7. Burns, M. and Collins, T., “VACCINE Overview,” Hoosier Science Association Conference, Indianapolis, Indiana, February 3-4, 2010.
8. Burns, M., “VACCINE and Education,” Regenstrief Center for Healthcare Engineering, West Lafayette, Indiana, March 1, 2010.
9. Burns, M. and Ebert, D., “VACCINE Overview,” March 2010.
10. Burns, M. and Ebert, D., “VACCINE Education CCICADA Kickoff,” December 2009.
11. Cleveland, W., “RBSA for Keystroke Detection,” NSA Invited Lecture, Fort Meade, Maryland, October 8, 2009.
12. Collins, T., “VACCINE Center,” Army Research and Development Center, Picatinny Arsenal, October 2009.
13. Collins, T., “VACCINE Center,” Indiana Association of Emergency Management and Public Safety, October 2009.
14. Collins, T., “VACCINE Center,” American Society for Industrial Security International, November 2009.
15. Collins, T., “VACCINE Center,” TCIP Conference, February 2010.
16. Delp, E., "Rosetta Phone."
17. Delp, E., “Visual Analytics for Command, Control, and INteroperability Environments,” Motorola, December 2009.
18. Ebert, D., "Visual Analytics for Crisis Response and Emergency Management," Joint US/Germany Visual Analytics Conference, Konstanz, Germany, June 2009.
19. Ebert, D., “Applications of Visual Analytics to FAZD Center Activities,” FAZD, June 2009.

20. Ebert, D., “From Foundations to Applications of Visual Analytics,” IASTED Conference, July 2009.
21. Ebert, D., “Visual Analytics for Command, Control, and INteroperability Environments,” NMIC, July 2009.
22. Ebert, D., “Visual Analytics for Command, Control, and INteroperability Environments,” ArgonST, July 2009.
23. Ebert, D., “CCI Center of Excellence – Visualization Sciences Overview,” NVAC Consortium, August 2009.
24. Ebert, D., "Visual Analytics for Safety and Security," UK VAC, September 2009.
25. Ebert, D., "An International Perspective on Visual Computing," Opening of the Welsh Institute of Visual Computing (RIVIS), Cardiff, Wales, UK, September 2009.
26. Ebert, D., Maciejewski, R., "Data Modeling and Exploration Using Visual Analytics," National Defense University, October, 2009.
27. Ebert, D., "Visual Analytics: Powering Discovery, Decisions, and Actions from Floods of Data," Distinguished Lecture, Department of Computer Science and Engineering, Ohio State University, October, 2009.
28. Ebert, D., “Visual Analytics for Command, Control, and INteroperability Environments,” MD Fusion Center, November 2009.
29. Ebert, D., “Visual Analytics for Command, Control, and INteroperability Environments,” VACCINE Kick-off, November 2009.
30. Ebert, D., “Visual Analytics for Command, Control, and INteroperability Environments: Research Overview,” CCICADA Kick-off, December 9, 2009.
31. Ebert, D., “Visual Analytics for Command, Control, and INteroperability Environments,” CCICADA Kick-off, December 10, 2009.
32. Ebert, D., “VACCINE Research Transition,” Navy War College, December 2009.
33. Ebert, D., "Visual Analytics: Powering Discovery, Innovation, and Action in the Digital Age," Jackson State University, January, 2010.
34. Ebert, D. and Maciejewski, R., "Public Health Data Exploration, Analysis, Modeling, and Prediction Using Visual Analytics," U. S. Department of Health and Human Services, January 2010.
35. Ebert, D., "Visual Analytics for Command, Control, and INteroperability Environments," Midwest Anti-Terrorism Advisory Council, U.S. Attorney's Office, Chicago, IL, March 2010.
36. Ebert, D., "Visual Analytics for Command, Control, and Interoperability Environments," DHS University Programs Summit, Washington, DC, March 2010.
37. Ebert, D., “Visual Analytics for Command, Control, and INteroperability Environments,” Briefing to Admiral Thad Allen at Purdue University, March 2010.
38. Ebert, D., “Visual Analytics for Command, Control, and INteroperability Environments,” Argonne National Laboratory, March 2010.
39. Elmqvist, N., “GraphDice,” EuroVis, Bordeaux, France, June 2010.
40. Elmqvist, N., "CoE-Explorer: Visualizing DHS Centers of Excellence,” C2I Panel at the U.S. DHS University Network Summit 2010, Washington, D.C., March 2010.
41. Maciejewski, R., “Structuring Feature Space – A Non-Parametric Method for Volumetric Transfer Function Generation,” IEEE Visualization Conference, October 2009.

42. Maciejewski, R., “Data Aggregation and Analysis for Cancer Statistics – A Visual Analytics Approach,” The Hawaii International Conference on System Sciences, January 2010.
43. Maciejewski, R., Gates, C., Chaturvedi, A., “The Visualization of Security,” The Center for Education and Research in Information Assurance and Security (CERIAS) Annual Information Security Symposium, March 2010.

Simon Fraser University

1. Fisher, B., "VACCINE Meets Vaccine: Public Health Research at SFU/UBC VIVA."
2. Fisher, B., “Canada/US Collaboration on Visual Analytics for Command, Control and Interoperability,” Defence Security Innovation, Québec QC, 2009.

Stanford University

1. Gerth, J., "Stanford/Purdue: Cybersecurity Analytics."
2. Gerth, J., "Visualization for Network Cybersecurity," Stanford MediaX Visualization Course, Palo Alto, California, August 11, 2009.
3. Gerth, J., “Stanford Network Cybersecurity,” Palantir, Inc., Palo Alto, California, August 25, 2009.
4. Gerth, J., “Stanford/PNNL Tools,” US-CERT/DHS S+T PI Meeting, Rosslyn, Virginia, March 8, 2010.
5. Gerth, J., “Isis,” US-CERT Analyst Training, Arlington, Virginia, March 8, 2010.
6. Gerth, J., “Visualization for Network Cybersecurity,” DHS S+T PI Meeting, Rosslyn, Virginia, March 9, 2010.
7. Gerth, J., “Cybersecurity Analytics,” DHS UP Summit, Washington, D.C., March 10, 2010.

University of Houston Downtown

1. Alo’, R. A., “VACCINE,” California State University Long Beach, College of Engineering, August, 2009.
2. Alo’, R. A., “VACCINE,” California State University Dominguez Hills, College of Sciences, August, 2009.
3. Alo’, R. A., “VACCINE,” SACNAS Annual Meeting, Dallas, October, 2009.

University of Texas

1. Gaither, K., "Hurricane Routing/Disaster Relief," DHS UP Summit, Washington, D.C., March 10, 2010.

University of Washington

1. Haselkorn, M., Benson, A., Biggers, K., and Wall, J., "Adaptive Development of a Common Operating Environment for Crisis Response and Management," ISCRAM 2010, Seattle, Washington, May 4, 2010.

University of North Carolina Charlotte

1. Ribarsky, W., "Exploratory Multimedia Analysis Applied to Video Analytics," Video Analytics Workshop, IEEE VisWeek, Atlantic City, New Jersey, 2009.
2. Kosara, Robert, "Embedding Information Visualization in Visual Communication, Arizona State University," February 13, 2009.
3. Ribarsky, W., "A Mobile System for Effective Emergency Response and Situation Awareness in Large Indoor Environments," Precision Indoor Personnel Location and Tracking for Emergency Responders, Worcester Polytechnic Institute, Worcester, Massachusetts, August 2009.
4. Ribarsky, W. Distinguished Lecturer, "Developing a Visual Analytics Approach to Analytic Problem-Solving," School of Computational Science and Engineering, Georgia Tech, February, 2010.
5. Ribarsky, W., "Situation Aware Mobile Tools for First Response and Emergency Evacuation," Invited Presentation, DHS University Programs Summit, Washington, D.C., March 10, 2010.
6. Ribarsky, W. Invited speaker, "Mobile Emergency Evacuation for Urban Area Security," Workshop for Charlotte Fire Department and Regional Homeland Security Officials, Charlotte, April, 2010.
7. Fan, J., ``Human-centered multimedia computing for personalized image/video recommendation," WOCC (The 19th Annual Wireless and Optical Communications Conference) Shanghai, May 14-17, 2010.
8. Subramanian, KR, "Emergency Response Within Indoor Structures: Interactive Visualization Tools," Technical Working Group Meeting, National Institute of Justice, April, 2010, Portland, OR, USA.

Virginia Tech

1. Marathe, M., "Co-Evolving Complex Networks: Epidemics in social and wireless networks," Argonne National Laboratory, Argonne, Illinois, September 28, 2009.
2. Eubank, S., "The role of individual-based models in controlling infectious disease outbreaks," Invited presentation at NIMBioS Investigative Workshop on Optimal Control and Optimization for Individual and Agent based Models, Knoxville, Tennessee, December 1-3, 2009.
3. Eubank S., "Beyond degree distribution: local to global structure of social contact networks," Invited presentation to 2010 International Conference on Social Computing, Behavioral Modeling and Prediction, Washington, DC, March 30, 2010.
4. Barrett, C., "Policy & Decision Informatics of Complex Systems," Invited Presentation at University of Wollongong, Wollongong, Australia, September 6-19, 2009.
5. Barrett C., Marathe M., "Interaction-based modeling of population dynamics and infectious disease," Infectious Disease Modeling Meeting, John Hopkins University, Baltimore, Maryland, January 7, 2010.

6. Eubank S., "Indemics: A synthetic information environment for epidemiological decision support," Infectious Disease Modeling Meeting, John Hopkins University, Baltimore, Maryland, January 7, 2010.

D. Center Publicity/Marketing Efforts

VACCINE takes pride in our ability to educate the public about the innovative research we perform through a variety of media outlets and marketing tools. Our website (www.VisualAnalytics-CCI.org) is an informational tool that provides an in-depth overview of our Center. Our website contains details of our research and educational programs, in addition to our featured projects, publications and current events.

Many of VACCINE's activities and research have been featured in the following media outlets:

- Journal & Courier
- SuperComputingOnline
- EurekAlert!
- Science Blog
- MedGadget
- Swine Flu Pandemic
- Trak.In News
- Crytogon.Net
- Tri-City Herald
- Thaindian News
- KEPR TV
- Taragana
- Fraud Magazine
- UPI.com
- Adobe.com
- Purdue University News Service
- WLFI TV
- Security Director News
- Indiana Business
- HPC Wire
- Indy Star
- Laboratory Equipment
- VizWorld.com
- Homeland Security Today
- Purdue University Research Review

Copies of the referenced articles can be found in Appendix B – VACCINE News Releases.

E. Other Outreach Activities

In addition to presenting at many conferences and workshops, VACCINE researchers are frequently asked to present their work at special events. These events include presentations to the first responder community, public and private corporations, and government agencies. Examples include, but are not limited to the following:

- Argonne National Laboratories
- Chicago Office of Emergency Management
- National Maritime Intelligence Center
- Motorola
- Raytheon
- Technologies for Critical Preparedness Conference (TCIP)
- American Society for Industrial Security International
- Indiana Association of Emergency Management and Public Safety
- Army Research and Development Center, Picatinny Arsenal

IX. Integration

1. CCICADA Research and Education

VACCINE and CCICADA, as co-leads for the Command, Control and Interoperability Center of Excellence, are working together to identify high-impact areas for bringing combined data and visual analytic capabilities to bear on problems that are critical to the homeland security research and education mission.

VACCINE and CCICADA are collaborating to create a seamless; integrated "center" that is much more than a collection of partnering institutions. The ultimate aim is to become the leader in the field of Data Analytics and Visualization Sciences that addresses the knowledge gap and serves society through creativity, expanded research, and state-of-the-art innovations. Toward the achievement of such goals, the CCI Center of Excellence has been using its very unique and robust resources to undertake the challenge of critical DHS and societal needs.

Both VACCINE and CCICADA leverage existing expertise within their networks of partner schools for program development and dissemination, thereby reducing duplication of effort. VACCINE has aligned its research into the network of DHS Centers of Excellence as it continuously develops and engages partners to provide real-world validation of its work. In parallel with the research, VACCINE and CCICADA have maintained a collaborative education relationship with monthly conference calls, joint attendance and presentations at conferences and shared development of material and content. In fact, fully 45% of our research and education portfolio is in collaboration with either CCICADA or another Center of Excellence.

- MDRP 3: Situational Surveillance and In-Field Criminal Investigative Analytics (CCICADA)
- MDRP 13: GeoJunction (CCICADA)
- MDRP 17: START Center Visual Analytics (CCICADA, START)
- MDRP 18: Visual Analytics for the DHS Centers of Excellence (CCICADA)
- MDRP 7: Introducing Sustainable Visual Analytics into Command Center Environments (FAZD)
- MDRP 12: Foreign Animal and Zoonotic Disease Visual Analytics (FAZD)
- EP 1: Leadership and Coordination (CCICADA)
- EP 2: MSI Collaboration (CCICADA)
- EP 3: K-12 Programs (CCICADA)
- EP 7: Professional Training (CCICADA)

In addition, during this first year, we have been setting the stage to conduct joint research with NCFPD, CREATE, and DIEM and we expect these projects to begin in Year 2.