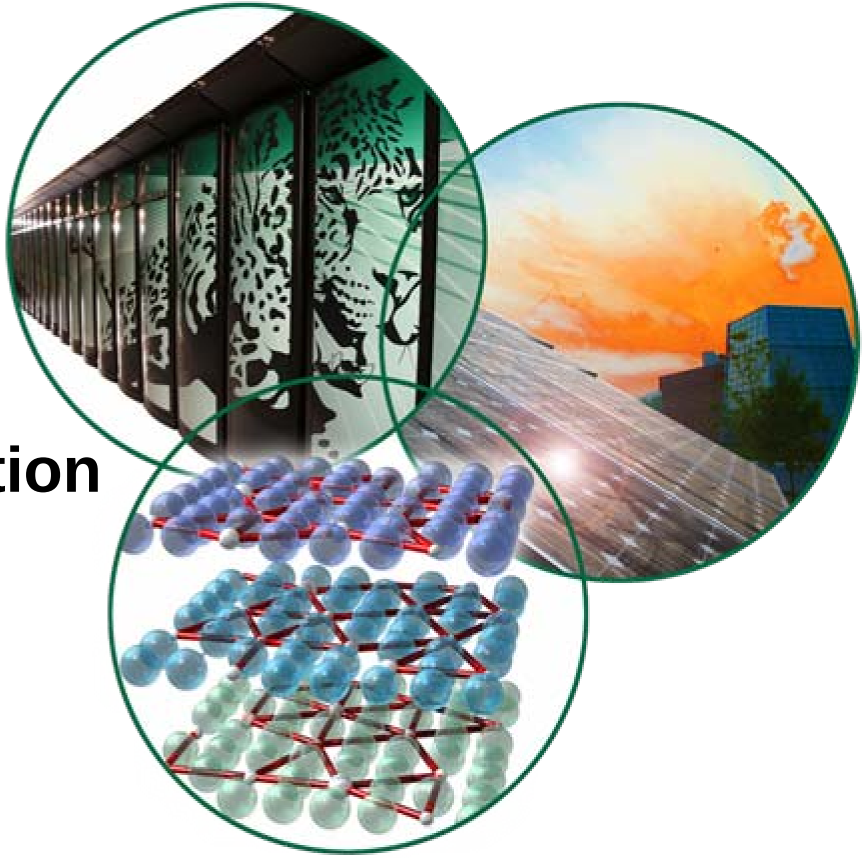
Oak Ridge National Laboratory
—The "not so foggy" future!



MG(R) Dennis K. Jackson
Director, Logistics Transformation
National Security Directorate

jacksondk@ornl.gov
(865) 574-7382

U.S. DEPARTMENT OF ENERGY

OAK RIDGE National Laboratory

Leading the development of ultrascale scientific computing

- **Leadership Computing Facility:**
  - **World's most powerful open scientific computing facility**
  - **Jaguar XT5 operating at 1.64 petaflops**
  - **Exascale system by the end of the next decade**
  - **Focus on computationally intensive projects of large scale and high scientific impact**
- **ORNL team won the Gordon Bell Prize at SC'08**
- **With the University of Tennessee, developing a second petascale computer for the National Science Foundation**

ORNL Is Committed to the Knowledge Discovery Agenda

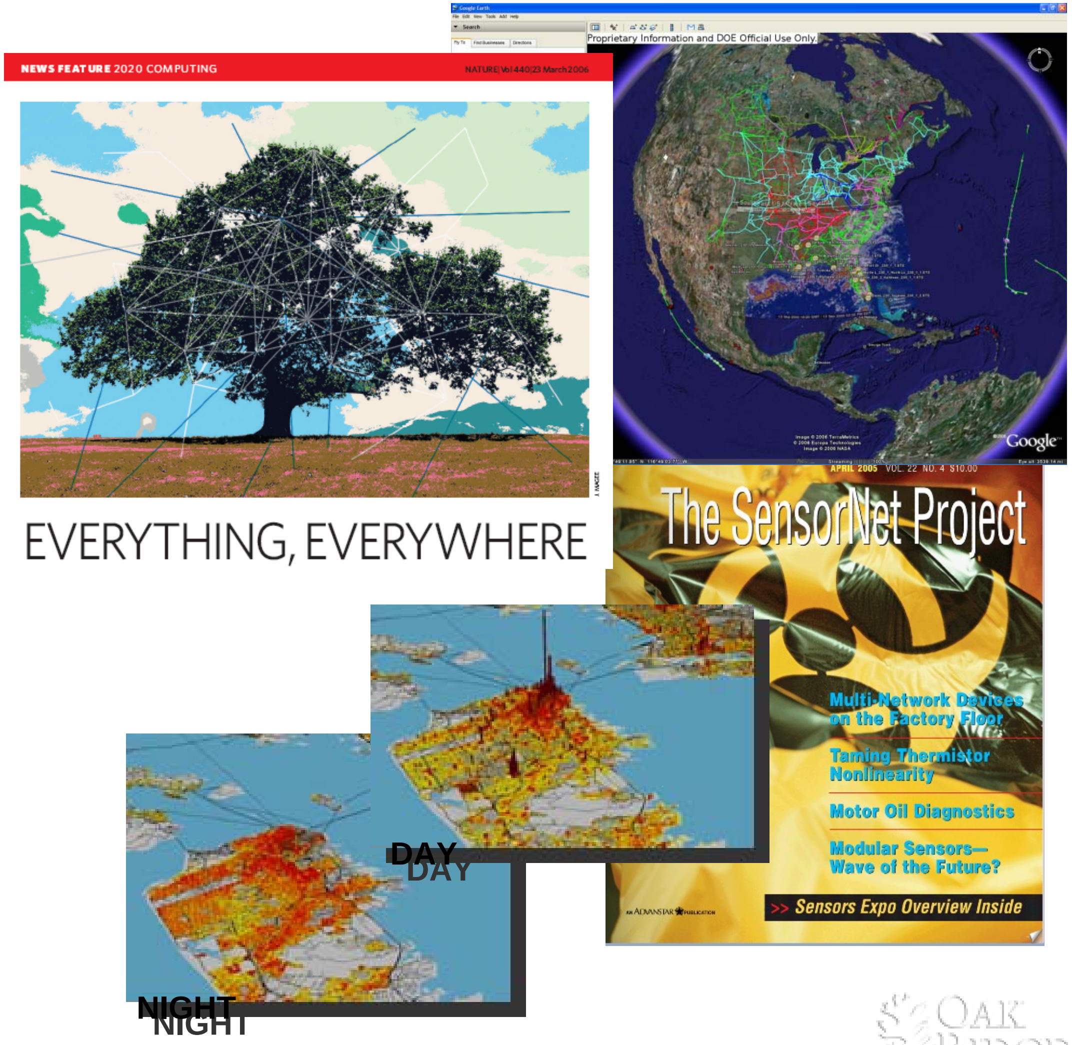
- **Entire Research Division Focused on Knowledge Discovery**
  - **130 full-time staff**
  - **50 subcontractors**
  - **50 students**

- **Outstanding Resources:**

  HPC, Networking, MRF, JICS

- **LDRD Initiative in Knowledge Discovery**

- **Programmatic efforts well-aligned with this science agenda**

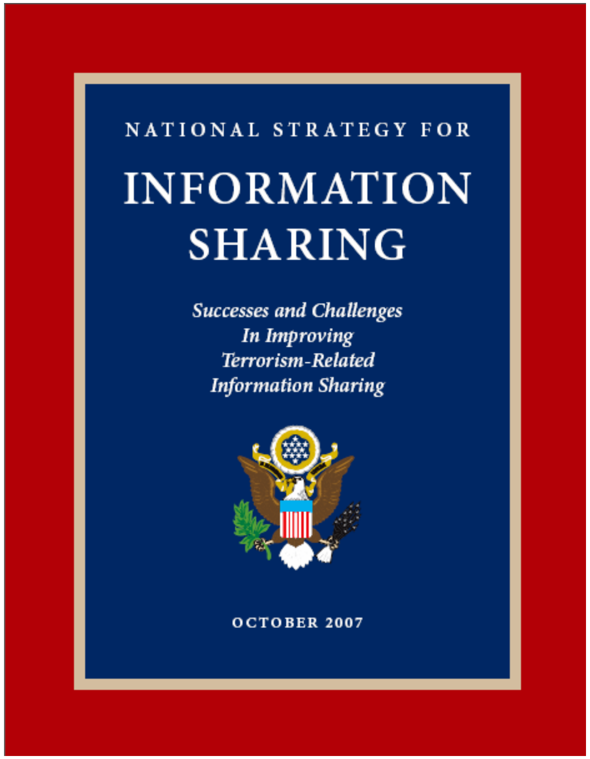

NEWS FEATURE 2020 COMPUTING
NATURE|Vol 440|23 March 2006
EVERYTHING, EVERYWHERE
Proprietary Information and DOE Official Use Only
Google
APRIL 2005 VOL. 22 NO. 4 $10.00
The SensorNet Project
Multi-Network Devices on the Factory Floor
Taming Thermistor Nonlinearity
Motor Oil Diagnostics
Modular Sensors—Wave of the Future?
>> Sensors Expo Overview Inside
AN ADVANSTAR PUBLICATION
DAY
NIGHT

ORNL's Focus in Knowledge Discovery…

- **Actionable insights from massive, dynamic, disparate data sources**
  - **Knowledge representation of disparate data sources**
  - **High speed analysis and fusion of text, video, audio, and sensor data streams**
  - **Geospatial and temporal data science**
- **Ability to ask more complex questions and detect more complex processes using increasingly higher data resolution**
  - **Population models and population data development**
  - **Modeling and simulation of emerging behavior in complex systems (e.g., social systems)**
  - **Real-time data driven simulations (take advantage of data resolution and availability)**

Knowledge Discovery Challenge

How to trigger and coordinate a discovery process across data held by industry, academia, and government agencies within and outside the United States

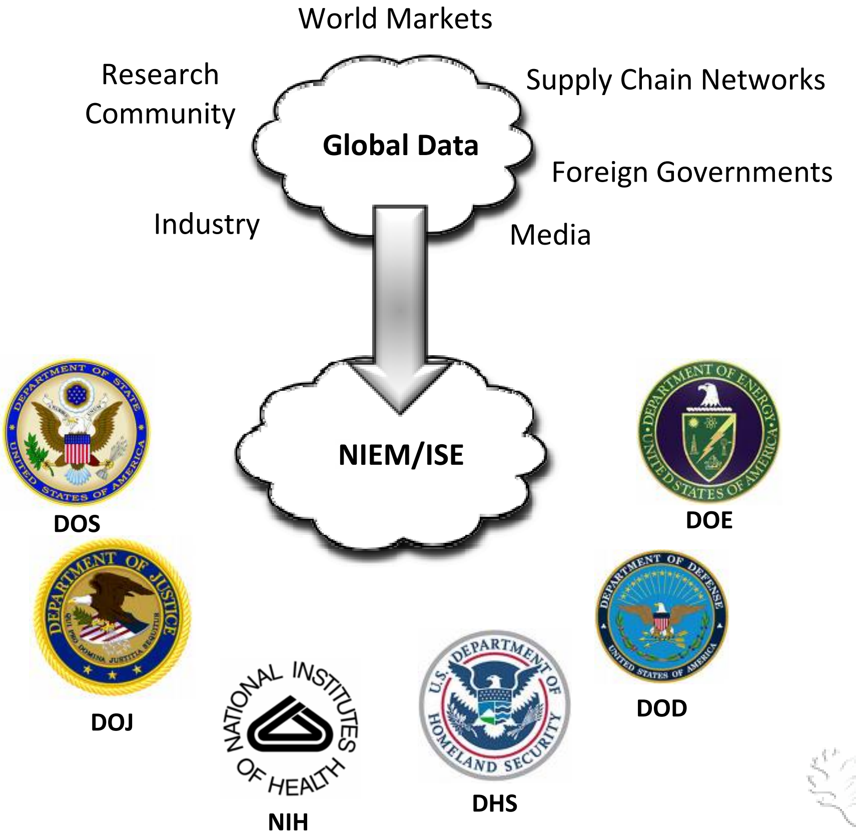


NATIONAL STRATEGY FOR

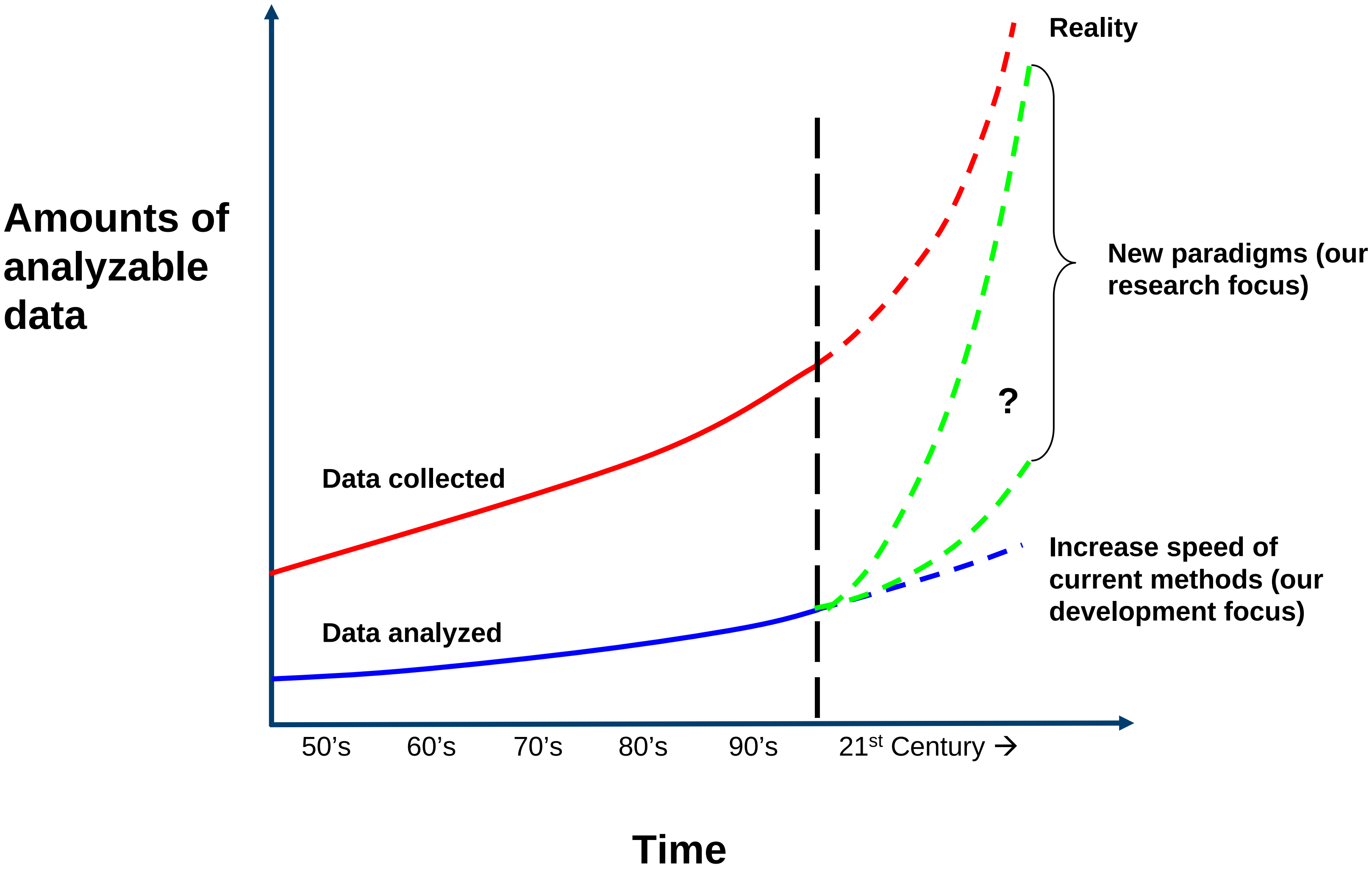
INFORMATION SHARING

Successes and Challenges In Improving Terrorism-Related Information Sharing

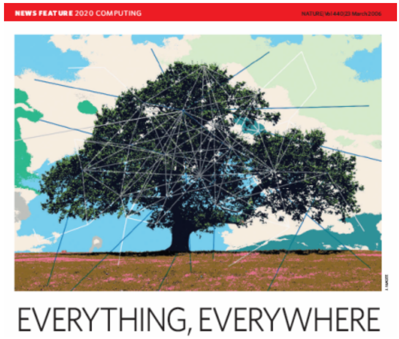
OCTOBER 2007



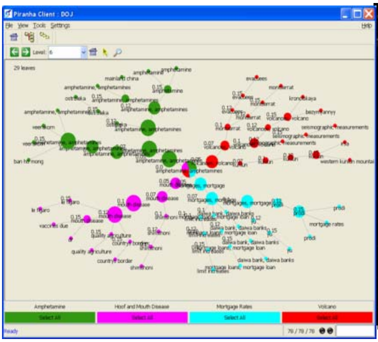
Knowledge Discovery Challenge



Research and Development Focus Areas

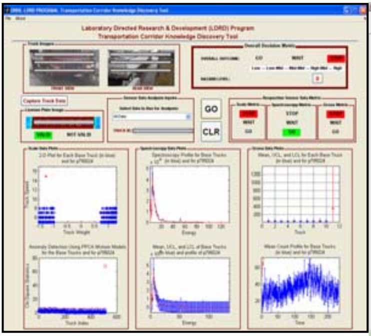


Sensor Networks

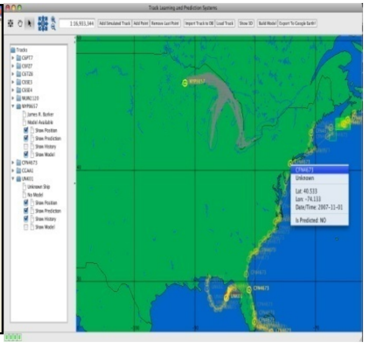


Analysis in Network

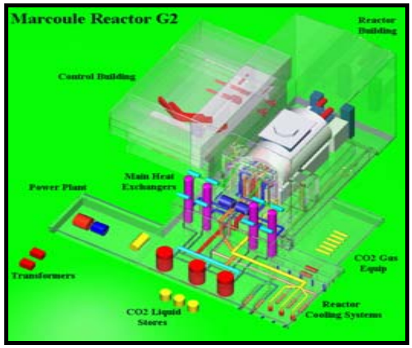
Persistent Surveillance


Laboratory Directed Research & Development (LDRD) Program
Transportation Corridor Knowledge Discovery Tool
GO
CLR

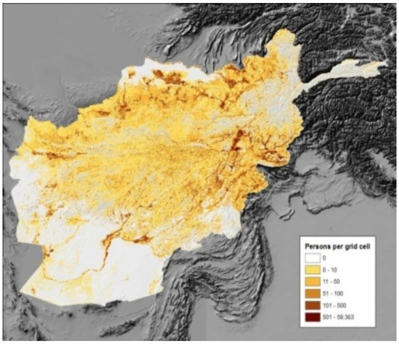
Data Fusion



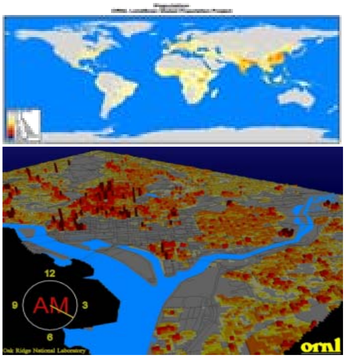
Anomaly Detection



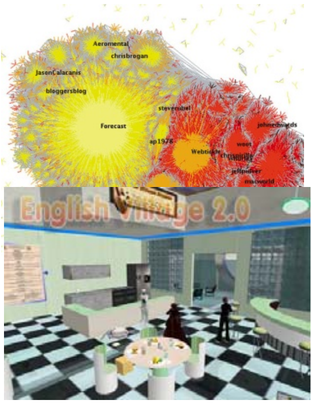
Predictive Analysis



Emergent Behavior



Population Dynamics



Social Data Analysis

Quantum Information

Our largest set of projects relate to collection, analysis, and dissemination of sensor data.

- **Interdiction, detection, emergency response**
  - **Mobile, Transportation Corridors, Ports, Military Bases**
- **Real-Time Data Management**
  - **Collection, Dissemination, Archiving**
- **Pre-deployment analysis**
  - **Cost, Performance Prediction, Risk vs Benefit**

YOUR TECHNICAL RESOURCE FOR SENSING, COMMUNICATIONS, AND CONTROL

sensors

www.sensorsmag.com
APRIL 2005 VOL. 22 NO. 4 $10.00

The SensorNet Project

Multi-Network Devices on the Factory Floor

Taming Thermistor Nonlinearity

Motor Oil Diagnostics

Modular Sensors—Wave of the Future?

>> Sensors Expo Overview Inside

- **Wide-area ubiquitous sensing, actuation, and deployment**
  - **Orchestrating the functionality across a large system of distributed sensors/processors (eg Electric Grid, Autonomous robotic systems)**
- **Cross-agency and cross-administrative boundary data-sharing and interoperability**
  - **Standards and policies**
- **Net-Centric Services**
- **Security, Access Controls**

INTELLIGENT SYSTEMS

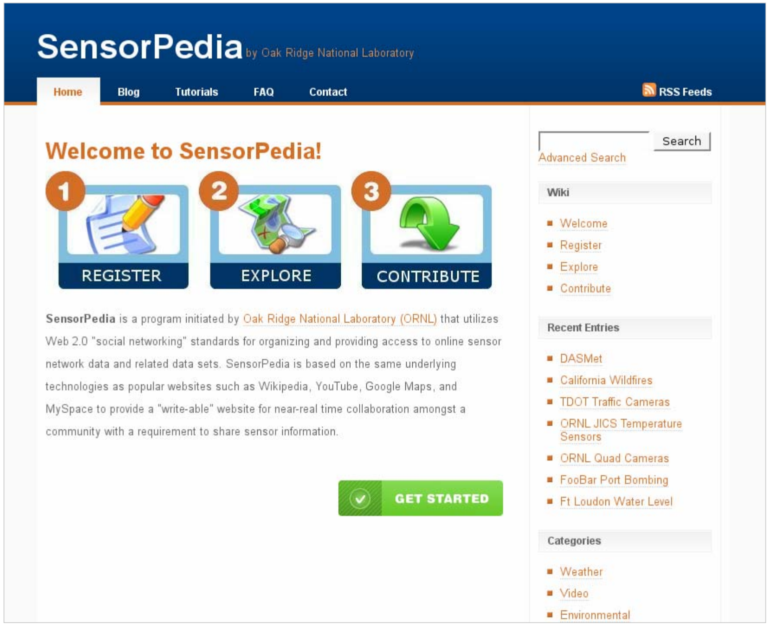
Advancing Sensor Web Interoperability

SensorNet is a framework to tie together sensor data from all over the country to create a real-time detection and alert system for various threats, whether they are chemical, radiological, biological, nuclear, or explosive.

Bryan L. Gorman, Mallikarjun Shankar, and Cyrus M. Smith, SensorNet Program, Oak Ridge National Laboratory

SensorNet is a vendor-neutral interoperability framework for Web-based discovery, access, control, integration, analysis, and visualization of online sensors, sensor-derived data repositories, and sensor-related processing capabilities. In other words, SensorNet attempts to create a [illegible] system to collect and analyze data from sensors all over the country to [illegible] and then alert agencies, emergency responders, and others as necessary. It is being designed and developed by the Computational Sciences and Engineering Division at the Oak Ridge National Laboratory (ORNL), in collaboration with the National Oceanic and Atmospheric Administration (NOAA), the Open Geospatial Consortium (OGC), the National Institute for Standards and Technology (NIST), the Department of Defense, and [illegible]

Social Network for Sharing Sensor Data



SensorPedia by Oak Ridge National Laboratory

Home | Blog | Tutorials | FAQ | Contact | RSS Feeds

Welcome to SensorPedia!

SensorPedia is a program initiated by Oak Ridge National Laboratory (ORNL) that utilizes Web 2.0 "social networking" standards for organizing and providing access to online sensor network data and related data sets. SensorPedia is based on the same underlying technologies as popular websites such as Wikipedia, YouTube, Google Maps, and MySpace to provide a "write-able" website for near-real time collaboration amongst a community with a requirement to share sensor information.

GET STARTED

Search

Advanced Search

Wiki

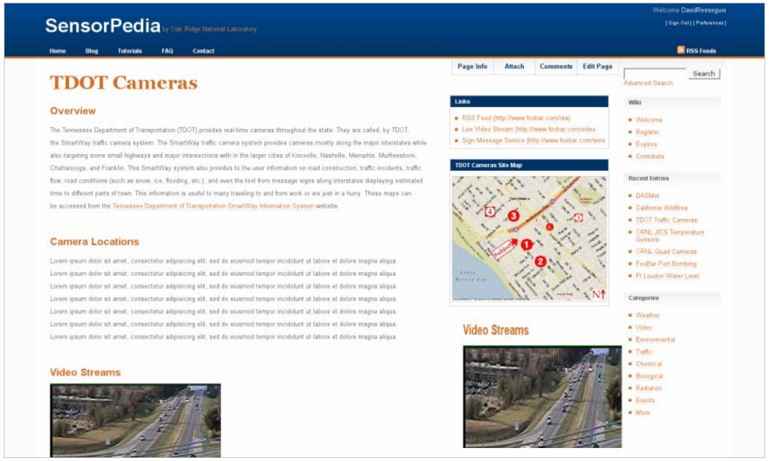
- Welcome
- Register
- Explore
- Contribute

Recent Entries

- DASMet
- California Wildfires
- TDOT Traffic Cameras
- ORNL JICS Temperature Sensors
- ORNL Quad Cameras
- FooBar Port Bombing
- Ft Loudon Water Level

Categories

- Weather
- Video
- Environmental



SensorPedia by Oak Ridge National Laboratory

Welcome DavidResseguie

[Sign Out] [Preferences]

Home | Blog | Tutorials | FAQ | Contact | RSS Feeds

Page Info | Attach | Comments | Edit Page

TDOT Cameras

Overview

The Tennessee Department of Transportation (TDOT) provides real-time cameras throughout the state. They are called, by TDOT, the SmartWay traffic camera system. The SmartWay traffic camera system provides cameras mostly along the major interstates while also targeting some small highways and major intersections with in the larger cities of Knoxville, Nashville, Memphis, Murfreesboro, Chattanooga, and Franklin. This SmartWay system also provides to the user information on road construction, traffic incidents, traffic flow, road conditions (such as snow, ice, flooding, etc.), and even the text from message signs along interstates displaying estimated time to different parts of town. This information is useful to many traveling to and from work or are just in a hurry. These maps can be accessed from the Tennessee Department of Transportation SmartWay Information System website.

Camera Locations

Lorem ipsum dolor sit amet, consectetur adipisicing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua
Lorem ipsum dolor sit amet, consectetur adipisicing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua
Lorem ipsum dolor sit amet, consectetur adipisicing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua
Lorem ipsum dolor sit amet, consectetur adipisicing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua
Lorem ipsum dolor sit amet, consectetur adipisicing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua
Lorem ipsum dolor sit amet, consectetur adipisicing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua
Lorem ipsum dolor sit amet, consectetur adipisicing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua

Video Streams

Links

- RSS Feed (http://www.foobar.com/rss)
- Live Video Stream (http://www.foobar.com/video
- Sign Message Service (http://www.foobar.com/sms

TDOT Cameras Site Map

Video Streams

Search

Advanced Search

Wiki

- Welcome
- Register
- Explore
- Contribute

Recent Entries

- DASMet
- California Wildfires
- TDOT Traffic Cameras
- ORNL JICS Temperature Sensors
- ORNL Quad Cameras
- FooBar Port Bombing
- Ft Loudon Water Level

Categories

- Weather
- Video
- Environmental
- Traffic
- Chemical
- Biological
- Radiation
- Events
- More

SensorPedia

Addresses the ability to access and fuse data from disparate sensor networks

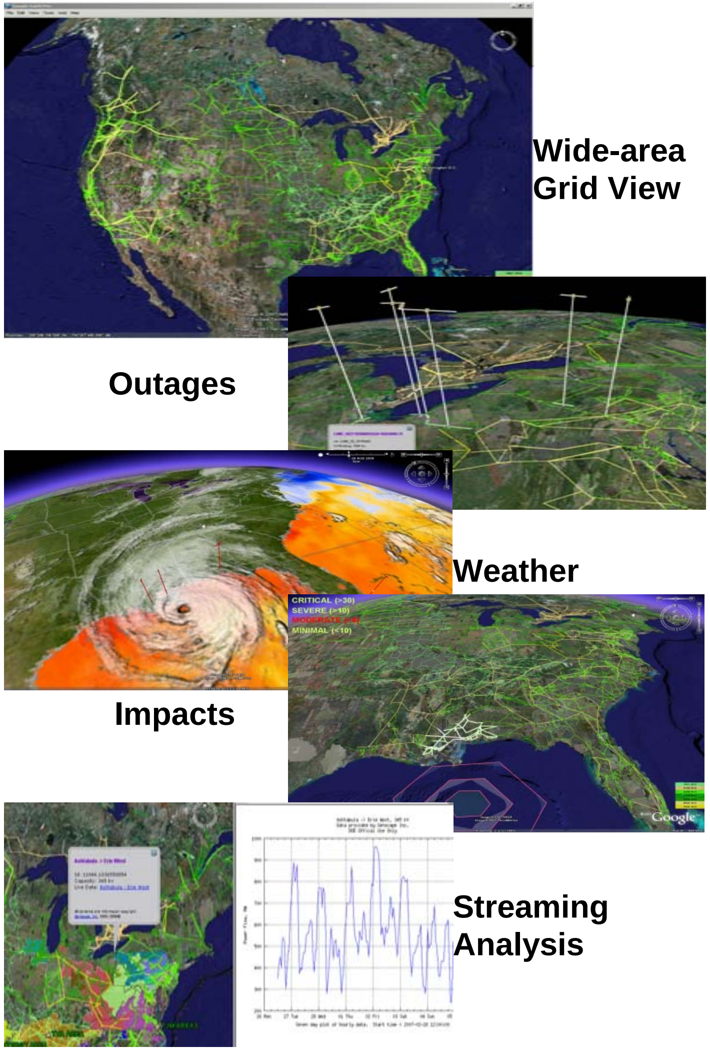
Use of Web 2.0 “social networking” technologies (e.g., RSS, wikis, podcasts, mashups, blogs, and streaming video)

Key identity management and credentialing standards

Data owner controls publishing and subscribing

Explores how volunteered sensor data is being used and shared

Knowledge Representation for Situation Awareness of the Electric Grid



Wide-area Grid View

Outages

Weather

Impacts

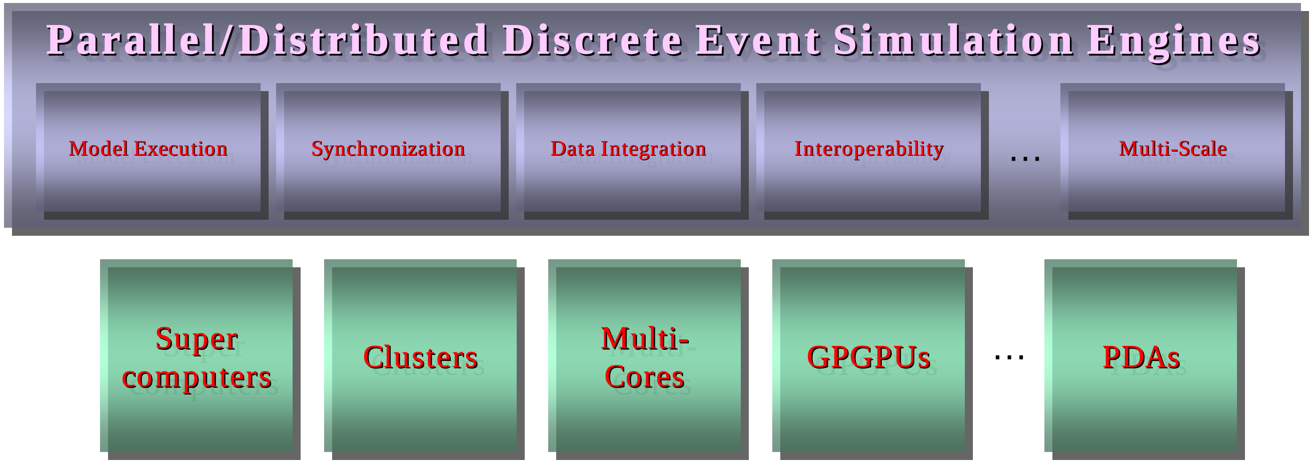
Streaming Analysis

- **Organize, stream, and fuse data from various sources through an analysis pipeline**

- **Present an intuitive visualization of the status to end-users**

Where are all my local, state, and federal assets?

- What assets can I track at all times?
- How well can I estimate the location of non-tracked assets?
- What computational resources will be required?
- What are the uncertainties?



IBM Blue Gene Award for scalable algorithms

Best Paper Award for agent-based methods

Tackling DTRA 10**5 persistent surveillance grand challenge

Event Spectroscopy: Natural Disasters

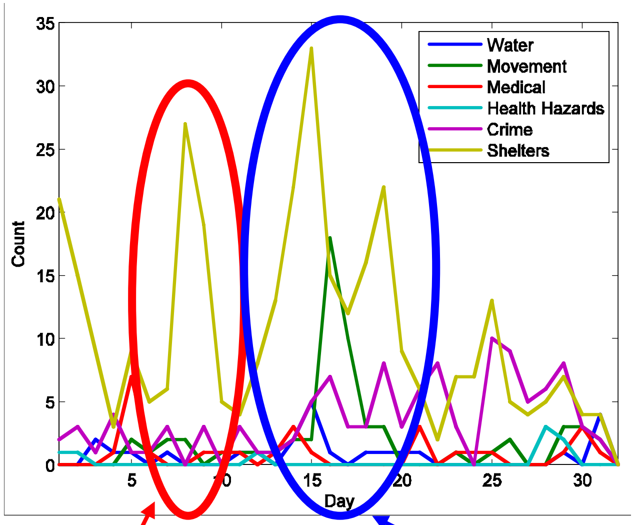
- ABC
- NBC
- CBS
- PBS
- CNN
- FOX News
- MS NBC
- REUTERS
Associated Press
Google News
moreover A VeriSign Company
RSS BLOG

Text
Time-series →

Textual Prism

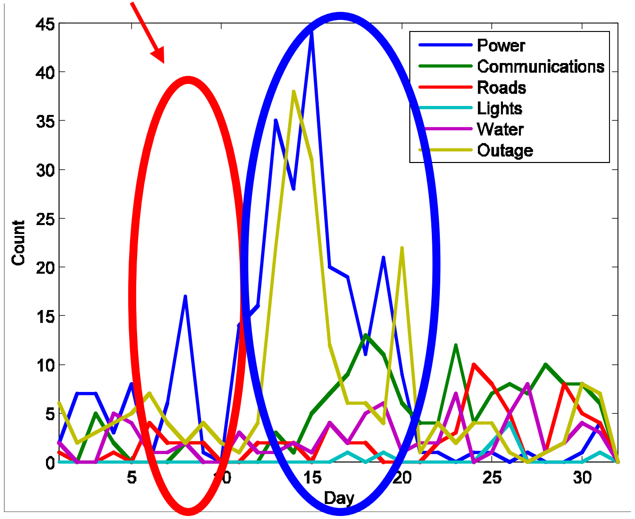
Technical	Social
Power	Water (health)
Communications	Movement
Roads	Medical
Lights	Health Hazards
Water (Systems)	Crime
Outage	Shelters

Social Spectral Components



Tornadoes
Jan 7, 2007

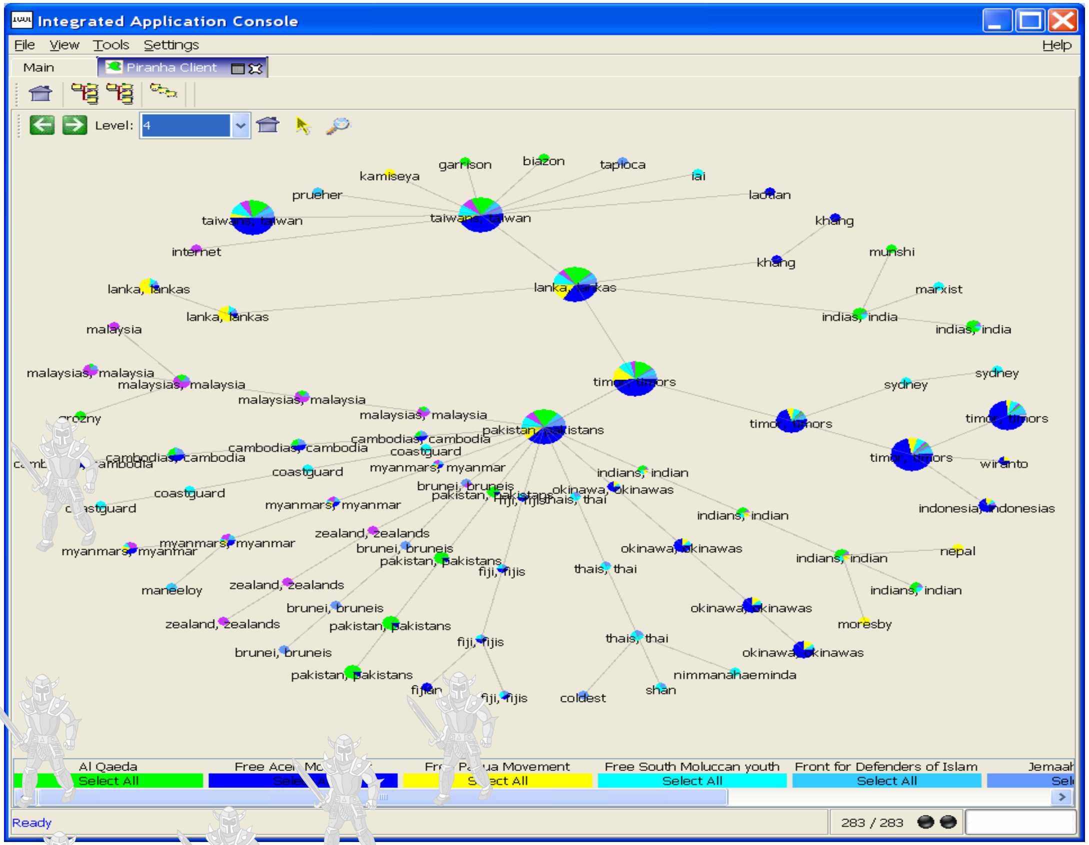
Ice Storm
Jan 14, 2007



Technical Spectral Components

Piranha

Knowledge Discovery Engine



US naturalized citizen indicted for hiding
6 Indonesians Barred From U.S.
Current and Former Military
Cold war leaves a deadly anthrax legacy
...bekistan -- In the ... of ... east of Moscow were ordered to undertake their most critical mission.

2007 R&D 100 Award Winner

Managed by UT-Battelle for the Department of Energy

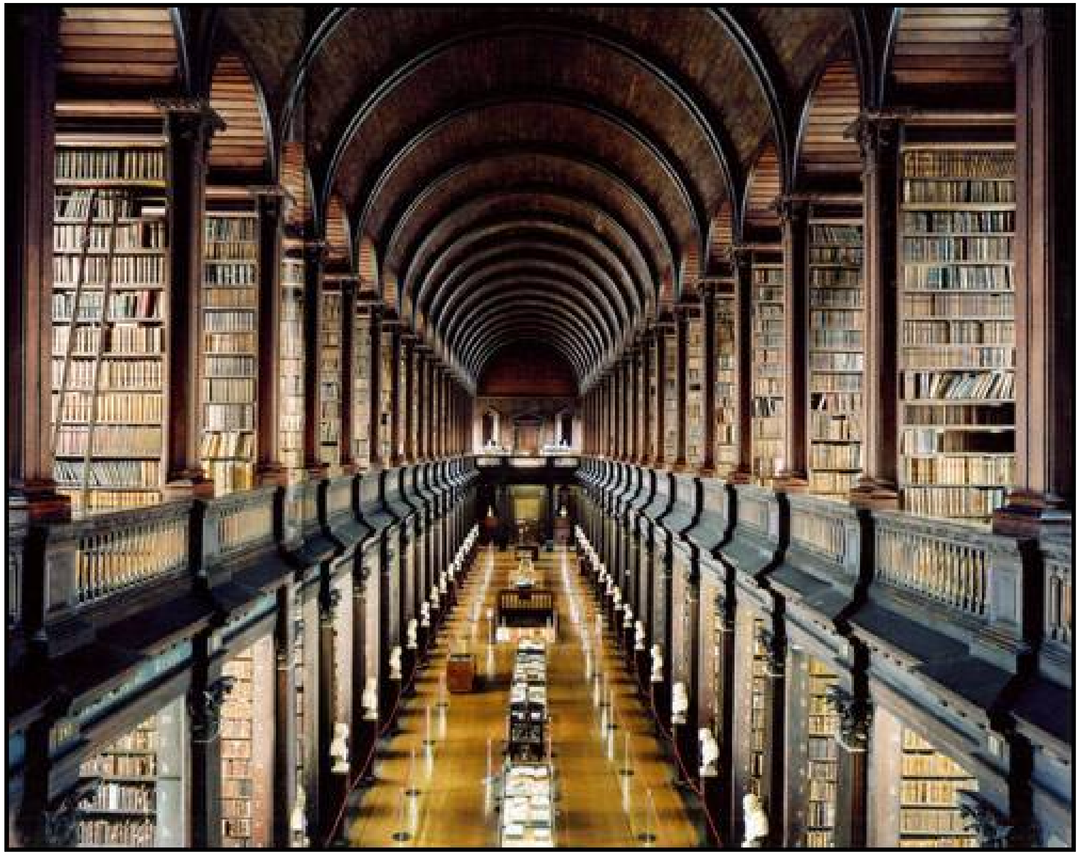
Textual Analysis

- We understand this problem
  - 8 years of research
  - 40+ Papers, 3 patents
  - Hands on experience with DHS, Military, IC, and Industry
- We are very good at it
  - $15M in research investment
  - 19 group members
  - R&D 100 Award (Oscars of invention) in 2007

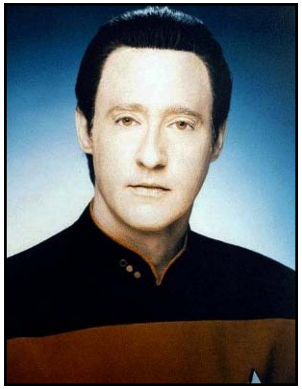
We can read a newspaper, but not a library … without help

THE VICTORIA GAZETTE.

VOL. I. VICTORIA, VANCOUVER ISLAND, FRIDAY, JUNE 25, 1858. NO. 1.



How can computers help?



- The smartest computer can not read a simple first grade level book

- But simple computers can help us find what we need when we need it

Overview of Text Analysis

- **Keyword Methods – Very fast, good for millions**
  - **Search**
    - **“Seafood in DC”**
    - **Good if you know what you are looking for and can find it on the top of the result list**
  - **Unsupervised Classification**
    - **“What were the main topics in message traffic last month?”**
    - **Good to get a general overview a set of messages, though topics may not be valuable**
  - **Supervised Classification**
    - **“What explosive and trigger messages were in last months traffic?”**
    - **Good for finding topics of interest, provided you can describe the topics**

Overview of Text Analysis

- **Full text methods – Slower, good for thousands**
  - **Clustering**
    - **“How are these set of documents related”**
    - **Good for organizing sets of documents done statistically, which may differ from human organization.**
  - **Term frequency Analysis**
    - **“What other words or concepts am I missing”**
    - **Good for linking terms and names, best suited for well written documents**
  - **Semantic Extraction – Slow but parallelizable**
    - **“I am out of ideas, what else can you find”**
    - **Good for the needle in a haystack analysis, but can be very slow.**

How computers can help

Document 1

The Army needs sensor technology to help find improvised explosive devices

Document 2

ORNL has developed sensor technology for homeland defense

Document 3

Mitre has won a
contract to
develop ho
defense se
for explos
devices

Terms

Army
Sensor
Technology
Help
Find
Improvise
Explosive
device

ORNL
develop
sensor
technology
homeland
defense

Mitre
won
contract
develop
devices

Term List

Army
Sensor
Technology
Help
Find
Improvise
Explosive
Device
ORNL
develop
homeland
Defense
Mitre
won
contract

Vector Space Model

	Doc 1	Doc 2	Doc 3
Army	1	0	0
Sensor	1	1	1
Technology	1	1	0
Help	1	0	0
Find	1	0	0
Improvise	1	0	0
Explosive	1	0	1
Device	1	0	1
ORNL	0	1	0
develop	0	1	1
homeland	0	1	1
Defense	0	1	1
Mitre	0	0	1
won	0	0	1
contract	0	0	1

Documents to vectors

Textual Clustering

Vector Space Model

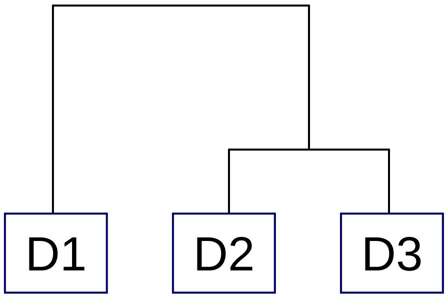
	Doc 1	Doc 2	Doc 3
Army	1	0	0
Sensor	1	1	1
Technology	1	1	0
Help	1	0	0
Find	1	0	0
Improvise	1	0	0
Explosive	1	0	1
Device	1	0	1
ORNL	0	1	0
develop	0	1	1
homeland	0	1	1
Defense	0	1	1
Mitre	0	0	1
won	0	0	1
contract	0	0	1

Similarity Matrix

	Doc 1	Doc 2	Doc 3
Doc 1	100%	17%	21%
Doc 2		100%	36%
Doc 3			100%

Documents to Documents

Cluster Analysis



Most similar documents

TFIDF

$$W_{ij} = \log_2\left(f_{ij} + 1\right) * \log_2\left(\frac{}{n}\right)$$

$$d_2(\mathbf{x}_i, \mathbf{x}_j) = \left(\sum_{k=1} (x_{i,k} - x_{j,k})\right)$$

Time Complexity

$O(n^2 \mathrm{Log}\ n)$

Vectors to trees

Challenge

- **Current computer algorithms that process text work well for small sets of documents**
  - **Average newspaper story .0001 seconds**
- **Not as well for medium size sets**
  - **Encyclopedia Britannica 2.3 days**
- **Infeasible for large sets.**
  - **British newspapers from 1800 – 1900 requires 317 years of processing**

ORNL Breakthrough…

$$W_{ij} = \log_2\left(f_{ij}+1\right) * \log_2\left(\frac{C+1}{c+1}\right)$$

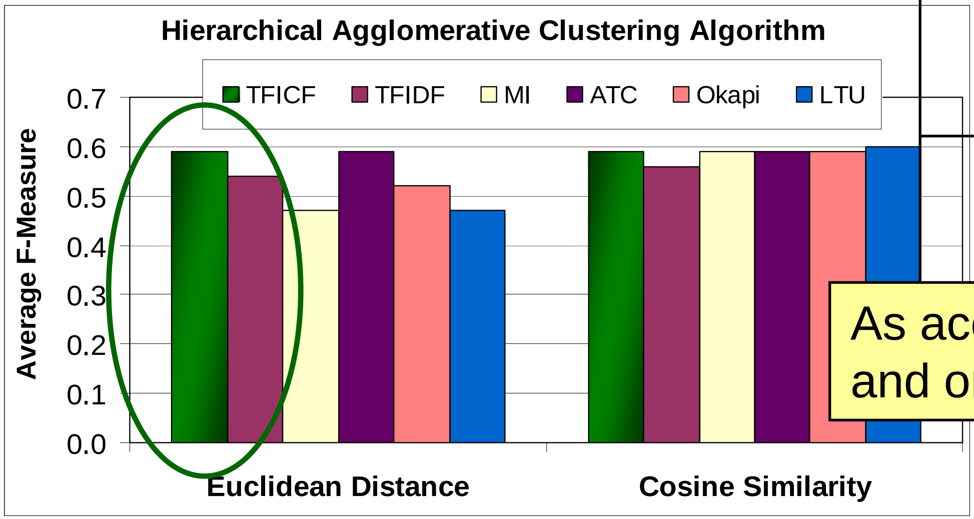
Inverse Corpus Frequency

Test Data Sets

Data Set	# of Docs	# of Classes	Largest Class	Smallest Class
Reuters	2349	58	1041	1
SMART	3891	3	1460	1033
20 News	4650	12	399	385

Term Weighting Schemes

Name	Term Weighting Scheme
TF-IDF	$wij = \log(fij) \times \log(N / nj)$
MI	$wij = \log \dfrac{\dfrac{fij}{N}}{\dfrac{\sum_{i=1}^{N} fij}{N} \times \dfrac{\sum_{j=1}^{M} fij}{N}}$
ATC	$wij = \dfrac{\left(0.5 + 0.5 \times \dfrac{fij}{max\_f}\right)\log\left(\dfrac{N}{nj}\right)}{\sqrt{\sum_{i=1}^{N}\left[\left(0.5 + 0.5 \times \dfrac{fij}{max\_f}\right)\log\left(\dfrac{N}{nj}\right)\right]^2}}$
Okapi	$wij = \left(\dfrac{fij}{0.5 + 1.5 \times \dfrac{dl}{avg\_dl} + fij}\right)\log\left(\dfrac{N - nj + 0.5}{fij + 0.5}\right)$
LTU	$wij = \dfrac{(\log(fij) + 1.0)\log\left(\dfrac{N}{nj}\right)}{0.8 + 0.2 \times \dfrac{dl}{avg\_dl}}$



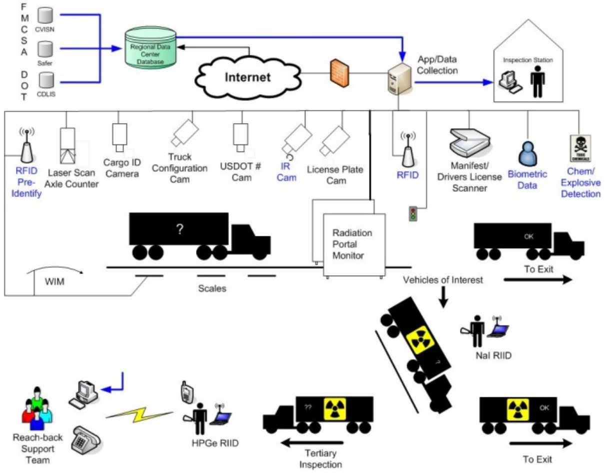
As accurate as current methods and orders of magnitude faster

Capability overview

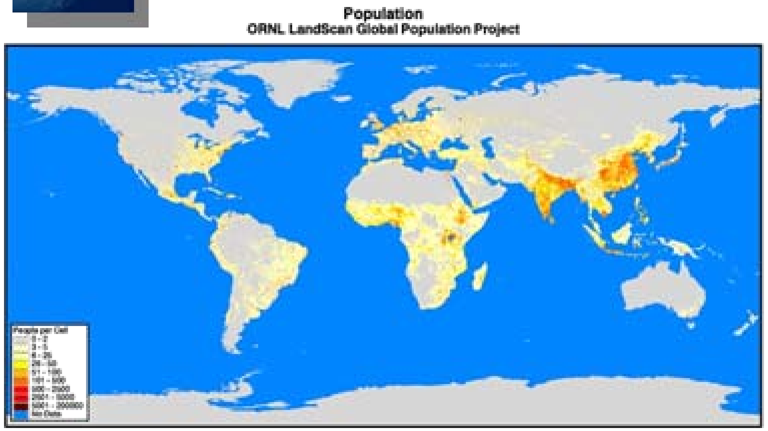
Capability	Capacity in documents	Piranha	Search Engines	Natural Language Processing Tools
Search	100M+	Yes	Yes	No
Unsupervised classification	1M	Yes	Some	No
Supervised classification	1M	Yes	No	No
Clustering	100K	Yes	No	No
Term Frequency Analysis	100K	Yes	Yes, but not available to user	Yes
Semantic Extraction	1000	Yes	No	Yes

Large scale data exploration constrained by wall-clock time to provide decision support.

- **Detect anomalies**
- **Data dip into structured and unstructured data**
- **Inductive hypothesis generation**
- **Human interaction enhanced by real-time data support**
- **Threat anticipation**

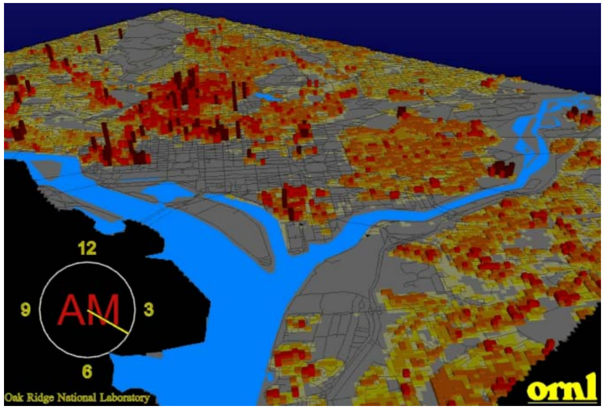


Population Data and Models

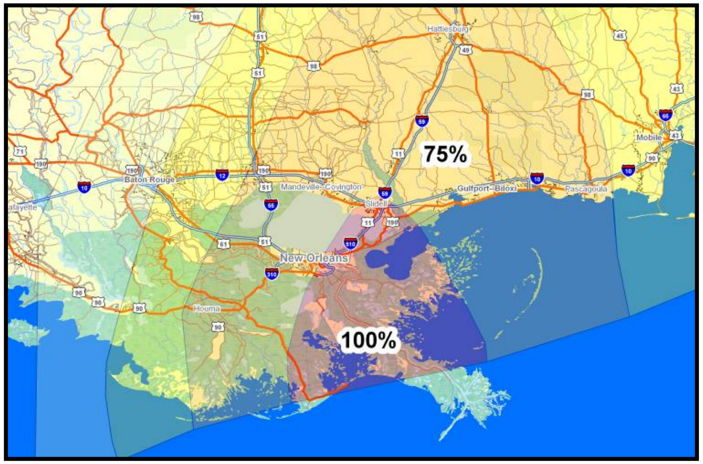


LandScan Global 30”x30”

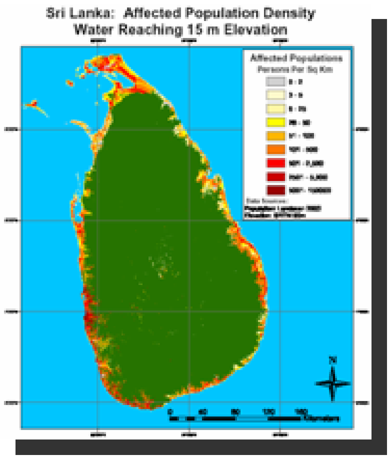
LandScan USA Day/Night 3”x3”



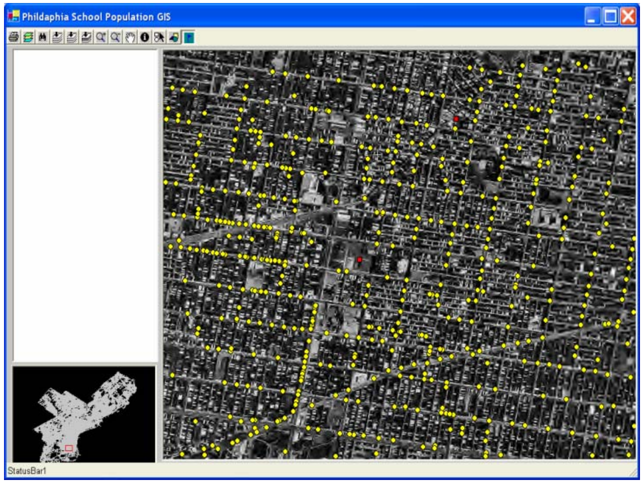
Nominal 24-hour variation



Hurricane Impacts


Sri Lanka: Affected Population Density
Water Reaching 15 m Elevation
Affected Populations
Persons Per Sq Km
N

Tsunami Impacts

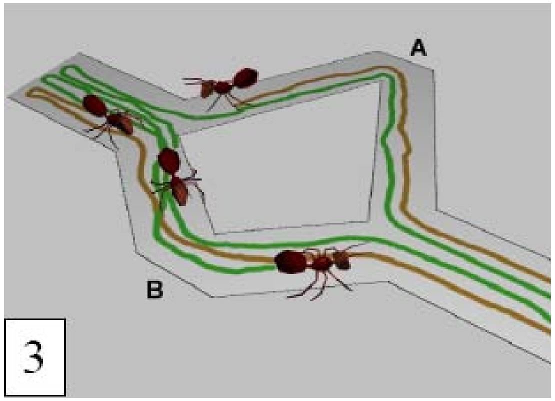


Exposure Impacts

Emergent Behavior in Social Systems

Photograph by José L. Gómez de Francisco
© 2005 National Geographic Society. All rights reserved.
Visions of Earth
National Geographic magazine, March 2005

Birds flocking



Ant pathways

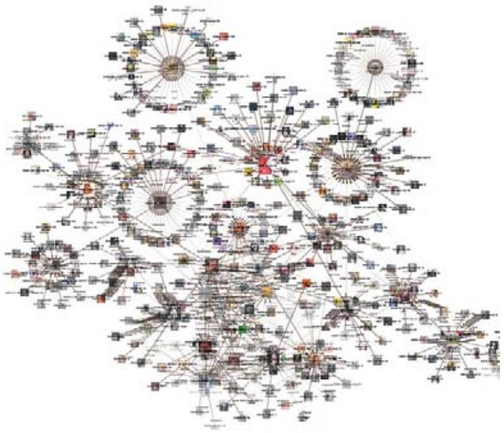
Human response

Agent-based simulations

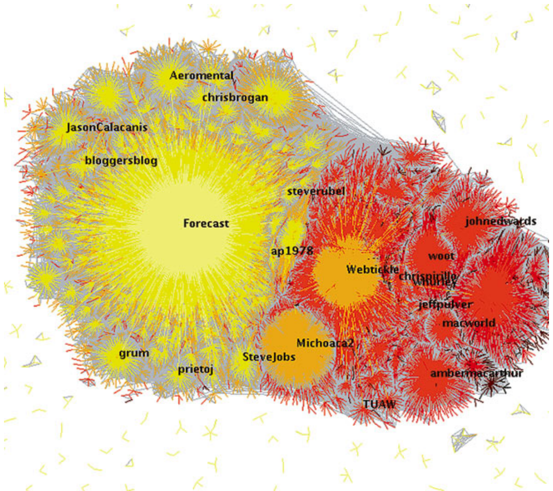
Discrete-event simulations

Social Networks Analysis

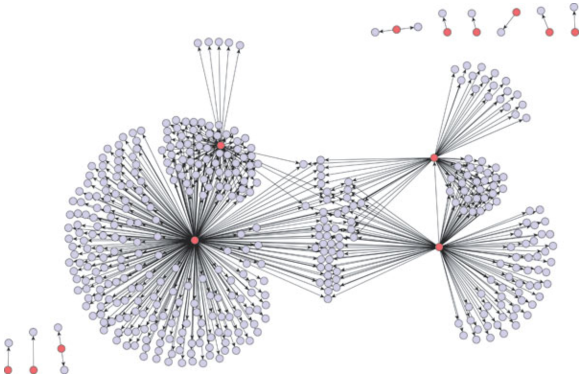
Blogosphere



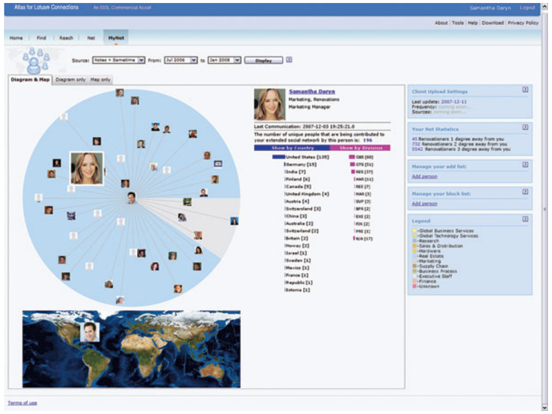
Comment Flow



Twitter Social Network



Viral Marketing



Atlas for Lotuse Connections An ISSL Commercial Asset Samantha Daryn Logout

About | Tools | Help | Download | Privacy Policy

Home | Find | Reach | Net | MyNet

Source: Notes + Sametime From: Jul 2006 to Jan 2008 Display

Diagram & Map | Diagram only | Map only

Samantha Daryn
Marketing, Renovations
Marketing Manager

Last Communication: 2007-12-03 19:25:21.0

The number of unique people that are being contributed to your extended social network by this person is: 196

Show by Country	Show by Division
United States [135]	GBS [60]
Germany [15]	GTS [51]
India [7]	RES [37]
Finland [6]	HAR [11]
Canada [5]	REE [7]
United Kingdom [4]	MAR [3]
Austria [4]	SUP [3]
Switzerland [3]	BPR [2]
China [3]	EXE [2]
Australia [2]	FIN [2]
Switzerland [2]	FRE [1]
Britain [2]	N/A [17]
Norway [2]	
Israel [1]	
Sweden [1]	
Mexico [1]	
France [1]	
Republic [1]	
Estonia [1]	

Client Upload Settings
Last update: 2007-12-11
Frequency: coming soon...
Sources: coming soon...

Your Net Statistics
45 Renovationers 1 degree away from you
732 Renovationers 2 degree away from you
5542 Renovationers 3 degree away from you

Manage your add list:
Add person

Manage your block list:
Add person

Legend
- Global Business Services
- Global Technology Services
- Research
- Sales & Distribution
- Hardware
- Real Estate
- Marketing
- Supply Chain
- Business Process
- Executive Staff
- Finance
- Unknown

Terms of use

Workplace Networks

Images from Technology Review, Vol 111/No 2. March/April 2008

Virtual Worlds to Explore Social Behaviors

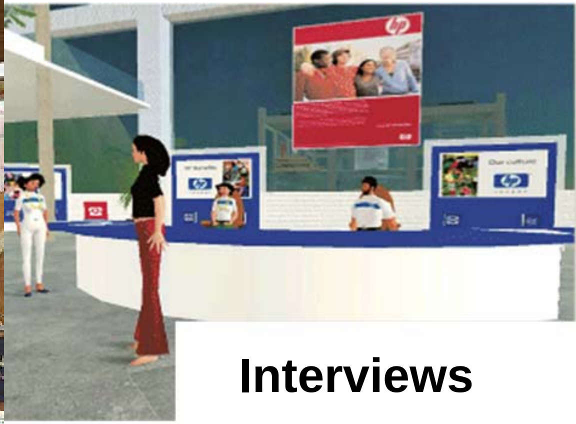
Second Life – Linden Lab

Education

Tourism

Collaboration

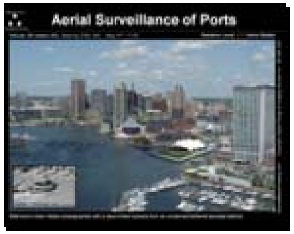
Shopping



Interviews

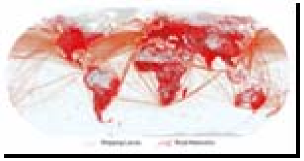
Achieving Systematic Situation Awareness

1 Collect



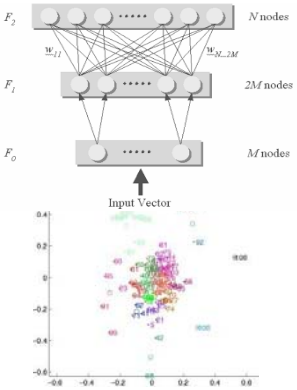
Syndromic Surveillance

Field Sensors



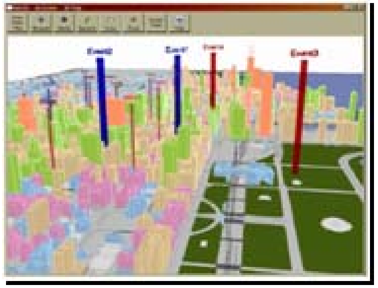
Monitoring

2 Integrate



Knowledge Base

3 Detect

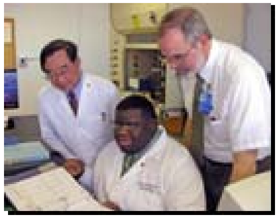


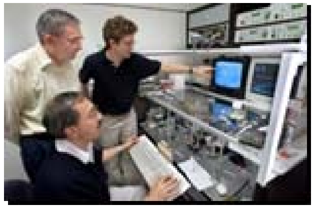
Anomaly Detection

4 Network/Share

Across Agencies

5 Evaluate/Assert





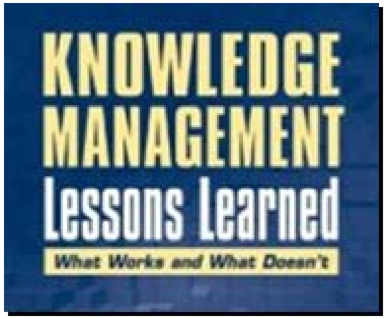
Cross Check

6 Respond

NATIONAL INSTITUTES OF HEALTH

U.S. DEPARTMENT OF HOMELAND SECURITY

7 Reconcile



KNOWLEDGE MANAGEMENT

Lessons Learned

What Works and What Doesn't

Summary

- **Current technology cannot yet solve emerging national challenges in knowledge discovery**
- **Intelligent software agents and associated research areas comprise significant breakthrough technology**
- **Results indicate high-potential to help solve these national challenges**
- **We have a progression of significant and successfully deployed agent systems and research to our credit**