



Networks Grand Challenge News and Science

Tracking Report

Issue #4 (Mid-January through Mid-May 2009)

**A selection of industry, funding, and research news relevant
to the SNL Networks Grand Challenge (NGC) Team**

This issue of the Networks News and Science Tracking Report from Perspectives covers material primarily from mid-January through mid-May 2009, although some important material from other periods is included. For example, older material uncovered as part of the tracking research may be included if it has not been discussed in previous reports.

This document contains items abstracted (edited, direct text or summaries of source material) from the news or other sources. Links are provided to the full text of source material. Occasionally, Perspectives' comments are included (indicated by italic type). Emphasis is ours and is indicated by bolding or underlining. Items of particular interest to us are indicated by a highlighted star.

A table of contents for this tracking report is shown on the next page. The report reflects revised priorities for the NGC team, and includes a section on text analytics / visualization. In addition, we have included a special topical focus section on "digital forensics." The next issue of this report is expected around late July; among other topics it will include a review of issues relating to data loading.

The full text of any source item is available. Contact Ann Miksovic: ann@perspectivesweb.com or (505) 881-0370.

Table of Contents

I. PRIORITY APPLICATIONS	3
A. Terrorism / Intelligence Analysis / Nonproliferation.....	3
B. Text Analytics / Visualization.....	5
1. General Information	5
2. Entity Extraction.....	8
3. Text Summarization.....	9
4. Document Clustering / Automatic Clustering.....	10
C. Cybersecurity / Computing.....	10
D. Big Data.....	13
E. Funding	14
II. OTHER APPLICATIONS AND ITEMS OF INTEREST	16
A. Epidemiology / Medical / Life Sciences / Pharma	16
B. Business Analytics	16
C. Other Items of Interest	17
III. COMPANY NEWS, IN BRIEF	20
IV. RESOURCES / OVERVIEWS	25
A. Reviews and Overviews.....	25
B. Resources on the Web.....	26
V. DIGITAL FORENSICS	27
A. About the Field of Digital Forensics	27
B. Digital Forensic Tools and Toolsets.....	29
C. Other Interesting Work: Automation, Datasets, Visualization, Analytics	31
D. Compendium of Information on Digital Forensics and Commercial Tools.....	34
Table 1: List of Selected Digital Forensic Tools / Toolkits, by Company.....	34
Table 2: Forensic Tool Reviews	41
Table 3: Forensic Tool Lists for Reference	42
Table 4: Conferences, Publications, and Information Sites	43

I. PRIORITY APPLICATIONS

A. Terrorism / Intelligence Analysis / Nonproliferation

- **★ [Can a Computer Model Prevent a War?](#)** – UA engineering is receiving \$2M from the U.S. Army for research into computer models of unconventional warfare. The U.S. Army has awarded another \$2 million to **University of Arizona** Professor Jerzy Rozenblit to fund phase 2 of a project to design intelligent software that can analyze the behavior and customs of political and cultural groups. In 2007, the Army awarded Rozenblit \$2 million to fund the recently completed phase 1 of the **Asymmetric Threat Response and Analysis Project**, known as ATRAP. **The ATRAP software will enable intelligence analysts to build up three-dimensional maps of interactions between conflicting groups.** By mapping behavior, relationships, resources, events and timelines, analysts hope to be able to predict, and therefore prevent, eruptions of violence. ... ATRAP's three-dimensional behavior modeling is based on maps – satellite images, for example – with two dimensions: latitude and longitude. The third dimension, time, is added to create what Brian Ten Eyck [UA department of electrical and computer engineering] describes as a "thought space." This third dimension consists of people, groups and events and their locations in time, but also more abstract entities such as relationships and affiliations. "There's a person here, there's an organization there, an explosion took place here," said Ten Eyck. "These things are all connected and the thought space shows you lines that connect the place where the bomb blew up and where the bomb maker lives, for example. And there's a slider that lets you move back and forth through time so you can see how events unfold or change over time in a particular region." ... ATRAP's potential spreads far beyond the limits of defense, said Rozenblit, and into the financial world, for example, "to combat financial fraud." He also cites disaster relief and epidemiology as areas that could benefit from ATRAP's analytical and mapping capabilities.
- **[DHS publishes concept of operations for fusion centers.](#)** This CONOPS outlines DHS processes relating to intelligence and operational information flows and interactions, deployment of officers, component integration, and identification of SLFC requirements, technical assistance and training. (Full text report [here.](#))
- **[MindCite connects the dots to solve crime.](#)** Billions of dollars have been spent on new tools to fight crime and terror in the US alone. One major obstacle in the road to cracking a case or preventing criminal or terrorist activity, however, is the difficulty of getting data from thousands of agencies who keep their own records and data bases. Enter Israel's [MindCite](#). "Since 9/11, the need to monitor Open Source Intelligence, such as blogs and chat rooms, became apparent," Hadar Himmelman, CEO of MindCite, a data mining intelligence company, says. "Our pioneering technology gathers a huge amount of information on key topics, integrates it with data from various sources, and presents a coherent visual map with precise, focused information to intelligence officers. The system is multi-lingual. No one else offers a full semantic solution." The company's platform technology is described [here](#) and homeland security solutions [here](#).
- **["Postmodern Intelligence: Strategic Warning in an Age of Reflexive Intelligence,"](#)** Myriam Dunn Cavelty and Victor Mauer (ETH Zurich), *Security Dialogue*, 40(2):123-144 (2009). (For a blog review of this article, see [here.](#))

Providing strategic warning to policymakers is a key function of governmental intelligence organizations. Today, globally networked challenges increasingly overshadow their historical state-centric counterparts so that warning has become considerably more difficult. It is recognized in parts of the intelligence community that many of the current problems for warning arise from continued reliance on analytic tools, methodologies and processes that were appropriate to the static and hierarchical nature of the threat during the Cold War. However, even though alternative analysis techniques have begun to be applied, this article argues that the intelligence community could benefit from the understanding that more than just the ontology of threats has changed, that in fact it is in the epistemological area that the

most meaningful changes have taken place: Society has seen the replacement of the previous means—end rationality by a reflexive rationality. The notion of reflexive security can provide a valuable conceptual framework for understanding the current changes, and it could be instrumental in adapting intelligence sources and methods to a new era. In particular, an awareness of both complexity sciences and postmodernism might increase understanding of the limitations of knowledge and lead to the establishment of a political discourse of uncertainty.

- [American Counterinsurgency](#). This article at *Inside Higher Ed* summarizes the views of a vocal critic of the U.S. Army “Human Terrain” program, Roberto J. González, an associate professor of anthropology at San Jose State University and author of a recently published book, *American Counterinsurgency: Human Science and the Human Terrain*. A few comments from Gonzalez: “I wanted to go beyond the headlines, to examine the development of the human terrain concept and how it has been transformed over the years. ... Modeling and simulation programs and dynamic social network analysis are the latest fads in human terrain research. Engineers, computer programmers, and social scientists seek to integrate ethnographic data into predictive computer programs. Each year the Pentagon spends tens of millions of dollars in a quest to find a technological holy grail that forecasts political hot spots – organized protest marches, riots, or full-blown terror attacks. Researchers at the University of Pennsylvania, Dartmouth, Purdue, and other universities are competing with private corporations for these funds. It’s become a real growth industry.”
- **Other items of interest:**
 - The **Electronic Frontier Foundation (EFF)** has published the “[Report on the Investigative Data Warehouse](#),” about the FBI’s single largest repository of operational and intelligence information, after suing the agency under the Freedom of Information Act. The report describes some of the tools in use in this system, as well as nearly 40 databases searchable by the system. (See also [press release](#).)
 - [Clandestine defense hub prepares to open at UM](#). Research site to develop tools to fight future threats: The idea of the Intelligence Advanced Research Projects Activity (IARPA), under construction at [University of Maryland’s] M-Square research park near the main campus, is to investigate new ideas for intelligence agencies that are too preposterous for government bureaucrats or private contractors to consider. The incomplete, \$40 million building that will house IARPA was dedicated in April ...
 - [ODNI Releases Second Data Mining Report](#). The Office of the Director of National Intelligence (ODNI) has released the second annual report, “[Data Mining Report 2009](#),” covering the data mining activities of all elements of the ODNI from January 31, 2008 through January 31, 2009. Constituent elements of the Intelligence Community are reporting their data mining activities to Congress through their own departments or agencies.

B. Text Analytics / Visualization

1. General Information

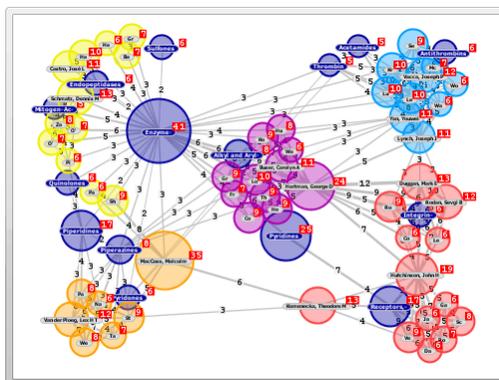
- ★ ['Map of Science' Shows Scientists' Virtual Trails through Online Services](#). Scientists have produced [what they describe as] the world's first Map of Science – a high-resolution graphic depiction of the virtual trails scientists leave behind when they retrieve information from online services. Most studies of scientific activity rely on citation data, which takes a while to become available because both the cited publication and the publication of a particular citation can take years to appear. In other words, citation data observes science as it existed years in the past, not the present. Bollen and colleagues from **LANL** and the **Santa Fe Institute** collected usage-log data gathered from a variety of publishers, aggregators, and universities spanning a period from 2006 to 2008. Their collection totaled nearly 1 billion online information requests. The study is part of the MESUR (Metrics from Scholarly Usage of Resources) project of which Bollen is the principal investigator. **The MESUR usage database is now considered the largest of its kind.** (Research is published as “Clickstream Data Yields High-Resolution Maps of Science,” 2009. Bollen J, Van de Sompel H, Hagberg A, Bettencourt L, Chute R, et al. *PLoS ONE*, 4(3): e4803. doi:10.1371/journal.pone.0004803. Full text, including a high resolution image of the map of science is available [here](#).)
- VizPattern from the Institute for Information Sciences at University of Southern California:** Two papers about ISI's latest visual analytics system “VizPattern” have been submitted to IEEE Symposium on Visual Languages and Human-Centric Computing and IEEE Symposium on Visual Analytics Science and Technology: “[QueryMarvel: A Visual Query Language for Temporal Patterns Using Comic Strips](#),” and “[VizPattern: Interactive Querying of Temporal Patterns by Example](#).” (A video is also available at this [link](#).)

VizPattern allows users to easily and intuitively construct patterns by dragging snippets from these visualizations and composing them into a comic strip. VizPattern provides powerful result visualizations to help users analyze the results, and to use these results to further refine patterns.
- ★ [Visual Data Mining: Theory, Techniques and Tools for Visual Analytics](#) (John Risch of **Future Point Systems, Inc.**; Anne Kao, Stephen R. Poteet, and Y.J. Wu of **Boeing Phantom Works**), Lecture Notes In Computer Science, pp. 154-171 (2008).
- The [Sun Center of Excellence for Visual Genomics](#) at the **University of Calgary**, led by [Christoph Sensen](#), has developed [Bluejay](#) for graphical queries of the genomic system. Bluejay differentiates itself from genome browsers such as OmniGene and Apollo because Bluejay has no built-in expectation of the structure of the data passed to it, and transforms non-XML conformant data on the fly, allowing the user access to most data sources in molecular biology. The flexibility of input will also make it appropriate in the near future as a MOBY Web Services interface. While the primary focus of Bluejay is visualization of data in molecular biology, the developers report that the linear display infrastructure provided by Bluejay could be used in visualizing XML encoded timelines in the humanities, or seismic data.

- iDashboards** provides interactive information dashboards that [Fern Halpern](#) finds “quite engaging and useful” [*we agree*]. A static screenshot of a dashboard is shown to the right, but see this [link](#) for the dynamic dashboard.



- [Touchgraph](#) – provides users with visualizations of links between websites. Try it out [here](#).



- **★ International Science and Engineering Visualization Challenge.** The National Science Foundation (NSF) and *Science* created the International Science & Engineering Visualization Challenge. The spirit of the competition is for communicating science, engineering and technology for education and journalistic purposes. One category is informational graphics. The deadline for all entries for the 2009 competition is September 15, 2009. The first round of judging starts October 5, and the winning entries of the 2009 competition will be published in the February 19, 2010, issue of the journal *Science*.
- **Docsllogic** is a Swiss company that provides visual analytics software that helps healthcare industry professionals quickly discover insights from knowledge about local and regional health networks. A “pioneer in portals for networking, collaboration, and knowledge exchange,” Docsllogic is “now leading the way in visual analysis tools for gaining insights into the increasingly complex and collaborative environment of networked healthcare. Docsllogic has developed the most advanced platform for mapping and analyzing influence networks.” Examples of some of their network mapping and visualizations are [here](#).
- **Google** has posted an “experimental” [flu trends map](#) for Mexico, in the wake of the swine flu threat there. It is unclear how useful these maps would be in providing a marker to possible developing epidemics (see [flu trends maps](#) for various countries and regions).
- **Bioalma** launched [novo|seek](#) in February, a free, text-analytics reliant search engine for biomedical literature in Medline and US Grants. The company aims to compete with Google Scholar and PubMed by combining named entity recognition, information extraction, and knowledge discovery to produce conceptually meaningful results with a greater ease of use than its competitors.
- **Interview with Seth Grimes: Text Analytics (3-part series).** This interview with one of the high profile names in the field includes these highlights:

[Visionary potential] The greatest potential for text analytics is in enabling a computer to pass the Turing Test. Text analytics will decode whatever the tester sends to the computer, it will mine a corpus for response material, and it will support generation of a credible and convincing response. It will do all this throughout a contextualized, conversational exchange complete with noise, external references and anaphora, and multiple topics and voices. That is, text analytics has the potential to enable a computer to understand and talk to people.

... The current hot topics are long-standing applications to life sciences, for instance pharmaceutical drug discovery, and intelligence, and newer applications for functions that include customer support; marketing; media and publishing; insurance, risk, and fraud; search enrichment, etc. ... I see a \$350 million 2008 market, and that figure does not consider the value created by university and industrial research, systems integrators, and custom development, nor [does it consider] the value of the products and capabilities enabled by text analytics.

- ★ **TEMIS** is an award-winning text analytics company headquartered in France that claims to be the European leader in text analytics. This company has an interesting pedigree. Founded 2000 by a team from **IBM** that had developed such products as Text Knowledge Miner and IBM Technology Watch, they signed an initial licensing agreement with **Xerox** for XeLDA®, IBM's "acclaimed linguistic engine." TEMIS later acquired Xerox's linguistics operations. The company offers "breakthrough solutions for indexing and organizing collections of documents and extracting information." Its [core products](#) include an information extractor for text documents, document clustering based on semantic similarity, document classification (see how it works [here](#)), and the multilingual engine, XeLDA, that models and standardizes unstructured documents. XeLDA technology uses Xerox's XFST linguistic technology ([Xerox Finite State Transducers](#)). TEMIS has 50+ employees and operates subsidiaries in Germany and the US. TEMIS products are distributed world-wide through its partner network.

2. Entity Extraction

- ★ **Infoglide** provides a website called "[Identity Resolution Daily](#)" (IRD). This portal, while salted liberally with postings from Infoglide, also tracks other interesting information about entity extraction and related topics. For example,
 - A tidbit from an article about the surnames from the *New York Times*: "By some estimates, 100 surnames cover 85% of China's citizens. Laobaixing, or 'old hundred names,' is a colloquial term for the masses. By contrast, 70,000 surnames cover 90% of Americans." (For another interesting view, see "The Name's Du Xiao Hua, But Call Me Steve – What's up with Chinese people having English names?" in an [article](#) in *Slate* by Huan Hsu.)
 - Dr. John R. Talburt, of the Laboratory for Advanced Research in Entity Resolution and Information Quality ([ERIQ](#)) at the **University of Arkansas** at Little Rock, conducts research addressing important problems related to entity resolution and information quality. (For example, see article by Talburt about "[Solving the False Negative Problem](#)." From the ERIQ website, we learn that an upcoming conference presentation by Talburt deals with "Attributed identity resolution for fraud detection and prevention.")
- Steven Arnold's blog points out two interesting Google patents: 1) US[7536408](#), for phrase detection, important in content processing; and 2) US[7536382](#), for query rewriting with **entity extraction**.
- **Lexalytics**:
 - [Lexalytics Releases Saliency 4.1 Offering Entity Management Toolkit for Better Entity Detection](#). The Entity Management Toolkit allows users to mark up documents specifying entities that are important to them, and the system will then make suggestions on other entities. Once enough documents have been marked up to build a model, the system is ready to process the entire corpus of data. The user doesn't need to write a single line of code...
 - **What is Text Analytics** ([part 1](#), [part 2](#)), a blog article from **Lexalytics**:

... a number of vendors, Lexalytics included, have significantly improved their entity recognition technology in recent months to utilize techniques like "grammatical parsing" and "Max Ent" models to do a better job of extracting entities. I did a complete post a little over a month ago about our new [Entity Management Toolkit](#) which explains how users can now build their own entity recognizers. We aren't the only ones pushing hard on entity extraction, other companies are working on this as well. Especially on **grammatical parsing using anaphora resolution** where "John Smith" and "He" are recognized as the same entity.
- **Northern Light** claims to be able to extract meaning-loaded entities using its MI Analyst capabilities and to provide better, meaning-loaded search results with important text summarization.

Search engines must evolve to have an in-depth understanding of the searched material. It is necessary that the search engine grasp the business purpose for the search and that search goes beyond presenting document lists to users. The search engine must interpret and analyze the search results and then present findings that would be considered most significant by the user if the user were able to read all of the retrieved documents. – Northern Lights CEO (See blog post [here](#) for further explanation.)

And now with MI Analyst 3.0, users see immediately the concepts critical to assessing a document in the list of search results. This might include the technologies mentioned in the article, companies mentioned, business issues mentioned (regulatory action, strategic partnerships, emerging market, etc.), venture-funded companies mentioned, and even bookmark info ("This document has been saved by 1 person with tags like Case studies, Cisco, voip"). And users can click to see an expanded document detail page with a full listing of extracted entities-all loaded with meaning. ([Source](#))

- A concept-connecting information visualization, created by **Evri**, may be of interest. This is used sometimes by the *Washington Post online* in certain articles for persons of interest. For an example, see this [treatment](#) of Ben Bernanke, Federal Reserve Chairman. According to one posting on a blog:

The site does a number of really interesting things. Firstly, it attempts to solve the named entity extraction problem in a broad way. Named entity recognition is often limited to person names, places and organizations. Evri doesn't seem to have any limit to the types of things it discovers - music, bands, movies, books. Secondly, it looks for relationships between those entities. This is largely via collocation in a document. Thirdly, it attempts to disambiguate concepts with more than one possible type, thus Blue, which could be a film, a band or an album (not to mention a colour) is disambiguated. Finally, it gives access to the web via an interface which allows the user to both search and wander across the relationships between entities.

3. Text Summarization

- **Cirilab** has released the Cirilab Knowledge Generation Engine™ (KGE) which is at the heart of a series of [products](#) designed to "read" unstructured text and provide insights into that text. SpeedRead, for example, provides quick text summarization (described [here](#)) and KnowledgeView extracts key themes, and provides a summary that is 80% smaller than the original, according to the company. The ThemeReader product transforms the themes into a visual map (see also a blog entry about the products [here](#)). The company uses "highly complex algorithms" and "multidimensional semantic spatial indexing" technology.
- **Document Summarization using Wikipedia** (Krishnan Ramanathan, Yogesh Sankarasubramaniam, Nidhi Mathur, and Ajay Gupta in a *HP Technical Report*) – Small screens used by mobile devices are creating a demand for document summarization:

Although most of the developing world is likely to first access the Internet through mobile phones, mobile devices are constrained by screen space, bandwidth and limited attention span. Single document summarization techniques have the potential to simplify information consumption on mobile phones by presenting only the most relevant information contained in the document. In this paper we present a language independent single-document summarization method. We map document sentences to semantic concepts in Wikipedia and select sentences for the summary based on the frequency of the mapped-to concepts. Our evaluation on English documents using the ROUGE package indicates our summarization method is competitive with the state of the art in single document summarization.

4. Document Clustering / Automatic Clustering

- [Weitkamper Technology](#) is a German firm that has been getting press recently on its search offerings. The company offers one facility, [XSEARCH Clustering](#), which groups results into thematically similar clusters, which are generated dynamically, “on the fly,” “in real-time at the moment of search. XSEARCH Clustering Engine shows all results at one glance, grouped thematically. Irrelevant information can be immediately sorted out. The Clustering Engine combines intelligent semantic algorithms with complete linguistic preparation, according to the company. *Perspectives ran a search using their online demo for < “automatic document clustering” OR “automatic clustering” > with some interesting, but limited, results. See [here](#).* The company also offers [clusterpat](#), which uses their clustering function to search, merge and cluster patent hits from the USPTO and European patent collections (Espacenet, WIPO). An interview with the CEO about the company’s search products may be found [here](#).
- ★ [Content Analyst](#) (Reston, VA), funded by [In-Q-Tel](#) in the past, owns important intellectual property related to **Latent Semantic Indexing** (originally developed by **Bell Labs**) and has worked with text analytics of interest to the intelligence community. **SAIC**, a previous owner of the company’s assets, maintains a minority interest. It claims to be “unequaled in its ability to categorize or classify documents based on representative examples, and then sort or identify unstructured information either in a batch operation or instantly, on-the-fly as information is encountered.” In general, it provides automatic categorization, concept search, cross-lingual search, automatic summarization, and instant context technology to its customers. With regard to document summarization, it uses “advanced mathematics to allow users to find similar, related, and relevant documents based purely on the concepts those documents **are discussing – without using keywords, and without getting back ‘matching’ yet irrelevant content.**” [dtSearch](#) is one of its [technology partners](#). (More on the technologies [here](#).) A blog [Q&A](#) was recently published about reusing categories developed in eDiscovery sampling and classification projects as “seeds” for later projects.

C. Cybersecurity / Computing

- [Agent Logic Announces Multi-INT Analyst Solution to Bridge the Gap Between Cyber Threat Data and Traditional Intelligence Sources](#) – Using its flagship product RulePoint®, Agent Logic provides non-technical cyber analysts with a real-time multi-INT operational view of cyber threats. capabilities to deliver real-time intelligence to non-technical staff while bridging the gap between cyber threat information and traditional sources of intelligence.

Agent Logic has harnessed its software products, experience, and best practices in multi-INT processing for national security to deliver a tailored solution that helps organizations understand and thwart tactical and strategic threats to cyber infrastructure.

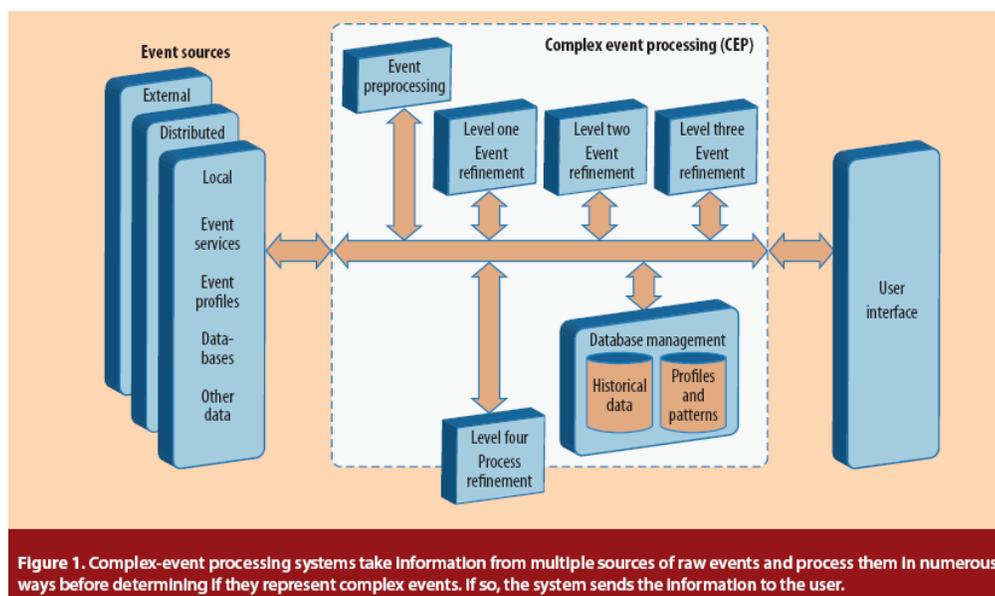
... analysts cannot wait for IT to re-program rules or systems. Using Agent Logic’s cyber solution, non-technical cyber analysts can correlate data collected by the tools used by traditional network security analysts with sensor feeds, geospatial data, field reports, internal databases, and other relevant data sources. This multi-INT analysis capability enables the cyber analyst to develop the same operational picture that characterizes conventional intelligence disciplines, such as counter-terrorism, maritime domain awareness, and weapons proliferation. Leveraging RulePoint’s web-based capabilities, analysts with different expertise, such as low-level network analysts and strategic intelligence analysts, can easily collaborate and share relevant information as needed across domains.

The cyber solution uses RulePoint to streamline the collection, analysis, and dissemination of key intelligence events across cyber and non-cyber data. Analysts create rules that ingest events, derived from traditional network security tools such as intrusion detection systems, firewalls, and other forensic type tools, and combine these events with other intelligence sources to identify and correlate events of interest, escalate responses, pass along data to other systems, add to watch lists, or present in geospatial contexts.

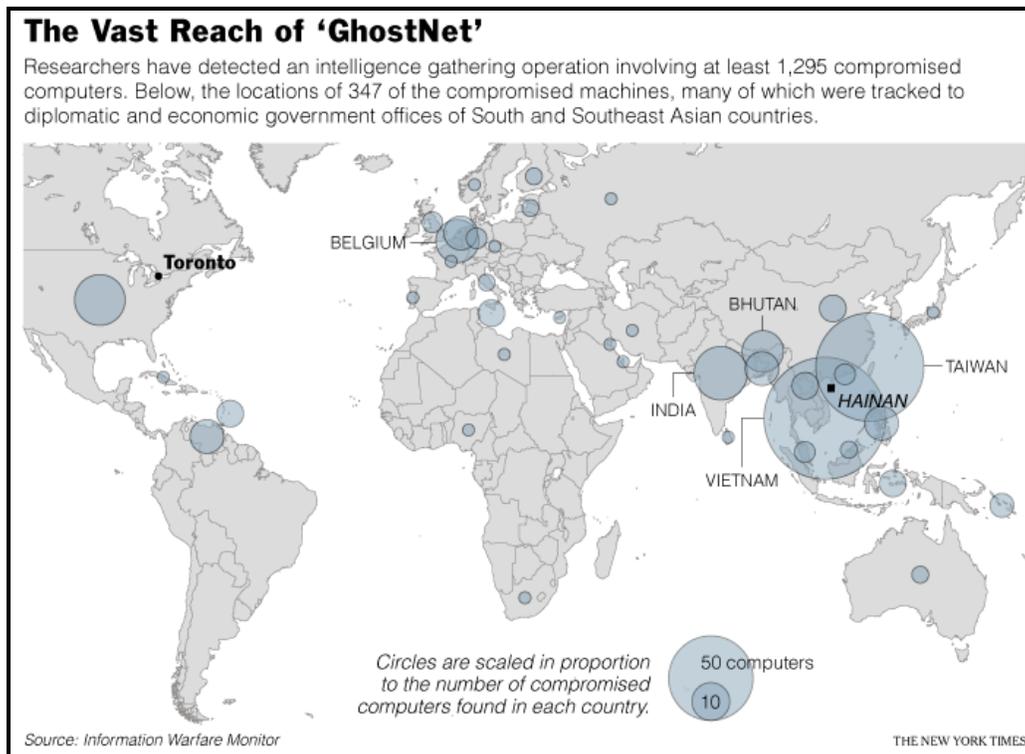
"Agent Logic has a great reputation and track record for solving complex intelligence problems," said Bob Flores, former Chief Technology Officer of the Central Intelligence Agency, and currently CEO of Appicology. "They have once again risen to the challenge by utilizing RulePoint to improve situational awareness of critical events. Their solution will be a key part of any organization's arsenal where the mission is to minimize or eliminate cyber threats," he said.

Michael Appelbaum, Agent Logic's CEO, said, "Using our software, non-technical cyber intelligence analysts can instantly create and change personalized rules that correlate incoming data feeds on-the-fly. By applying these capabilities simultaneously across disparate sources such as network threat tools, traditional intelligence data, internal repositories, and watch lists, analysts are provided with a more accurate, actionable, and timely threat picture. This powerful advance in automated threat detection, opportunity discovery, and collective intelligence enables the government to effectively connect cyber threat data with existing information assets--creating situational dominance and ultimately improving the security of our country."

- IEEE's *Computer* featured **Agent Logic** in a recent article, "[Complex Event Processing Poised for Growth](#)." According to the article, CEP technology could be used for numerous applications. For example, it could analyze events within a network, power grid, database, or other large system to determine whether it is the target of an attack, is performing optimally, or is experiencing problems. Data from simulations could be run through CEP systems to analyze proposed business processes and other activities to determine whether they would be inefficient, cause legal problems, or violate customers' service-level agreements. The technology could examine internal corporate activities to determine whether they conform to government regulations and corporate policies. Similarly, regulators could use CEP to look through organizations' business activities to determine whether they are violating laws, such as by committing insider stock trading. Barriers to use include cost – CEP systems are estimated by one source to cost between \$100 and \$250k. However, prices are expected to come down, and the market revenues generated by CEP products are estimated to more than double from \$180 million in 2008 to \$370 million this year and increase to \$460 million in 2010. According to IBM's Etzion, "The need for CEP is obvious. It represents one of the next big competitive advantages." A schematic of the way CEP systems work is shown below.

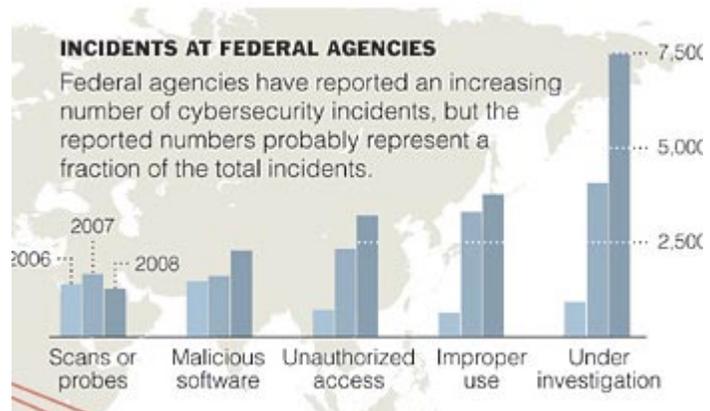


- [The Flourishing Fraud Economy](#). Online scammers who sell stolen credit card numbers, bank account data or Social Security information haven't felt much pain from the economic problems facing the rest of the world, according to new research. One factor is a steady supply of raw materials – phishers have targeted the growing number of people looking for mortgage and credit fixes with specially crafted scams to steal their info. ... Stolen credit card numbers now go for as little as 6 US cents each, if they're bought 10,000 at a time. The price can be \$30 per card for smaller orders. Access to hijacked e-mail accounts: 10 cents to \$100. Bank account credentials: \$10 to \$1,000.
- [Vast Spy System Loots Computers in 103 Countries](#) (*New York Times*, March 31). This fascinating article reports on the **Information Warfare Monitor's** identification of GhostNet (see [report](#)), using **Palantir's** software. Part of the team working on the project from the **University of Cambridge** is releasing their own report (see [here](#)). [See also Palantir item [below](#).]



(Source: Information Warfare Monitor and the New York Times)

- [U.S. Steps Up Effort on Digital Defenses](#). A new international race has begun to develop cyberweapons and systems to protect against them. (Part of a new series by the *New York Times* on "In Cyberweapons Race, Questions Linger over U.S. Offensive Capabilities.")



D. Big Data

- ★ **Cloudera / Hadoop:** [Hadoop, a Free Software Program, Finds Uses Beyond Search](#). (*New York Times*) Three top engineers from Google, Yahoo and Facebook, along with a former executive from Oracle, announced a start-up called **Cloudera**, based in Burlingame, Calif., that will try to bring Hadoop's capabilities to industries as far afield as genomics, retailing and finance. Hadoop is the open source implementation of MapReduce, a powerful tool designed for deep analysis and transformation of very large data sets. Cloudera raised \$5M in Series A funding, and its investors and advisors include Diane Greene (former CEO **VMware**), Mike Abbott (senior VP, **Palm**), Caterina Fake (co-founder, **Flickr**), Dr. Qi Lu (president of the Online Services Group, **Microsoft**; former executive vice president, **Yahoo!**), Marten Mickos (former CEO, **MySQL**), Jeff Weiner (president, **LinkedIn**; former senior vice president, Yahoo!), Gideon Yu (**Facebook** CFO; former CFO at **YouTube**). ([Source](#))
- ★ **I.B.M. Unveils Real-Time Software to Find Trends in Vast Data Sets.** IBM spent close to six years working on the software and has just moved to start selling a product based on it called **System S**. The company expects it to encourage breakthroughs in fields like finance and city management by helping people better understand patterns in data. "Instead of creating separate large databases to track things like currency movements, stock trading patterns and housing data, the System S software can meld all of that information together. In addition, it could theoretically then layer on databases that tracked current events, like news headlines on the Internet or weather fluctuations, to try to gauge how such factors interplay with the financial data."
- Pervasive Software Launches Pervasive DataRush.** DataRush is "a high-performance, embeddable software platform for the next generation of data-intensive processing and analytics. Pervasive said the transformative technology addresses the gap between proliferating multicore processors, exploding volumes of data and the lack of analytic software to fully exploit industry hardware and efficiently extract useful intelligence."
- The Unreasonable Effectiveness of Data,** ([full text](#)) by Alon Halevy, **Peter Norvig**, and Fernando Pereira, (Google), *IEEE Intelligent Systems*, March/April 2009. "We should stop acting as if our goal is to author extremely elegant theories, and instead embrace complexity and make use of the best ally we have: the unreasonable effectiveness of data." *This article is about semantic interpretation using big data and has caused a lot of discussion – for example, see this [blog](#).*
- Brain Mapping Time Reduced from Years to a Few Months with New Technology.** The Scientific Computing and Imaging Institute at the **University of Utah** developed software to automatically merge thousands of images into gigabyte-scale mosaics and align the mosaics into terabyte-scale volumes. And in parallel, teams at the **Moran Eye Center** developed TEM-compatible molecular probes and classification software to tag every cell with a molecular signature, creating "color" TEM imaging. As part of this project, the authors will soon reveal the very first molecular level map of the entire retina and neuronal networks in both a normal mammalian retina and genetic models of retinal degeneration. More than 92% of this 400,000-image volume has been built and should be complete by mid-April. (Journal reference: Anderson et al. "[A Computational Framework for Ultrastructural Mapping of Neural Circuitry](#)." *PLoS Biology*, 2009; 7 (3): e74.)
- The Cellphone, Navigating Our Lives.** **Google** started a location-aware friend-finding system called **Latitude** in 27 countries early this month. On its face, Google's new service – available on dozens of mobile systems – is simply a way for friends to keep track of one another and meet up, for families to stay in touch or for parents to find comfort in knowing where their children are. But it will generate a gold mine of new information about where millions of people travel each day, and there is no doubt that Google and others are planning to dig in that mine. ... Recently, for example, Sam Altman, a 23-year-old Stanford University computer science graduate and the founder of Loopt, a pioneering friend-finding service, was having dinner in Palo Alto, Calif., when he noticed from the screen on his phone that his freshman college roommate was having dinner just two restaurants

away. The two met after dinner at a bar, where they were joined by another former Stanford student who noticed on his display that they were socializing together.

- ★ [Science gleans 60TB of behavior data from Everquest 2 logs](#). Thanks to a partnership with **Sony**, a team of academic researchers have obtained the largest set of data on social interactions they've ever gotten their hands on: the complete server logs of Everquest 2, which track every action performed in the game. ... Noshir Contractor described how the data was allowing him to explore social network dynamics within the game. He described a variety of factors that are thought to influence the growth and extent of social networks, such as collective action, social exchange, the search for similar people, physical proximity, friend-of-a-friend (FoaF) interactions, and so on. Because these are well-developed concepts, statistical tools exist that can extract their signature from the raw data by looking at interactions like instant messaging, partnerships, and trade.
- Big Data: Technologies and Techniques for Large Scale Data** ([excerpt](#), [blog interview](#)).

E. Funding

- New NSF Grants:**

Title	Start Date	Principal Investigator	Organization	Awarded Amount to Date
A Bayesian framework for analyzing genetic and gene expression architecture of complex phenotypes	09/01/2009	Mezey, Jason	Cornell University - State	\$516,392
Complexity Regularization in Statistical Learning Theory	08/01/2009	Koltchinskii, Vladimir	GA Tech Research Corporation - GA Institute of Technology	\$217,989
DAT: A Visual Analytics Approach to Science and Innovation Policy	07/15/2009	Ribarsky, Martin	University of North Carolina at Charlotte	\$746,571
SBIR Phase I: Large-Scale Social Network Analysis Software Services for the Telecommunications Industry	07/01/2009	Eagle, Nathan	NDM Labs (New Mexico)	\$100,000
SGER: 2- and 3-D Visualization of Ecological Phenomena - Transition to the Petabyte Age	05/01/2009	Cushing, Judith	Evergreen State College	\$74,997
CAREER: An Integrated Approach For Efficient Privacy Preserving Distributed Data Analytics	02/01/2009	Kantarcioglu, Murat	University of Texas at Dallas	\$80,000
SBIR Phase I: Correlating Opinions with Outcomes in Business and Industry: Statistical Modelling of Natural Language Data	01/01/2009	Pierce, David	Jodange Inc	\$100,000
CAREER: A Framework for Sparse Signal Reconstruction for Computer Graphics	06/01/2009	Sen, Pradeep	University of New Mexico	\$495,513
CAREER: Categorical Shape Reconstruction	03/01/2009	Zhang, Li	University of Wisconsin-Madison	\$112,011

- **★ EU VisMaster Project** (“visual analytics, mastering the information age”), is a 24-month study funded for € 797.4k (\$) and involving some 25 organizations, including **Business Objects S.A. (SAP)**. It is aimed as forming a strong visual analytics research community in Europe. ([Project abstract](#))

With VisMaster, we want to push the limits of today's Visual Analytics. ... Specifically, the working groups will join excellence in the fields of data management, data analysis, spatial-temporal data, and human visual perception research with the wider visualisation research community. The VisMaster Project main goals are to: 1) form and shape a strong European Visual Analytics community, 2) define the **European Visual Analytics Research Roadmap**, 3) expose public and private stakeholders to Visual Analytics technology, and 4) set the stage for larger follow-up Visual Analytics research initiatives in Europe. ([Website](#))

Put simply, visual analytics entails appropriately combining the strengths of intelligent automatic data analysis with the visual perception and analysis capabilities of the human user. One application, for example, is in the **analysis of large-scale network traffic**. The problem of malicious activities in internet traffic has resulted in the need for companies and public institutions to carefully analyse their traffic share. Visual analytics can be used to gain a faster understanding of situations that could be of threat. ([Press release](#))

Funding Opportunities:

- **Military Networking Technology for Global Information Exchange (GIE)**, [BAA-03-15-IFKA](#). Open and effective until September 30, 2008. AFRL Rome Research Site is soliciting white papers for concept developments, experiments, and demonstrations involving new and innovative approaches to support future Air Force networking, services, and information assurance requirements in the context of a Global Information Grid (GIG). BAA includes: Network Modeling and Simulation – Efforts should research and develop modeling and simulation tools that are trustworthy to predict, with known and characterizable accuracy, network behavior over a broad range of time scales, network sizes and technology composition. .
- **Special Capabilities in Information and Surveillance (SCIS)**, [BAA-08-07-RIKA](#). Open through September 30, 2012. The Air Force Research Laboratory (AFRL), Sensors Directorate – Rome, NY is soliciting white papers under this BAA for the performance of research, development, design, and testing that directly supports its core mission. Topics include: Novel approaches to monitor, visualize and recommend alternative courses-of-action in the management and defense of massive, heterogeneous wide-area networks (5 million nodes); Computer forensic and software protection technologies; Innovative methods to visualize complex, self-organizing systems; and Novel three-dimensional data visualization and projection methods.
- **Rapid Architecture and Web Support**, [BAA-CTO-09-01](#). Open through January 14, 2010; reviews will be conducted continuously on receipt of concept papers. The Defense Information Systems Agency (DISA) is soliciting innovative research proposals in areas of interest. Topics include: innovative proposals that provide enhancements that focus on providing the Intelligence Community with enhanced Computing and Communications Infrastructure capabilities providing agile, adaptive, and capabilities-based IT. These capabilities include wideband networking integrated with smart remote data storage, data conferencing and collaboration, and search and visualization.
- **★ Web-Enabled Temporal Analysis System Tool Kit (WebTAS TK)**, [FA8750-09-R-0022](#) (responses to the RFP were due in April). AFRL Rome has a requirement to develop and integrate new technology into real-time spiral releases of the Web-Enabled Temporal Analysis System Tool Kit (WebTAS TK) software... WebTAS TK [is] a software toolkit that provides a general data visualization and analysis infrastructure for the analysis of temporal, spatial, entity and association information. The flexible WebTAS TK architecture provides access to multiple data sources, data mining, and collaboration tools that facilitate trend, link, pattern, and distributed analysis to support activity prediction and customer unique capabilities **[NOTE: The [solicitation text](#) has a review of the system's current capabilities and a list of selected areas where WebTAS has been used – see “statement of scope”.]**

II. OTHER APPLICATIONS AND ITEMS OF INTEREST

A. Epidemiology / Medical / Life Sciences / Pharma

- Upcoming lecture by **Albert-Laszlo Barabasi** from the Center for Complex Networks at **Northeastern University**:
 The evolution of networks is governed by simple and generic laws, which result in apparently universal architectural features. In his lecture, Barabasi will highlight how understanding this architecture shows not just the potential of these networks, but also their vulnerabilities. This leads to understanding the weaknesses in massive networks like the Internet and how to disrupt the network of a disease, such as cancer. Barabasi, who previously was a professor in the Department of Physics at the University of Notre Dame, will also discuss the amazing order that characterizes these connections, and the resulting implications for communications and medicine. ([Source](#))
- ★ [Washington company raised early alarm about swine flu](#). **Veratect** raised the first warning about a possible outbreak of swine flu in Mexico more than two weeks before the World Health Organization offered its initial alert. Veratect "automatically searches tens of thousands of Web sites daily for early signs of looming medical problems or civil unrest anywhere in the world. Anything of interest is turned over to a team of 35 analysts to determine its significance and post on the company's Web site. The company markets access to its Web site to government agencies, businesses and others and has tried unsuccessfully to sell its service to the CDC, WHO, and DHS." ([Product list](#)) [*Government agency officials appear to believe the approach is interesting but untested.*]

B. Business Analytics

- [Data Mining Case Heads to the Supreme Court](#). Two major publishers of health care data filed a petition in late March at the Supreme Court, raising cutting-edge questions about whether increasingly widespread data mining that is used for commercial purposes is protected by the First Amendment.
- [Facebook Does Some Science, Concludes Redesign is Good for Ads](#). Facebook conducts social network analysis to determine how many people a typical user communicates with, and adds a stream feed to essentially increase indirect communications, which will positively affect advertising revenues.

Facebook's latest design and interface changes have been variously despised and applauded by its members. One of the biggest and most controversial changes was to make every user's homepage look and feel more like Twitter, which is a continuous "stream" of data that updates in real time ... The [average communication] figures were lower than you may expect: a user with 150 friends typically communicates reciprocally (conversation-style) with just five people, and is in direct communication with nine people in total. But if you factor in "stream communication," which implies that by reading the stream of data flowing in from your various contacts you're ephemerally "in touch" with them, then the average user communicates with 20 people. ... the stream feed could be viewed as a potent advertising tool that's between two and four times more effective than other communications systems within Facebook. By using in-stream messages, an advert could access a far greater audience and do so very swiftly.

C. Other Items of Interest

- ★ [Random Network Connectivity can be Delayed, but with Explosive Results, New Study Finds](#). Mathematicians studying networks in which the formation of connections is governed by random processes have provided new evidence that super-connectivity can be appreciably delayed. But the delay comes at a cost: when it finally happens, the transition is virtually instantaneous, like a film of water abruptly crystallizing into ice. The team's findings – described in a paper, “[Explosive Percolation in Random Networks](#),” in the March 13 issue of the journal *Science* ([abstract](#)) – could be useful in a number of fields: from efforts by epidemiologists to control the spread of disease, to communications experts developing new products. “We have found that by making a small change in the rules governing the formation of a network, we can greatly manipulate the onset of large-scale connectivity,” said Raissa D'Souza, an associate professor of mechanical and aeronautical engineering at **UC Davis**.
- ★ “[Navigability of complex networks](#)” (Marián Boguñá, Dmitri Krioukov & K. C. Claffy) *Nature Physics* 5, 74 - 80 (2009).

Routing information through networks is a universal phenomenon in both natural and man-made complex systems. When each node has full knowledge of the global network connectivity, finding short communication paths is merely a matter of distributed computation. However, in many real networks, nodes communicate efficiently even without such global intelligence. Here, we show that the peculiar structural characteristics of many complex networks support efficient communication without global knowledge. ... Our findings suggest that real networks in nature have underlying metric spaces that remain undiscovered. Their discovery should have practical applications in a wide range of areas where networks are used to model complex systems.
- [Smarter Searches: Technology Merges Images, Data and Knowledge](#). This article, focusing on the work of researchers at **UT-Dallas**, reports that members of the School of Engineering, the Cybersecurity Research Center, and other groups, are working with social network analysis and other techniques to integrate geospatial information with other types of information. Some interesting information from the article: members are jointly developing tools with **Raytheon**; work is funded by **IARPA** and the **National Geospatial Intelligence** agency; the team has already developed a system called DAGIS (Discovering Annotated Geospatial Information Services), a semantic Web framework for geospatial information that the team subsequently extended to handle queries related to police blotter data; and the team – consisting of Bhavani Thuraisingham, Latifur Khan and their computer science colleagues Murat Kantarcioglu and Kevin Hamlen – have brought in more than \$9 million in contracts and grants in the past four years (including a number of grants from DOD agencies).
- An interview with [David Bader](#), Executive Director of High-Performance Computing in the College of Computing at **Georgia Institute of Technology** contains a discussion of his work on **SNAP** – an open-source graph analysis framework that builds on Georgia Tech's SWARM library and supplements existing static graph algorithms with relevant ideas from dynamic graph algorithms, social network analysis, and parallel and multicore processing. “SNAP provides a simple and intuitive interface for the network analyst, effectively hiding the parallel programming complexity involved in low-level algorithm design from the user while providing a productive high-performance environment for complex network queries.” ([Source](#))
- “[Reproduction of Hierarchy? A Social Network Analysis of the American Law Professoriate](#),” Katz, Daniel Martin, Gubler, Joshua, Zelner, Jon, Provins, Eric A. and Ingall, Eitan M., SSRN. For a blog post commenting on the article, and comparing results to the *US News and World Reports* rankings, see [here](#). Paper abstract:

... Leveraging advances in network science and drawing from available information on the more 7,200 tenure-track professor employed by an ABA accredited institution, we explore the

topology of the legal academy including the relative distribution of authority among its institutions. Drawing from social epidemiology literature, we provide a computational model for diffusion on our network. The model provides a parsimonious display of the tradeoff between "idea infectiousness" and structural position. While our model is undoubtedly simple, our initial foray into computational legal studies should, at a minimum, motivate future scholarship.

- ★ [A Graph Analysis of the Linked Data Cloud](#), by Marko A. Rodriguez (March 2, 2009). ([full text](#))
 Abstract: The Linked Data community is focused on integrating Resource Description Framework (RDF) data sets into a single unified representation known as the Web of Data. The Web of Data can be traversed by both man and machine and shows promise as the *de facto* standard for integrating data worldwide much like the World Wide Web is the *de facto* standard for integrating documents. On February 27th of 2009, an updated Linked Data cloud visualization was made publicly available ([here](#)). This visualization represents the various RDF data sets currently in the Linked Data cloud and their interlinking relationships. For the purposes of this article, this visual representation was manually transformed into a directed graph and analyzed.

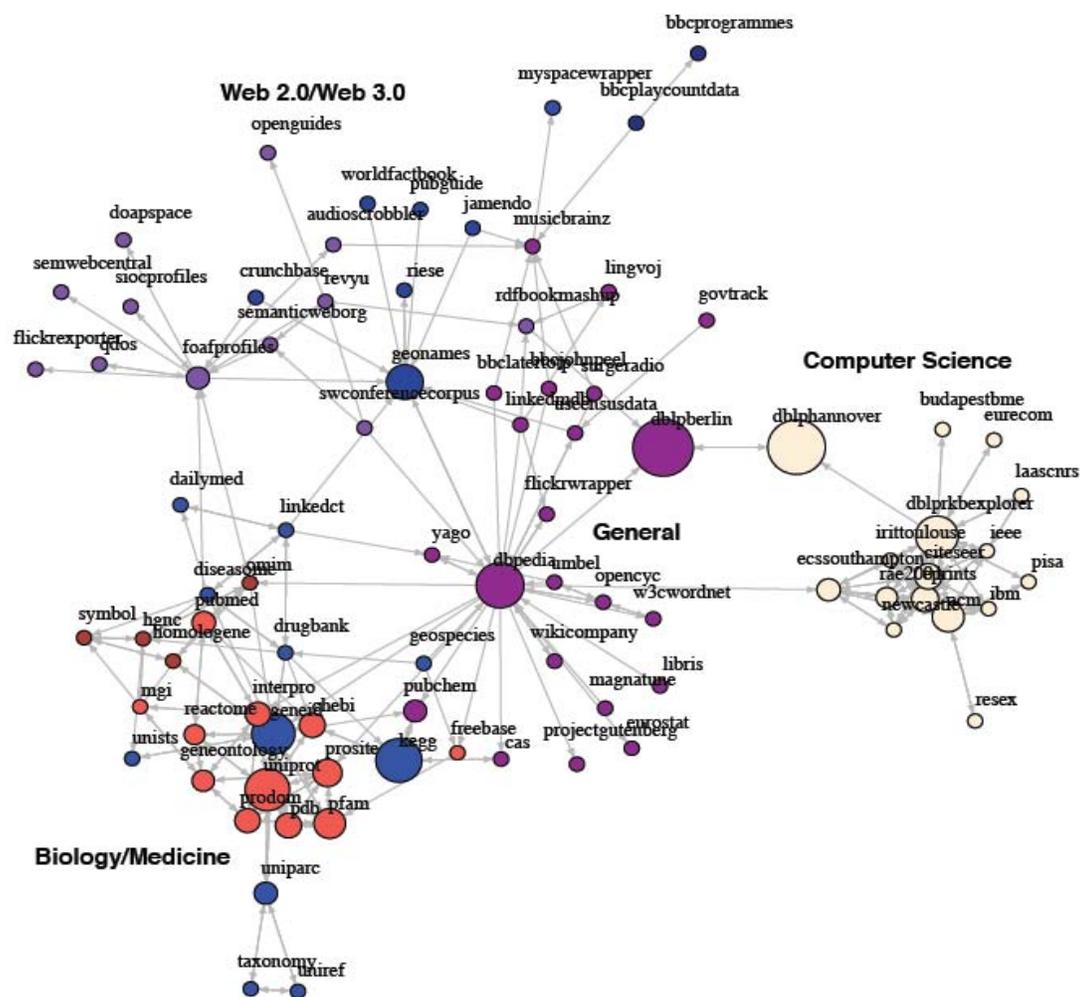


FIG. 4: A graph representation of the March 2009 Linked Data Cloud. Each vertex denotes a Linked Data data set. Each edge denotes whether one data set makes reference to another. The size of the vertices are determined by their PageRank centrality according to a $\delta = 0.85$ [22]. The vertex colors denote the structural communities as identified by the leading eigenvector community detection algorithm [18]. Finally, the Fruchterman-Reingold layout algorithm was used to visually render this representation [23].

- [Analysis of Complex Networks: From Biology to Linguistics](#). Mathematical problems such as graph theory problems are of increasing importance for the analysis of modeling data in biomedical research such as in systems biology, neuronal network modeling etc. This book follows a new approach of including graph theory from a mathematical perspective with specific applications of graph theory in biomedical and computational sciences. The book is written by renowned experts in the field and offers valuable background information for a wide audience.
- [Informatics team finds simple rules that explain universal laws of written text](#). Alessandro Flammini and Filippo Menczer (**Indiana University**), along with M. Ángeles Serrano from the **University of Barcelona**, have authored a paper entitled “[Modeling Statistical Properties of Written Text](#)” that has been published in the *PLoS One* [[abstract](#)]. The paper introduces and validates a generative model that explains from simple rules the simultaneous emergence of patterns of written text observed in many languages.
- [An Interview with Jon Kleinberg](#) (a top expert in social network analysis).
- “[The application of social network theory to animal behaviour](#),” by Mark Haley (Bath Spa University, UK), *Bioscience Horizons*, March 2009 ([full text](#)).
- [As Data Collection Grows, Privacy Erodes](#). As illustrated in Alex Rodriguez's positive steroid test result, personal information that seems anonymous is often traceable, and potentially life-changing.
- [Why normalization matters with K-Means](#). (K-means clustering in Clementine)
- An online [notebook on complex networks](#) from **Kosma Shalizi** may be a useful source.
- [Social Network Analysis Comes of Age](#), feature article at *TCMnet* by Aidan Connolly, CEO of **Idiro Technologies**. General article with a focus on customer relationships.
- [Semantic Web Promises a Smarter Electricity Grid](#). Working in the EU-funded [S-TEN project](#), a team, led by Bernhard Schowe-von der Brelie at the FGH research institute in Mannheim, Germany, developed a generic framework for novel ICT architectures and applied semantic web technologies to make networks ‘self-describing’ so that each component – be it a volt meter on a wind turbine or a thermometer on a weather station – autonomously publishes information about what it is, where it is, and what it does.
- [Tracking the Digital Traces of Social Networks](#). Noshir Contractor at Northwestern University has studied the massive online virtual world Second Life to test social theories ... Second Life has more than 15 million registered accounts and differs from other massive online multiple-player games in that there is no real goal – people create virtual avatars of themselves and then chat with other people, and buy and sell items.

III. COMPANY NEWS, IN BRIEF

- **Agilex:**

[Agilex Names New President of Intelligence, Security and Defense](#) (ISD). Agilex, which provides semantic-based text analytics and works with the ISD sector, has appointed a 20-year veteran of intelligence agency work as President of the ISD sector. John G. Edwards most recently worked with the Director of Central Intelligence and also spent close to 15 years managing technical operations and engineering support within the Central Intelligence Agency.

- **Analytix:**

[Analytix On Demand Partners with Pervasive Software to Deliver Powerful SaaS Business Intelligence to Healthcare Industry](#). ... the solution is designed to give healthcare enterprises a comprehensive, real-time view of business performance across clinical, financial and operational aspects of healthcare management. Architected for scalability, the solution is built for managing large volumes of data from multiple disparate sources.

- **Attensity:**

★ [Attensity, Empolis, Living-e Merge to Deliver Combined Customer Analytics](#). (April 20, 2009) The three companies have merged into a single company, called the [Attensity Group](#), that will offer software for analyzing customer feedback and comments across internal and external channels, including the Web. Attensity's technology will be integrated with [Empolis](#)' information management applications and [Living-e](#)'s text-analysis software for incoming customer communications, whether from e-mails, Web forms, SMS messages, letters, faxes and call logs. . The Attensity Group, which will be headquartered in Palo Alto, will employ about 300 people in offices in the US, Germany and London. *[Fern Halpern offers a commentary on whether this move makes sense, [here](#). Curt Monash's [blog](#) obviously had some detailed insider information on the deal. He reports that "Inxight managers have been brought in to run the whole thing. Specifically, **Ian Bonner** will be CEO, and **Ian Hersey** will be EVP of Products and Technology" ... and ... "Existing investors (employees aside) have largely been bought out. Most of the stock is owned by Aeris, an investment vehicle for **SAP co-founder Klaus Tschira**."]*

[enherent and Attensity Announce Partnership to Deliver Solutions Using Text Analytics Technology](#). The complementary capabilities of enherent and Attensity enable "the effective management of brand reputation..."

Attensity made the [Red Herring's "Global 100" list of top tech start-up companies](#). "Technology industry executives, investors, and observers have regarded the Red Herring 100 lists as an invaluable instrument to discover and advocate the promising startups that will lead the next wave of disruption and innovation."

- [Blue Spider Analytics:](#)

The company announced the release of RC 0.8.2.0 to beta users on March 20th (no further detail available).

- **Detica / BAE:**

[Detica and CEOP launch new collaborative partnership to fight online child abuse](#). Working with the UK's Child Exploitation and Online Protection (CEOP) Centre, Detica will support it and the Virtual Global Taskforce by providing its expertise in information intelligence, on a *pro bono* basis, to develop new ways to identify and stop online child abuse networks.

- **FMS Advanced Systems Group:**

FYI: FMS offers a side-by-side comparison of its Sentinel Visualizer vs. i2's Analyst Notebook, [here](#).

- **ISS (Intelligent Software Solutions):** *(Most news items below could not be directly hyperlinked; access them from the "news" tab at the top of their homepage. For more information on any of the products mentioned below, use the "solutions" tab accessible from the homepage.)*

Intelligent Software Solutions: Making a difference every day. This article, appearing at *Technology-Digital.com*, notes that "One of the specialties in **WebTAS** is predictive analysis. ISS is currently engaged in major enhancements to this subsystem."

FYI: ISS is developing **PANDA**, or the Predictive Analysis for Naval Deployment Activities.

Managed out of DARPA's Information Exploitation Office, the PANDA program envisions a global-scale system providing anticipatory situation awareness on 50K+ entities based on motion and emission based change detection and adaptive context modeling. While the current focus is on the maritime domain, it is anticipated that PANDA will have wide-spread application. ISS is focused on applying advanced machine learning, fuzzy computing, and probabilistic reasoning technologies to discover individual vessel SIGINT emission patterns and generate vessel forecasts and normalcy assessments from these patterns.

AKMC (Automated Knowledge Model Creation) technologies are leveraged to discover emission patterns and to model the patterns using expressive Temporal Transition Model (TTM) and Bayesian Network representations. Statistical time series analysis techniques will be applied to discover emission usage rate patterns across various temporal and spatial dimensions. Near real-time level 2+ fusion components will generate vessel forecasts and normalcy assessments using probabilistic inference and predictive analysis techniques to correlate observations of vessel activities against the learned emission behavior models. The system will facilitate automated model learning, relearning, and adaptation. These automated learning and situation assessment components are being developed using state of the art Service Oriented Architecture (SOA) and Java EE methodologies and technologies.

- **KnowledgeMiner Software:**

★ KnowledgeMiner Software in partnership with **Intel** and **Microsoft** has released a major new data mining application for the Mac called **KnowledgeMiner (yX) for Excel**. KnowledgeMiner (yX) for Excel takes advantage of multi-core processors through the use of parallelization to achieve speeds more than 600 times faster than previously. ...

The goal of using KnowledgeMiner (yX) for Excel is to obtain useful knowledge from large collections of data. The discovered knowledge can be equations describing properties of the data, frequently occurring patterns, clusterings of the information, etc. ... KnowledgeMiner has as a foundation its own unique set of inductive learning and self-organizing modeling technologies: GMDH Neural Networks, Fuzzy Rule Induction, and Analog Complexing pattern recognition for time series prediction, classification, modeling, clustering, and diagnosis of complex ill-defined systems. Multileveled self-organization and validation approaches make it not only possible to generate reliable individual and hybrid models in an objective way, but also provides an explanation component on the fly in the form of algebraic or difference equations, fuzzy rules, or a set of similar patterns. This extracted knowledge can be directly used to improve model results, to get new insights into the system, and to aid decisions.

[\(Source\)](#)

- **KXEN:** “KXEN, The Data Mining Automation Company™ delivers next-generation Customer Lifecycle Analytics to enterprises that depend on analytics as a competitive advantage.”

[New Social Network Analysis Module \[KSN\] Strengthens KXEN Automated Data Mining.](#) A new KSN social network module makes exploiting customer connections easier. “With KSN you can identify multiple networks (friends, co-workers, family...), and use that knowledge to reveal new information about your customers’ interactions. Letting you better anticipate their next move. Helping you make your next campaign even more profitable.” ([Source](#))

An interview about this development with KXEN’s VP for Strategic Business Development, describing KXEN as “the first to launch social network tools for analytics,” appears [here](#). According to the interview, the product uses graph theory, and the company considers its only competitor to be SAS.

We think our solution is unique in its ability to handle very large volumes (more than 40 m nodes and 300 m links).

- **Leximancer:**

[Leximancer Forms OEM Partnership with Polecat Ltd.](#) London-based Polecat’s MeaningMine draws strategic marketing insight from external and internal data sources, including intranets, blogs, customer feedback, audio and video, and analyst reports. The combined platform (with Leximancer’s text analytics) “will derive actionable customer insight from unstructured data and provide key insights for customer service, brand management and customer intelligence professionals. ‘To be perfectly candid, we chose Leximancer’s platform because it is years ahead of any other text analytics platform we’ve seen on the market today,’ said James Lawn, co-founder of Polecat.”

- **OmniViz:** BioWisdom Ltd is a private company headquartered in Cambridge, UK. ... In December 2006 [it] merged with OmniViz Inc, a leading supplier of data visualization software.

- **Palantir Technologies:**

★ *Palantir recently refreshed its website and added new content that may be of interest to the NGC team.* Palantir shows an [example use](#) of its software to understand Hezbollah, and offers other examples of its use, e.g., identifying hot spot populations for West Nile Virus, analysis of conflict in the Congo, and uncovering the GhostNet cyber espionage network (see [here](#)); mortgage fraud, implementing the Troubled Assets Relief Program (TARP), and investigating money laundering activities in online payments (see [here](#) and more on the TARP [here](#)); and analysis of Al Qaida foreign fighters in Iraq (see [here](#)).

[Palantir Wins Interactive Visual Analytic Environment Award from IEEE.](#) Palantir received the award for its participation in the 2008 Visual Analytic Science and Technology (VAST) competition, sponsored by the Institute of Electrical and Electronics Engineers (IEEE).

[Application Platform - SNA and Map.](#) In the 2.0 release, Palantir:

... takes the concept of Enterprise solutions to the next level with its customizable Application Platform. The Platform allows organizations to extend their experience with distinct user applications. Organizations can embed applications within Palantir, running the application as if it were a pre-existing package in the Application Platform. For an enterprise with large investments in existing software applications, Palantir’s Application Platform becomes an invaluable asset ... As with embedded applications, Palantir also allows custom ‘helpers’ to support applications within the platform. The Platform already comes pre-packaged with existing helpers like a histogram that bins objects on property types. These helpers add to the power of embedded applications and can operate across applications.

- **SAP:**

[SAP Launches Investigative Case Management Software to Help Police Solve Crimes](#). The new software package addresses a wider range of law enforcement needs across the complete investigation lifecycle – from instigation and initial investigation through secondary investigation, case finalization and review ... If evidence, a person or a crime scene is connected to multiple cases, this association will be instantly recognizable to detectives. SAP Investigative Case Management for Public Sector and complementary offerings, such as **text analytics** and business intelligence software from the SAP® BusinessObjects™ portfolio, allow a law enforcement agency to deploy SAP software as a common enterprise-wide platform for managing investigations.

- **SAS:**

[SAS Offers New Fraud Framework Featuring SAS® Social Network Analysis](#) offers detection and alert and case management to help reduce fraud loss and prosecute fraudsters. The framework includes many technologies to prevent, detect and investigate fraud including business rules, anomaly detection, predictive models and social network analysis. Designed to help detect and prevent both line-of-business fraud and cross-channel enterprise fraud, organizations can use the framework to help increase fraud detection rates, reduce false positives and streamline investigative resources. At the core of the framework is a profiling engine that scores individuals, accounts, products and networks based on rules, fraud scores and links to known fraudsters. An enterprise view of fraud exposure and risk is provided by consolidating alerts from multiple systems.

- **SPSS:**

[SPSS Rebrands Its Analytical Offerings](#). ... doing away with the Clementine product branding in favor of an umbrella label for its entire portfolio: “PASW,” short for “predictive analytics software.” ... the new versions of SPSS products focus on usability – and not just for data experts ... moving beyond the data analyst audience is “where you get the real power.” (More detail [here](#).)

[“Turning Customer Interactions into Money: Using Predictive Analytics to Achieve Stellar ROI,”](#) a white paper from SPSS, is available for download.

- **Tech-X:**

[Tech-X Corporation](#) develops scientific modeling and distributed computing software products and advanced technologies for research, engineering and education. Its products enable clients, collaborators and partners to “build a greater understanding of physical processes and increase their design and development productivity.” Their products include the OOPIC Pro for integrated simulation and visualization, and FastDL, which “helps solve the challenge of analysis and visualization of very large data sets on clusters by coordinating IDL script code through an MPI interface (mpiDL) or by scheduling autonomous tasks (TaskDL).” Last year, Tech-X was one of the recipients of a SciDAC award for visualization at the annual meeting of the U.S. DOE's Scientific Discovery through Advanced Computing (SciDAC) program. Most recently, it released GPULib v1.0. This software library executes vectorized mathematical functions on graphics processing units (GPUs) from **NVIDIA**, “bringing high-performance numerical operations to everyday desktop computers.”

- **TIBCO:**

[TIBCO Helps Government Agencies Spot Hard-to-See Fraud, Risks and Threats](#) -- TIBCO Spotfire Analytics Software Makes Hidden Relationships in Intelligence, Financial and Network Data Visible to Government Analysts. Using TIBCO Spotfire, government agencies can ...use the software's capabilities to make analysis a part of any process, and the ability to automatically broadcast updated analyses and share data insights to information consumers inside the organization

or out in the field. The latest version of the software even allows analysts reviewing complex multi-dimensional data to zoom and rotate in three dimensions, enabling them to analyze their data in a whole new way, and quickly spot new insights or hidden patterns. In addition, TIBCO has recently announced new products to combat the unique data analysis challenges of federal agencies. TIBCO Spotfire [Networks Analytics](#) (introduced in October of 2008) is a new product that aids in any type of relationship analysis, from identifying email and cell phone traffic patterns to understanding social network analysis, to analyzing counter-party or supply chain relationships. TIBCO Spotfire S+ products allow government agencies to incorporate and deploy predictive analytics on very large data sets in a wide range of applications, including drug safety and intelligence.

- **Visual Analytics, Inc. (VAI):**

[CODY Systems and Visual Analytics Partnership to Extend Data-Sharing Technologies](#) – Alliance Provides End-to-end Data-sharing Solution for the Public Safety, Law Enforcement and Federal Sectors. [CODY Systems](#)' technology provides tactical, real-time data sharing to the edge of the network, while Visual Analytics' technology compliments CODY's services via strategic, multi-tiered visual and geospatial analysis of data, allowing the information to be searchable on many levels and for a variety of purposes. With this new, end-to-end data fusion, sharing, and analysis platform, the two companies claim that they will fill a known gap in the data fusion landscape by bringing together First Preventers, tactical operatives and officers on the street, investigators, and strategic fusion center analysts on one common, seamlessly-integrated platform for the real-time fusion, sharing and analysis of data from critical, disparate data sources. Combined, these technologies will allow agencies to synch data in real-time, at the most granular level, from disparate sources (including disparate RMS systems), analyze it, and make it searchable to users of any participating agency.

IV. RESOURCES / OVERVIEWS

A. Reviews and Overviews

- [***Making Sense of Data II: A Practical Guide to Data Visualization, Advanced Data Mining Methods, and Applications***](#) (Glenn J. Myatt and Wayne P. Johnson), Wiley, 2009. 291 pages. Billed as an easy-to-use guide for non-specialists.
- ★ [**"Top Visual Search Engines: The Most Interesting Ways to Visually Explore Search Engine Results"**](#) (blog post), reviews and compares 15 search engines. Another post, ["How difficult is it to change,"](#) reports on the new, [searchme.com](#) visual search engine and discusses user resistance to changing search habits.
- ★ ***Complexity: A Guided Tour***, by Melanie Mitchell of the **Santa Fe Institute**: (Book Review in *Nature* 458, 411, March 26, 2009) Mitchell offers a valuable snapshot of the growing field of complex-systems science ...

Especially valuable is the book's exploration of recent attempts to categorize the dynamics of cellular automata - simple systems that act as models for the study of rich dynamics. Some of this work, under the name of computational mechanics and linked to the ideas of Mitchell's former colleague, the late Jim Crutchfield, probes the fundamental 'information physics' of complex systems in general. This focus of the book is commendable, as much of the literature of complex-systems research dwells on more expansive philosophical themes at the expense of the 'boring' details of specific models. Yet intense scrutiny of such models may ultimately reveal clues to solving Bedau's mystery. Mitchell touches on the many practical applications of this science, ideas put into practice by forward-looking companies such as Cisco and Capital One. The book is timely, given that many analyses of the present financial crisis have concluded that the key issue is how markets have outstripped our ability to understand them. It has become fashionable in recent years to criticize complex-systems science for generating too much hype and not offering enough practical insight. But insights into truly complex problems do not come easy. Mitchell's welcome book makes it clear that this field is making steady, if slow, progress. ([On Amazon](#))
- [***Mathematical Tools for Datamining***](#) (free March 2009 ebook), authored by Dan A. Simovici and Chabane Djeraba, is described as follows:

The maturing of the field of data mining has brought about an increased level of mathematical sophistication. Such disciplines like topology, combinatorics, partially ordered sets and their associated algebraic structures (lattices and Boolean algebras), and metric spaces are increasingly applied in data mining research. This book presents these mathematical foundations of data mining integrated with applications to provide the reader with a comprehensive reference.
- [***Successful Enterprise Search Management***](#), Stephen E. Arnold and Martin White. White and Arnold write about the most important issues in the management of enterprise search procurement, deployment, and enhancement. The study mentions dozens of vendors, includes numerous case studies, and contains illustrations of business procedures.
- [***Driving Results through Social Networks***](#), a new book by Rob Cross ([webcast](#)) addresses such questions as "What do we know about the networks of high performers? What promotes knowledge worker productivity? What would you do if you could see these networks? What would you do differently?" The book is discussed on this [blog page](#).

- **Washington University's** Center for the Humanities in Arts & Sciences has announced **Asad Ahmed**, Ph.D., assistant professor of Arabic, as a 2010 Faculty Fellow. Ahmed's book is tentatively titled "*Empire and Periphery: A Social Network Analysis of the Hijazi Elite in the Early Islamic Period.*" It aims to reconstruct the sociopolitical history of the elite families of the Hijaz, a province in the Arabian Peninsula, for the entirety of the first and part of the second Islamic dynasty (661-833). The work includes quantitative and **social network analysis** to explore the social structure and sociopolitical history of the most prominent elite families. [The book will not only provide] a detailed provincial history but also bring to light the ways in which the central authorities managed a vast empire in the early history of Islam. ([Source](#))
- **Valdis Krebs** has a number of interesting recent posts on his [TNT blog](#), e.g., "[Contagion Amongst Banks](#)," "[Madoff Feeder Funds](#)," and "[The Network Structure of Swine Flu Pandemic](#)."
- Full-text presentations from the April 2009 "**Search Engine Meeting**" are available [here](#).

B. Resources on the Web

- ★ Presentations from SBP09 (the second **International Workshop on Social Computing, Behavior Modeling, and Prediction**) are available [online](#). Among the many interesting presentations are: Terence Lyons from **AFOSR** on "[Collective Behavior and Social Cultural Modeling](#)"; Kathleen Carley and colleagues from **CMU** and **Lockheed** on "[Dynamic Networks: Rapid Assessment of Changing Scenarios](#)."
- Conferences:
 - *Two Mode Social Network Analysis*, September 30 – October 2, Amsterdam ([conference overview](#)).
 - [ICDAR 2009](#), 10th International Conference on Document Analysis and Recognition, July 26-29, 2009, Catalonia, Spain. The [AND 2009 Workshop](#) (3rd Workshop on Analytics for Noisy Unstructured Text Data) is held in conjunction with ICDAR.
 - 5th Annual [Text Analytics Summit](#), June 1-2, 2009, Boston.

V. DIGITAL FORENSICS

The Networks Grand Challenge management team asked Perspectives to take a brief, preliminary look at (open source) information on software tools for hard drive recovery and especially for analysis of recovered data. We identified many off-the-shelf programs for recovering data from a hard drive for routine hard drive problems, and this area of commercial software appears to be basically commoditized. However, the more sophisticated disk recovery and analysis software – tools and toolsets often described as “Digital Forensics” or “Computer Forensics” used for situations where data on the disk has been purposely hidden, degraded or “destroyed” by the user in an effort to evade detection or prosecution – is an active and dynamic area of development and research. Digital forensics is essentially a big data problem, where the investigator usually does not know what in the mountains of data on a hard drive (or cell phone or other device) may be important to the investigation.

Perspectives’ research on the area of digital forensics contains some background on the field, overviews on the state of forensic tools, and information on developments in this area that struck us as particularly interesting, such as the work of Simson Garfinkel of the Naval Postgraduate School in developing a corpus of hard drives for analysis and automated forensics analysis, and current uses of visualization and/or network analysis. Perspectives developed tables for reference and further exploration, the most extensive being an annotated, detailed table of forensic software toolkits from more than 50 organizations. In addition, we have provided links to evaluations of some of these toolkits, and reference tables of toolkit lists, conferences, selected publications, and digital forensics portals.

A. About the Field of Digital Forensics

Digital forensic tools are used by law enforcement and security agencies, businesses, and legal firms. Some of the top applications for these tools involve investigation of:

- Criminal / terrorism activities
- Insider threats
- Fraud
- Employee / Employer issues: e.g., wrongful dismissal, harassment, improper access by terminated employees, computer misuse in the workplace

Golden G. Richard, III, a professor at University of New Orleans who also runs a consulting firm focused on digital forensics, recently (2008) commented on the future evolution of digital forensics tools. His predictions and comments are useful as an introduction to thinking in the field. They are reviewed at some length below and may be worthy of reading in entirety ([source](#)). Specific predictions for next-generation product capabilities include:

- Better file carving
- Better auditing (i.e., “digital evidence bags”)
- Going faster
- Live forensics (used on “running” computers or other devices)

Richard focuses on where forensically interesting data may reside (see right), and discusses the current challenges in forensics.

One of those key challenges is the time it takes to do the initial imaging of the hard drive and the audit log.

The time required – hours or even days – is problematic for investigators because during the imaging process, no further investigative steps can be taken. Investigators literally have nothing to do, according to Richard.

On “Traditional” Computer Systems

- Undeleted files, expect some names to be incorrect
- Deleted files
- Windows registry
 - e.g., USB device histories
 - e.g., recently accessed files, URLs
- Print spool files
- Hibernation files
- Temp files (all those .TMP files!)
- Slack space
- Swap files
- Browser caches
- Alternate partitions
- On a variety of removable media (floppies, ZIP, Jazz, tapes, ...)

Richard speaks at length about this problem in his presentation, and summarizes his comments about audit log generation as follows:

Too Slow: Symptoms

- Machines tied up for days doing preprocessing
- Painful to “think outside the box” (i.e., outside the index) during investigation
- Getting an answer to even a simple question
 - “Does this credit card number appear?”
 - “Did Joe send an email to Cassandra?”
- ...takes a long time
- If a regular expression search takes hours to complete, what do you do in the meantime?

Too Slow: Potential Solutions

- Reduce amount of evidence to process
- Smarter analysis techniques
- Automation of boring manual operations
 - evidence correlation
 - (photographic) image verification/sorting/clustering
 - automatic relevant keyword searches
 - automated password cracking
- Increase speed of processing

Evidence Reduction

- Better imaging (evidence copying)
 - Move intelligence closer to evidence acquisition
 - Copy only relevant evidence, leave the rest
 - Cryptographic hashing during acquisition
 - “Quick” matches during acquisition:
 - Bloom filters
 - Hashes of strategic file blocks
 - Similarity hashing
 - “Automatic” identification of interesting evidence
 - Better Triage
 - Live forensics
- } Do a better job “on the spot”...

B. Digital Forensic Tools and Toolsets

While sources differ on which digital forensics tools / toolsets are most widely used, some of the top commercialized tools come from **Guidance Software** (EnCase), **AccessData** (FTK), **Paraben** (P2 suites), **E Fence** (Helix), and **X-Ways Software** (WinHex). There are also some commonly-used open source tools as well, most notably the software developed by **Brian Carrier** – The Sleuth Kit, Autopsy Forensic Browser and mac-robber.

There are **approximately 150 different automated tools routinely used by law enforcement organizations** to assist in the investigation of crimes involving computers. These tools are used to create critical evidence used in criminal cases, yet there are no standards or recognized tests by which to judge the validity of results produced by these tools. – NIST

Ian Charters, a Senior Manager at Deloitte & Touche, LLP (and apparently former DIA and CIA – [source](#)) wrote an article on “[The Evolution of Digital Forensics: Civilizing the Cyber Frontier](#)” (based on a presentation given at a key NIST conference in October of 2008). This article provides an interesting perspective on the development of forensic tools. Charters notes that an important factor in the development of these tools is that they be forensically correct, in the sense that the collected data must be:

- Collected and maintained in accordance with a defined procedure
- Verifiable as authentic
- Verified as relevant
- Collected in a reliable manner
- Preserved (original data) to the extent possible

Charters’ overview of some of the top tools is interesting, and we quote him at length:

Once the tools and techniques of digital forensics were accepted as legitimate and effective, the market began to drive digital forensics from a point-based solution into an enterprise-based solution. This imposed some significant technical demands. It also created significant market opportunities. In general, the Enterprise phase of digital forensics can be characterized by: real-time collection; field collection tools tailored to the need of the collectors; [and] forensics as a service.

... The following are some of the landmark software packages that have helped define Enterprise Phase software-based solutions:

- Access Data – Known File Filter (KFF)
- National Drug Intelligence Center’s – Hashkeeper
- Guidance Software – EnCase Enterprise
- Access Data – Enterprise
- Brian Carrier’s – The Sleuth Kit (TSK)
- Mediant – Intelligent Response
- Clearwell – E-discovery Platform
- [Athena Archiver](#) [*which now makes a next generation email archiving and storage management system; “search millions of emails in seconds”*]
- LogLogic

Known File Filter and Hashkeeper are key tools that have established the foundation for the automation and standards-based forensic data collection and analysis. They incorporate a scientific, reproducible, and disciplined approach allowing the forensic professional to ignore large

volumes of known system and application files so that attention can be focused on potentially significant data and in the case of Hashkeeper, the development of know bad file signatures.

... **Guidance Software** really pioneered the move to Enterprise forensic solutions. Their EnCase Enterprise makes the collection of data in real-time of the Enterprise infrastructure possible and manageable.

Mediant's Intelligent Response may very well be establishing the next stage of evolution in digital forensics. Their Intelligent Response product is a rules-based appliance that takes automated forensic data collection to the next level. The evolution of this product and the response by the rest of the industry will be worth watching.

Clearwell's E-discovery Platform is another interesting offering that may be a trendsetter. The concept behind E-discovery Platform is that for some Enterprises, the proper management, archiving, and subsequent forensic data collection of e-mail is so burdensome, difficult, and disruptive that out-sourcing the entire function is attractive. With new Federal guidelines for e-discovery on the books, Clearwell may have gotten themselves in front of [an] increasingly vexing problem with a well thought-out and timed solution.

While not specifically a digital forensic tool, the company **LogLogic** is providing a ground breaking integrated log collection and analysis capability that is long overdue.

The Future: I believe **the future of digital forensics will be aimed at greater automation and interoperability**. In general it will be characterized by:

- Proactive collection and analysis: Appliance-based tools
- Government/Commercial partnerships (e.g., [InfraGard](#))
- Standards-based software architectures
- Standards-based reporting: Forensics version of SCAP [[Security Content Automation Protocol](#)]

Automation is a key element to the future of digital forensics. Automation will allow proactive collection and detection. This will translate into reduced cost for detection and mitigation. This will be a welcome capability considering the increasing concern over data leakage and the new rules for e-discovery. Automation will also reduce program costs. The creation of proactive forensic appliances is a clear move in this direction. If we can characterize forensic events effectively and embed those profiles in an automated appliance, it would allow not only for the real-time collection of forensic data, but also provide real-time prevention of related "events".

Other interesting tools are under development by **Basis Technology** (which received funding from In-Q-Tel), **Perlusto** (ILook), **Nuix** (US GALE, which is offered through **Digital Intelligence Inc.**), and **Backbone Security** and **ADF Solutions** (steganalysis tools).

Many companies offer digital forensics services to users and have developed proprietary approaches (i.e., often a combination of in-house developed toolsets used with some of the better commercially available tools). These firms tend to put out very little information about their technology; it is unclear if this is because they believe it to be a competitive differentiator, or if it is because they know that it is not a differentiator (and thus not worth talking about). One example is [Forensicon](#), which is focused on intellectual property matters such as trade secret misappropriation. Another company of interest here is [Detica Forensics](#). Detica offers specialized services that: "... routinely use a range of tools from industry standard forensic packages such as **EnCase©** and **FTK©** [applied] to **our own highly tuned data analytics software to highlight the most relevant items of data**. Data analytics will encompass all forms of storage types ... and may even necessitate the recovery of data hidden through the use of steganography or encryption. ... one of the most powerful forms of analytics is to show the inter-relationships between what may appear ... to be disparate items of data, but **when linked show the digital equivalent of social networks**." ([Source](#))

Perspectives has developed a compendium of information on toolkits (primarily commercial) shown in the final section of the report.

- **Table of digital forensic software tools** – an annotated table organized by company ([Table 1](#)). If interesting or sophisticated analytical / graphical tools were identified as part of the toolsets, we pointed this out.
- **Toolkit Evaluations (Table 2)**: For several years, *SC Magazine* has been providing detailed evaluations of forensic toolkits. The cost of each product at the time it was evaluated is listed by the magazine, and we have included those cost figures in a summary table. NIST has also performed evaluations of forensic tools according to an exhaustive set of criteria – some 46 evaluations to date – and a link to these assessments at the National Institute of Justice is included.
- **Other Resources**: For the readers' reference, we compiled a **list of online resources on forensic tools** that list products for various aspects of the forensics problem ([Table 3](#)).

C. Other Interesting Work: Automation, Datasets, Visualization, Analytics

There are a number of sources indicating that automation, visualization, analytics, and other techniques are slowly converging with the field of digital forensics. Here are some highlights:

- ★ [Simson L. Garfinkel](#) is an Associate Professor at the **Naval Postgraduate School** and an associate of the School of Engineering and Applied Sciences at **Harvard University**. His research interests dovetail nicely with NGC's focus (e.g., computer forensics, the emerging field of usability and security, personal information management, privacy, information policy and terrorism - see [wiki](#) on this interesting professional). Garfinkel works with "automated computer forensics." In 2007, he had a project with the goal of developing test corpora of 3,000 hard drives by the end of 2008, according to an online [project abstract](#). Using the hard drive corpus, Garfinkel and his team aimed to develop technologies and techniques to enable automated forensic processes. These include:

The [Advanced Forensic Format \(AFF\)](#), a new file format for archiving disk images and associated metadata. [See [AFF wiki](#) for tools developed as part of this format by Garfinkel and available as an open source - see official AFF [website](#).]

Cross-Drive Analysis, a technique for automatically identifying and correlating pseudo-unique information across forensically relevant media. We have developed algorithms based on CDA that can automatically determine the owner of a hard drive, and automatically identify multiple drives in a collection that were used by the same organization.

Carving Fragmented Files with Object Validation, a technique for reassembling files that have been split into multiple pieces in a disk image without reference to the file system metadata.

As of August of 2008, Garfinkel's research group had actually gathered some 2000 hard drives from around the world. This "Real Data Corpus", he says, is a one-of-a-kind scientific resource for:

- Developing and validating forensic and data recovery tools.
- Training students in forensics and data recovery.
- Developing and validating document translation software.
- Exploring and characterizing real-world computing practices, configuration choices, and option settings.
- Studying the storage allocation strategies of file systems under real-world conditions. ([Source](#))

Garfinkel reports:

... we are developing a new technique for mapping social networks among individuals whose data is on captured hard drives. These approaches could be used, for example, to allow the rapid and automated analysis of disk drives seized during the course of a police investigation or obtained as part of military operations.

Tools developed for automated and “bulk forensics” are described [here](#). Some of Garfinkel’s recent papers:

- Garfinkel, S.; Farrell, P; Rousev, V; and Dinolt, G., “[Bringing Science to Digital Forensics with Standardized Forensic Corpora](#)” (accepted to DFRWS 2009)
- Rousev, V., and Garfinkel, S., “[File Fragment Classification – The Case for Specialized Approaches](#),” [Systematic Approaches to Digital Forensics Engineering 2009](#), Oakland, CA (acceptance rate 30%)
- Garfinkel, S., “[Automating Disk Forensic Processing with SleuthKit, XML and Python](#),” [Systematic Approaches to Digital Forensics Engineering 2009](#), Oakland, CA.
- Garfinkel, S. and Cox, D., “[Finding and Archiving the Internet Footprint](#),” invited paper, presented at the First Digital Lives Research Conference: Personal Digital Archives for the 21st Century, London, England, February 9-11,2009.
- Garfinkel, S., “[Providing Cryptographic Security and Evidentiary Chain-of-Custody with the Advanced Forensic Format, Library, and Tools](#),” [The International Journal of Digital Crime and Forensics](#), 1(1):1-28, January-March 2009. [[Full text](#)]

- [Robert F. Erbacher](#) of **Utah State** is one academic working with the issue of visualization and forensics. He outlines the basic issues in one of his class presentations (see [here](#), in particular for sample block diagram and treemap views). He was the Chair of the 2009 [IEEE SADFE Workshop](#) (Systematic Approaches to Digital Forensic Engineering). Erbacher is working with **PNNL** and **AFRL** on projects related to digital forensics (see [here](#)) and apparently has a **DARPA** seedling grant. Erbacher is one author of the paper, “[Improving the computer forensic analysis process through visualization](#)” in the February 2006 issue of *Communications of the ACM* (49(2):71-75) in a [special issue](#) on “next generation cyber forensics.”

There are several widely used computer forensics toolkits, namely: ILook [Perlustro], Encase [Guidance Software], and Sleuthkit [freeware by Brian Carrier]. **The extent of visualization incorporated into these tools is essentially explorer type interfaces.** Thus, our work on visualization goes far beyond what these tools provide. However, these tools are very good at providing scripting and other low level analysis tools. Our goal is not to compete with these tools but rather, in the future, to integrate their results into the visualizations and thus make a more usable and effective set of capabilities.

–“[Foundations for Visual Forensic Analysis](#),”
2006

Sheldon Teerlink, Robert F. Erbacher
[\[full text link\]](#)

- The **VizSec 2008** conference papers ([table of contents](#)) include one by Moses Schwartz and L. M. Liebrock of NM Tech (Schwartz is also affiliated with SNL) on “[A Term Distribution Visualization Approach to Digital Forensic String Search](#).”

Digital forensic string search is vital to the forensic discovery process, but there has been little research on improving tools or methods for this task. We propose the use of term distribution visualizations to aid digital forensic string search tasks. Our visualization model enables an analyst to quickly identify relevant sections of a text and provides brushing and drilling-down capabilities to support analysis of large datasets. Initial user study results suggest that the visualizations are useful for information retrieval tasks, but further studies must be performed to obtain statistically significant results and to determine specific utility in digital forensic investigations.

- Mississippi State researchers, [T. J. Jankun-Kelly](#), [J. Edward Swan](#), and [David Dampier](#) (Director of the National Forensics Training Center), along with **Jeffrey Carver** of the University of Alabama, are working on Computer Forensic visualizations in a [NSF-funded project](#) that began in 2006. Carver describes the project as follows:

This goal of this project is to build a visualization that will enable computer forensics practitioners to more effectively gather and document electronic evidence, such as may be found on a computer hard drive. The first task in this process is to perform a task analysis of the way in which computer forensics experts currently perform their work. The results of this analysis will provide the requirements for the visualization tool. After the visualization tool is built, its effectiveness will be empirically validated against the original requirements. My expertise in empirical software engineering and experimental design is aiding in the initial task analysis and in the final validation.

Research Findings: We have focused on understanding how users with various levels of expertise use existing computer forensics tools like "Forensics Toolkit" or "Autopsy Forensic Browser" to gather evidence of computer crimes. Our initial study focused on the domain of Webmail and internet history in fraud cases. Using results from on-site surveys of computer forensics officers and user studies with students, an initial visualization prototype for assisting in email analysis was developed. The study canvassed Mississippi and involved student interaction with law enforcement officers in order to inform our study. Using those results, we developed some quantifiable measures that were used to pilot some of the visualization's features with students. These results form the basis of the visualization we're currently refining; the plans are then to validate the visualization with further testing and evaluation by law enforcement officials. ([Source](#))

- Greg Conti and Erik Dean** (West Point Military Academy) demonstrated at a recent Black Hat Conference how visual computer forensic methods can reduce the time it takes to review files. (More detail is available at the [article](#) at *Law.com* – see also quote at right.)

Computer forensics is a slow process. Examiners typically embark on a tedious file review process to determine each file's relevance to a particular case. This can quickly add hours and extra costs to computer forensics. ([Source](#))

- Clearwell Systems** is presenting on "The Analysis-Powered Forensic Investigation: How Analytics Can Find the Smoking Gun" at the upcoming Techno Security 2009 Conference.

... Advancements in information analytics are helping investigators quickly and accurately identify a case's relevant evidence and potential custodians. Attend this hands-on lab to learn how investigators are using new tools to rapidly identify and produce the native emails and digital files needed for corporate investigations, litigation requirements and regulatory inquiries. ... ([Source](#))
- FYI, as a matter of interest: The **Electronic Discovery Reference Model (EDRM)** is an industry group attempting to develop and establish practical guidelines and standards for electronic discovery (more information at the [website](#)).

D. Compendium of Information on Digital Forensics and Commercial Tools

This section contains information on a variety of aspects of computer forensics organized into four tables. [Table 1](#) is an annotated list of tools and toolsets (mostly but not exclusively commercial). The field has a great many suppliers, and we believe that skimming the entries will provide a useful snapshot of the state of the industry. More sophisticated analytical capabilities or visualization features are identified where these were obvious. Organizations are listed alphabetically and links to detailed product descriptions are provided. This table was developed, in part, from sources noted in Table 3.

[Table 2](#) contains links to detailed reviews of forensic tools from *SC Magazine* (including product costs as published by the magazine) and from NIST's Computer Forensics Tools Verification project.

The remaining tables are for the readers' reference for further exploration, generated as part of this research. [Table 3](#) contains a list of forensic tool lists. [Table 4](#) contains a list of some important conferences, publications, and portals / information sites related to digital forensics.

NGC team members may be interested in the upcoming **Techno Forensics** conference sponsored by NIST to be held in Gaithersburg, MD, on October 26 – 28, 2009 ([here](#)). While the agenda has not yet been posted for this conference, last year's conference (see agenda and full text presentations [here](#)) appeared to be quite interesting.

Table 1: List of Selected Digital Forensic Tools / Toolkits, by Company

Tool Company / Organization	Explanation
AccessData	FTK (Forensic Toolkit), Registry Viewer, Password Recovery Toolkit. One of the top companies in this area. "At the heart of the total law enforcement solution is our award-winning Forensic Tool Kit® (FTK®), which is well recognized throughout the community as best in class for digital forensic investigations. FTK, however, is just the beginning of our offerings. For organizations that need to do more than computer forensics, we also offer our fully integrated Mobile Phone Examiner for the acquisition and analysis of cell phone data ... "
ADF Solutions	This company has recently partnered with Backbone Security (see entry below), "the industry leader in advanced digital steganalysis tools." They are offering a steganography SearchPak® for use with ADF's digital forensic triage tools. These tools are currently deployed at law enforcement agencies, government organizations, and commercial corporations worldwide. ADF created the steganography SearchPak® from hash values extracted from the Steganography Application Fingerprint Database (SAFDB) created and maintained in Backbone's Steganography Analysis and Research Center (SARC). SAFDB is the world's largest commercially available hash set exclusive to steganography applications. Digital forensic examiners around the world are using hash values from SAFDB to detect the presence of steganography applications on seized media. First responders and lab examiners can use the SearchPak® with ADF's Triage-ID™ and Triage-Lab™ to quickly detect the file artifacts of 675 steganography applications. Triage-ID™ is a field tool for first responders to perform automated on-site analysis of suspect computers, with a bootable CD-ROM and USB drive. Triage-Lab™ is a Windows-based tool that performs automated analysis of drive images, network drives, stand alone suspect computers, DVDs, CDs, and other digital media.

Tool Company / Organization	Explanation
ASR Data	Offers a variety of tools for forensic analysis, including SMART and SMART for Linux , and several others for visualization, including Grok-NTFS (NTFS file system analysis tool with data visualization) and Grok-LNK (Grok-LNK is an NTFS file system analysis tool – point Grok-LNK to the root of a mounted NTFS file system and Grok-LNK will examine all the link files and provide the user with a tab delimited list of the links, the targets and their attributes – by combining Grok-LNK with SmartMount , Grok-LNK can examine ExpertWitness / Encase format E01 files, raw dd images, Virtual Disk images, FTK images and many others).
Backbone Security	"The industry leader in advanced digital steganalysis tools." (See also Steganalysis Wiki). The Steganography Analysis and Research Center (SARC) is a Center of Excellence within Backbone Security focused exclusively on steganography research and the development of advanced steganalysis products and services. The SARC has developed state-of-the-art steganography detection and extraction capabilities that address the needs of digital investigation specialists and information technology security personnel in law enforcement, government, military, intelligence, and the private sector. Provides a national repository of steganography application hash values, or fingerprints, and developing the most advanced steganalysis tools, techniques, and procedures to find and extract hidden information.
Basis Technology (received In-Q-Tel funding)	"The next generation of digital forensics tools." Capture – the Media Exploitation Kit enables experts and non-experts alike to capture data off hard disks, while also documenting the integrity and source of the data. Analysis- Odyssey Digital Forensics Keyword Searching System's smart search crosses language and file format "barriers." Analysts need not know all the languages of the data to perform searches... Portability – the Advanced Forensic Format® (AFF) for storing captured data is open and extensible to make that data available for analysis by any tool the investigator chooses. Products are integrated with the company's Rosette entity extractor, linguistics platform, and other tools.
BKForensics	Cell Phone Analyzer
Blackbag Technologies	BlackBag Macintosh Forensic Software . Company provides Mac-based data forensic and eDiscovery solutions to law enforcement and private sector clients. Downloadable demo here .
Brian Carrier's Forensic Tools (open source)	The Sleuth Kit (TSK) is a collection of command line digital investigation tools. The tools run on Linux, OS X, FreeBSD, OpenBSD, and Solaris and can analyze FAT, NTFS, UFS, EXT2FS, and EXT3FS. The Autopsy Forensic Browser is an HTML-based graphical interface for the command line tools in The Sleuth Kit. This makes it much easier and faster to investigate a system. mac-robber is a tool that will collect temporal data from mounted file systems. The data can be used to make a timeline of file activity on the system using tools from The Sleuth Kit. (Wiki) See also Brian Carrier's TASK data forensic program in RPM format here .

Tool Company / Organization	Explanation
Clearwell	<p>E-Discovery Platform (this platform is billed as a tool for forensics investigators) ... intelligently processes case datasets in hours instead of days – even with cases containing millions of documents ...</p> <p>Analysis functions: After processing, the Clearwell E-Discovery Platform applies its proprietary, patent-pending analysis algorithms known as Dynamic Content Analysis™ to create the Clearwell Master Index™. This index is much more sophisticated than a full text index and powers e-discovery analysis capabilities not available with any other e-discovery solution on the market today. Key analysis features of the Clearwell E-Discovery Platform include:</p> <ul style="list-style-type: none"> • Discussion threads: ... patent-pending algorithms dynamically link together all related messages into chronological threads ... • Topic clustering: ... patent-pending linguistic algorithms automatically organize documents into specific topics • People analytics: ... analyzes individual and group-to-group communications within your company, or to your customers, suppliers, and partners ... can easily access a list of top custodians for a search or monitor communications between regulated and non-regulated divisions... derives group information automatically from Active Directory, and users also have the flexibility to create their own custom groups. • File analytics: ... identifies duplicate files that may be attached to multiple emails or saved on a user's hard drive. File analytics allow investigators to easily determine everyone who possesses or has sent or received a file of interest and allow reviewers to review a file once instead of multiple times. • Term analytics: ... natural language algorithms analyze noun phrases to help users uncover secret project names and code words that may be relevant to a case or investigation.
CompuForensics (computer forensics training institute)	Seized Computer Analysis Boot (SCAB) software - Restricted [NO OPEN INFORMATION AT SITE].
Computer Forensic Training Center Online	FSUITE forensic software, sold by Key Computer, Inc.
ComputerCOP	Offers a suite of ComputerCOP Professional Field Forensic Tools.
Crucial Security Inc.	Crucial Vision: a digital forensics bulk-processing preview and holistic examination tool. Crucial Vision speeds time to analyze large volumes of data by providing examiners a holistic view across all of their data , resulting in prioritized work flow. Crucial Vision offers 3 to 5 times faster searching and processing performance than traditional products. (Appears to be a free download).
Cyber Security Technologies	Mac Marshal™ automates the analysis of Mac OS X file system images. The OnLine Digital Forensic Suite™ (OnLineDFS for short) is a software product for performing forensic-quality investigations of live computers in networked environments. P2P Marshal™ is a new computer forensic tool which automatically detects, extracts and analyzes P2P evidence on hard drive images .
Data Lifter	DataLifter - Forensicware Solutions™ forensic support tools (also offered by Digital Intelligence)

Tool Company / Organization	Explanation
DF Labs in Italy (PTK, an open source tool)	From the IRItaly project , " PTK is an alternative advanced interface for the suite TSK (The Sleuth Kit – see entry for Brian Carrier above). PTK was developed from scratch and providing the functions already present in Autopsy Forensic Browser. It implements numerous new features essential during forensic activity. PTK is not just a new graphic and highly professional interface based on Ajax technology, but offers a great deal of features like analysis, search and management of complex cases of digital investigation. The core component of the software is made up of an efficient Indexing Engine performing different preliminary analysis operations during importing of every evidence. PTK allows the management of different cases and different levels of multi-users. It is possible to allow more than one investigators to work at the same case at the same time. All the reports generated by an investigator are saved in a reserved section of the Database. PTK is a Web-Based application and builds its indexing archive inside a Database MySQL, using thus the construction LAMP (Linux-Apache-MySql-PHP)." [source]
DIBS USA Software	DIBS Analyzer & DIBS Mycroft V3
Digital Detective	NetAnalysis Forensic Internet History software, & other utilities
Digital Forensic Solutions, Inc.	Company formed by Golden Richard, who invented Scalpel (freely supplied). Scalpel is a fast file carver that reads a database of header and footer definitions and extracts matching files from a set of image files or raw device files. Scalpel is filesystem-independent and will carve files from FATx, NTFS, ext2/3, or raw partitions. It is useful for both digital forensics investigation and file recovery. Scalpel resulted from a complete rewrite of foremost 0.69, a popular open source file carver, to enhance performance and decrease memory usage.
Digital Intelligence, Inc.	DriveSpy, PDWipe, PDBlock, Image & Part. The company also offers Nuix US GALE product, which automates analysis and includes visualization tools , and Data Lifter tools (See also Data Lifter and Nuix entries in this table)
DMZ Services	F.I.R.E. (DMZS-Forensic and Incident Response Environment) is a portable bootable CDROM based distribution capable of providing an immediate environment for performing forensics analysis, incident response, data recovery, virus scanning and pen-testing.
dtSearch	dtSearch: The dtSearch product line can instantly search terabytes of text across a desktop, network, Internet or Intranet site. Used by a number of others to create forensic solutions, including AccessData , e.g., see here . Features here , include natural language searching.
Dutch National Police Agency (OFCA, an open source tool)	The Open Computer Forensics Architecture (OCFA) is a modular computer forensics framework built by the Dutch National Police Agency [KLPD/Dutch]. The main goal is to <u>automate the digital forensic process</u> to speed up the investigation and give tactical investigators direct access to the seized data through an easy to use search and browse interface.
e-Fense	Helix3 Pro forensics product.
Evidence Talks	"One of the most highly regarded digital forensic consultancies in the UK." Offers: <ul style="list-style-type: none"> • Remote Forensics, a secure architecture upon which full digital forensics, imaging and analysis, can be conducted, without pulling any of the data back across the network. • SPEKTOR Forensic Intelligence – a forensic triage solution ... Using a combination of unique portable hardware and software, SPEKTOR delivers the ability to interrogate computers and removable storage devices for the presence of relevant material quickly, accurately and within a forensically and legally acceptable framework. • Email Forensics: LoPe quickly extracts email messages and attachments from multiple Outlook PST files.

Tool Company / Organization	Explanation
Forensic Innovations, Inc.	Forensic Innovations, Inc. is now able to detect TrueCrypt, and other encryption data hiding techniques. The newly released File Investigator version 2.23 continues in their plan of supporting more file types (currently supporting 3,312) than any other product in the industry. Along with the 141 added file types, the accuracy of 78 existing file types was improved. Add to that the ability to identify more encryption file formats (including the elusive TrueCrypt), and this version is truly a leap forward in advancing the state of the art in the analysis and filtering of Electronically Stored Information (ESI).
Forensic Server Project (Harlan Carvey)	The Forensic Server Project (FSP) is a proof-of-concept tool for retrieving volatile (and some non-volatile) data from potentially compromised systems. The FSP consists of several Perl scripts and third-party utilities. The server component of the FSP is run on an investigator or administrator's system, and handles all data storage and activity logging. The client components (i.e., FRU.pl and supporting Perl scripts and tools) of the FSP are burned to a CD, and run from the CD drive of the potentially compromised system. Data is copied to the server component via TCP/IP. It should be noted that while the FSP is used for incident response and forensic audits of Windows systems, it is also an open source project. The server component is written in Perl, and can be run from other systems that support Perl (with minor modifications). Client components can be written in Perl, or any other scripting language. Apparently this comes with a book published in 2007 by Harlan Carvey, see here .
Foundstone	Free Forensics Tools – Pasco, Galleta, Rifiuti, NTLast, Forensic Toolkit, ShoWin, BinText, PatchIt, and Vision
Funduc Software	"Several useful utilities" for computer forensics (shareware and freeware).
Guidance Software	Bills itself as the world leader in forensics investigation . Offers the Encase family of software products. For example, " EnCase® Enterprise is the only truly secure and scalable network-enabled forensic investigation platform in existence. It allows you to peer deeply into your network to understand at a forensic level what is stored or occurring on your machines, and if necessary, remediate without disrupting operations. EnCase® Enterprise can be customized to meet your organization's unique needs, or you can deploy EnCase® Information Assurance, which is preconfigured to complete your defense-in-depth strategy." See new product announcement, " Advanced Algorithms Enlisted To Fight Cyberwars ."
Hot Pepper Technology	Email Detective for investigation of AOL email (also offered by Digital Intelligence). Also offers eChat Locker for investigators.
KrollOnTrack	Offers OnTrack software for data recovery and search. "... recognized leader of data recovery, advanced search, and legal technology solutions, Kroll Ontrack offers intuitive, scalable, efficient and cost effective software products to address your most critical and time sensitive data needs." Ontrack® Engenium® is "an award-winning and market leader in Conceptual Information Access Technology ..." Semetric® – Conceptual Search . Autometric® – Automatic Clustering. Cometric™ - Search Results Clustering. CometricSP™ – Search Results Clustering for SharePoint .
LastBit Software	Primary offerings are password recovery tools, e.g., Secret Explorer allows the ability to "locate hidden information in any Windows-based system."
LC Technology International	Forensic Utility Suite™ is the first-of-its-kind allowing forensic and computer crime professionals to effectively recover data on all Microsoft® operating systems. The suite is a compilation of FILERECOVERY® for Windows® , FILERECOVERY® Professional and PHOTORECOVERY® for Digital Media .

Tool Company / Organization	Explanation
LogLogic	Offers an appliance-based and managed services-based platform for collecting, storing, reporting and alerting on 100 percent of user and system activity data from virtually any log data source.
Mares and Company (the Norcross Group)	Maresware Computer Forensic software & Hash Sets offered by Dan Mares. This software is apparently commonly used in the community. Flagship product is Maresware: The Suite – forensic capabilities include discovery of "hidden" files (such as NTFS Alternate Data Streams) for incident response purposes; evaluation of timelines; powerful key word searching and comparing ; files verification; keyboard locking; forensic diskette imaging; file reformatting; completely documenting the examiner's steps and procedures; disk wiping to overwrite a hard drive to DOD standards; and adding Bates-style identifying numbers to files for evidentiary use.
Mandiant	Intelligent Response (v1.3) : helps you find evil by automating and solving computer security incidents. It is designed to be a collaborative work environment – allowing multiple users to identify, collect, analyze and report on data simultaneously. According to Ian Charters, the product is a rules-based appliance that takes automated forensic data collection to the next level . The evolution of this product and the response by the rest of the industry will be worth watching.” (Additional news articles here and here .)
National Drug Intelligence Center - DOMEX (DOJ)	The National Drug Intelligence Center (NDIC) Document and Media Exploitation (DOMEX) Branch has developed a uniquely efficient approach that allows analysts to quickly organize and assimilate significant amounts of seized documentary and electronic evidence. <ul style="list-style-type: none"> • NDIC created Real-time Analytical Intelligence Database (RAID) to manage large quantities of data gathered during DOMEX operations. RAID is a relational database used to record key pieces of information and to <u>quickly identify links</u> among people, places, businesses, financial accounts, telephone numbers, and other investigative information examined by our analysts. It can be used to analyze any type of information from any kind of investigation or as a case management tool. • The HashKeeper expedites the analysis of electronic media. HashKeeper is a software application that quickly eliminates known operating system files and focuses on electronic files created by the user/subject of the investigation.
New Technologies, Inc. (NTI)	Offers Stealth™ Suite and Computer Incident Response Suite (“DOD tested and certified”).
NTSecurity.nu Toolbox	Freeware utilities. “These tools are intended for white hat use only. Use them for security testing, for hacking in a lab environment, and so on. I [Arne Vidstrom] certainly do not condone any illegal or immoral use, and in several cases I have (on purpose) made them easier to detect and/or harder to hide.”
Nuix: FBI	An Australian-owned and developed forensic software tool, in wide use by government and private sector. Powerful search & visualization features , email analysis, and inter-operability with key forensic tools like Encase (see Guidance Software entry, above). US GALE simplifies, automates and dramatically speeds up the ability to find crucial evidence and graphically demonstrate what really happened. The result is faster output, greater productivity, reduced cost and increased quality control. The solution comes complete with Nuix’s revolutionary Universal Indexing™ Engine and Evidence Aware Searching™ abilities. Single cases can scale to sizes in excess of 10TB and include any combination of the above formats. ... Unique visualization shows clearly how information was sent into, through, and out of an organization, including full email addresses, times and dates of emails. Reviews the social and business network of people, then drills down into specific relationships to better understand them.

Tool Company / Organization	Explanation
Oxygen Software	Oxygen Forensic Suite 2 is a mobile forensic software that goes beyond standard logical analysis of cell phones, smartphones and PDAs. Using advanced proprietary protocols permits Oxygen Forensic Suite 2 to extract much more data than usually extracted by logical forensic tools, especially for smartphones.
Paraben Software	Offers comprehensive forensics software solutions for handheld, hard drive and enterprise forensics. Maker of P2 tools, notably the P2 Commander , a comprehensive hard drive forensic suite that uses plug-in architecture to create specialized engines that focus on such things as E-mail, Network E-mail, Chat Logs, File Sorting, and more all while increasing the amount of data that can be processed and utilizing resources through multi-threading and task scheduling. The company offers a comparison chart of its product versus Guidance's EnCase and AccessData's FTK .
Perlustro	ILook Forensic Analysis Tool (apparently this was recently commercialized and originally developed by Elliot Spencer and U.S. Dept of Treasury, Internal Revenue Service – Criminal Investigation). Offers a tool called IXIMAGER (v3) that "can image more systems than any product on the market. It doesn't require a box full of hardware writeblocks and it doesn't write to any device unless given the authority to do so by the operator. Not even cold booted linux software raids can knock a dent in it. It has the capability to boot 15 year old Apple PPC machines to the 8 cores of G5's running Intel chips. It avoids the normal necessity to do physical removals from laptops and there is no necessity to do target firewire mode on Apples." The Analysis pack offers visualizations , see here , including e-mail link analysis.
Loglogic, Inc.	LogLogic® Open Log Management "leads the industry" with an appliance-based and managed services-based platform for collecting, storing, reporting and alerting on 100 percent of user and system activity data from virtually any log data source.
Sanderson Forensics	A computer investigative services firm that sells some of its forensics software, e.g., RevEng (a fully featured hex viewer), SkypeAlyzer ; PmExplorer (mobile phones), VidReport (video), and KaZAlyser .
SubRosaSoft.com / MacForensicsLab	Provides MacForensicsLab and MacLockPick . MacForensicsLab is a complete suite of forensics and analysis tools in a cohesive package combining the power of many individual functions into one application in order to provide a single solution for law enforcement professionals.
Microsoft Sysinternals	Several useful, free utilities from Microsoft.
TCT (The Coroner's Toolkit)	Freeware / shareware: Post mortem analysis of a UNIX system (a collection of programs by Dan Farmer and Wietse Venema).
Tech Assist, Inc.	ByteBack , BringBack & Detective PC Data Sleuth – on a website called "Tools that Work" by Tech Assist, Inc.
Technology Pathways	ProDiscover family of computer security tools. For example, see the ProDiscover DFT forensic tool for law enforcement.
WetStone Technologies	One of the top companies in digital forensics. Gargoyle Investigator products, Stego Suite , and more.
X-Ways Software Technology AG	Offers X-Ways software line, including WinHex , Davory , Evidor & X-Ways Forensics . "Davory undeletes files and recovers files from logically corrupted or formatted drives. Incorporates some of the data recovery techniques from WinHex and concentrates on ease of use. Offers two separate, fully automated data recovery mechanisms to maximize your chances of success. "

Table 2: Forensic Tool Reviews

Source	Products Rated / Product Cost (<i>Company name if not specified in product</i>)
Group test: Digital forensics (SC Magazine) May 8, 2009	<p> Cyber Security Technologies OnLine Digital Forensic Suite \$9,000 single user (\$3,000 for law enforcement); \$25,000 for multi-user HBGary Responder Field Edition; \$979 LogLogic MX 2010; \$35,000 LogRhythm LR-1000-XM; starts at \$20,000 Mandiant Intelligent Response v1.2; \$86,500 Paraben Device Seizure; \$1,095 Prism Microsystems EventTracker; typical 50 server setup: \$19,995 Splunk - indexes & searches all information in data center environment, giving easier access to logging utilities for ... network forensics; enterprise license starting at \$7,500 Technology Pathways ProDiscover IR v5.5; \$12,995 </p>
Media forensics (SC Magazine) May 1, 2008	<p> Helix 1.9 (<i>e-Fence, free</i>) Forensic Toolkit v2.0 (<i>Access Data</i>); \$2,995 Gargoyle Investigator Enterprise Module (<i>WetStone Technologies</i>); Starts at \$1,995 Paraben Device Seizure; \$895 STRSRCH and URL_SRCH (<i>Mares and Company</i>); \$95 plus a one-time dongle fee of \$25 Technology Pathways ProDiscover Forensics 4.9; \$12,995 </p>
Network forensics (SC Magazine) May 1, 2008	<p> LogLogic LX 2010 v4.2; \$68,995 LogRhythm v4.0; \$20,000 Niksun NetDetector; starts at \$10,000 Technology Pathways ProDiscover Incident Response; \$12,995 WetStone Technologies LiveDiscover Forensic Edition; \$1,995 </p>
Forensic tools 2007 (SC Magazine) April 1, 2007	<p> Guidance Software EnCase Forensic v. 6; \$3,000 for a corporate license, plus support LogRhythm LR1000 v. 3.5 - log analysis appliance; \$30,000, plus support Paraben Device Seizure v. 1.1; \$895, plus support Paraben P2 Enterprise Shuttle; \$6,995, plus support Technology Pathways ProDiscover IR v 4.9; \$7,995, plus support WetStone Technologies Gargoyle Investigator; \$995, plus support WetStone Technologies LiveWire Investigator v. 3.1.1C; \$8,995, plus support </p>
Forensic tools 2006 (SC Magazine) July 11, 2006	<p> Coroner's Toolkit (<i>Open source/Dan Farmer and Wietse Venema</i>); Free EnCase Forensic (<i>Guidance Software</i>); \$3,000 Forensic ToolKit (<i>Access Data</i>); \$1,095 i2 Analyst's Notebook; \$3,652, inc. one year's support LogLogic LX 2000 - log analysis tool; \$49,999 Mandiant First Response; Free NetWitness - a network traffic security analyzer (<i>Mantech Int'l</i>); \$30,000 ProDiscover Incident Response; \$7,995 Sleuth Kit & Autopsy Browser (<i>Open source/Brian Carrier</i>); Free </p>
NIST Computer Forensics Tool Testing Site	<p> The Computer Forensics Tools Verification project provides a measure of assurance that the tools used in the investigations of computer-related crimes produce valid results. It also supports other projects in the National Institute of Justice's overall computer forensics research program, such as the National Software Reference Library (NSRL). </p> <p> List of NIST forensic tool evaluations at NIJ: as of the publication of this report, some 46 evaluations are posted on this site. </p>

Table 3: Forensic Tool Lists for Reference

Tool List	Explanation
e-evidence info	Links to well-known digital forensics toolsets and other useful tools .
Forensic Tool References	Extensive list of individual tools from the Electronic Crime Partnership (organized by tool name) – April '08, annotated as to function.
TUCOFS	The Ultimate Collection of Forensics Software : A complete resource for cyber law enforcement technologies. Provided by Cyber Enforcement Resources, Inc. , a non-profit organization, and the provider of <i>Cybersnitch</i> , a free service for the public and law enforcement, which they claim is the world's leading computer crime reporting system.
Computer Forensics WIKI	Extensive list of tools for forensics investigators, includes commercial and shareware, freeware. Includes a section on data recovery, as well as: Disk Imaging, Disk Analysis, Live CDs, Metadata Extraction, File Analysis, Network Forensics, Anti-Forensics, and other tools.
Computer Forensics Toolkits	Annotated list of software suites from the <i>forensics.nl</i> site. This list has short explanations of each tool. The list includes a mix of commercialized products and products developed by individuals, some of which may be freeware or shareware.
Mares list of Forensic Hardware and Software Resources	Lists software alpha, as well as by these functions: Disk manipulation, formatting, partitioning; Data recovery specialists; Disk/text/hex editing; General purpose software for a variety of uses; Graphic viewers & processing; Hashing – CRC, SHA calculations; Linux(*UX tools); Windows 9X administrative tools.
Computer Forensics Products	From Timberline Technologies, list organized alphabetically ... “provided as a free service to those seeking commercial information security products or shareware tools.”
Computer Forensics	Private security page of Alexander Geschonneck.
Open Source Digital Forensics Tools	List of open source tools (from <i>opensourceforensics.org</i>) that are deemed useful during a digital investigation as well as having source code that is "easily accessible". Windows- and Unix-based tools are listed in the following categories: Bootable Environments (software that you can use to boot a suspect system into a trusted state), Data Acquisition / IR Tools (used to collect data from a suspect system); Media Management Analysis Tools (used to examine the data structures that organize media, such as partition tables and disk labels); File System Analysis Tools ; Application Analysis Tools (used to analyze the file content.); and Network Analysis Tools (used to analyze network packets and traffic. This does not include logs from network devices).
Efficient Forensic Tools for Handheld Devices	2008 paper by authors from Virginia State and Texas A&M. “In this paper, we present an overview of forensic tools and discuss the challenges involved in the design of forensic tools with the steps needed to develop better toolkits in the digital forensic world.” The paper covers tools for PDAs, mobile phones, cameras, music players, and GPS devices.
Steganalysis Tools	List of tools from <i>StegoArchive.com</i> .

Table 4: Conferences, Publications, and Information Sites

Conferences	
Techno Forensics conference 2008 (NIST) – many presentations available online (here).	
Techno Forensics & Digital Investigations conference 2009 (NIST, October 26-28, 2009) (here).	
DFWRS – Digital Forensic Research Workshop (2009 program)	
IWDW 2009: The 8th International Workshop on Digital Watermarking (IWDW09) is a premium forum for researchers and practitioners working on novel research, development, and applications of digital watermarking, steganography, steganalysis, and forensics techniques for multimedia data (2009 conference topics).	
Computer Forensics Show (trade show) August 3-5, 2009	
SANS WhatWorks Summit in Forensics and Incident Response, July 6-14, 2009	
ICDF2C 2009 – International Conference on Digital Forensics & Cyber Crime, Sept. 30 – Oct. 2, 2009	
List of upcoming events from <i>Digital Forensic Investigator News</i> .	
Publications	
Conference Presentation (slides and video)	" Current and Next Generation Digital Forensics ," 2008, Golden G. RICHARD (University of New Orleans, Certified Forensics Investigator, also co-founder of Digital Forensics Solutions) (video and slides) – "state of the art in digital forensics is still pretty low tech."
Conference Presentation	" The Latest in Write Blockers & Drive Erasers " by Greg Dominguez, Vice President, Forensic Computers, Inc. (Techno Forensics 2008 conference presentation)
Article	" The Truth is in There: Sleuthing for Data with Digital Forensics " (2007 article).
Article	" Antiforensics: When Tools Enable the Masses " (2007 article).
Article	" The Evolution of Computer Forensics " by Ian Charters (2009 article).
Paper	" Applying Topic Modeling to Forensic Data " (2008 paper).
Reference Book	<i>NIJ Electronic Crime Scene Investigation Guide for First Responders</i> (2008 book).
Reference Book	<i>Computer and Intrusion Forensics</i> (2003 book).
Reference Book	<i>Advances in Computer Forensics II</i> (2006 book).
Report	"CSI Computer Crime and Security Survey, 2008 " (2008 report).
Information Sites of Interest	
Computer Forensic Wiki – has list of tools, including for steganalysis.	
Digital Forensics Investigator – news	
Champlain College Center for Digital Investigation ; Gary C. Kessler is the expert here.	
Institute of Computer Forensic Professionals	
Computer Crime Research Center	
e-evidence info (a personal portal with a lot of good info)	
Computer Forensics World	
Computer Forensics – including Cybercrime and Steganography Resources	
GCK's Cybercrime and Cyberforensics links	
Digital Forensics Links (personal portal)	
Digital Forensics Computing	