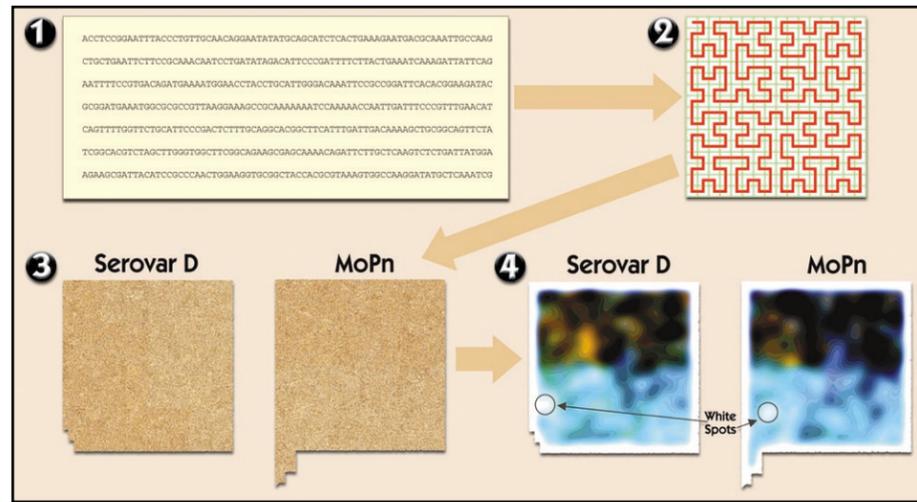


## Simplifying Scientific Datasets

Scientific datasets can be extremely large, which makes analyzing, computing, and learning from the data extremely time consuming and resource intensive. PNNL has created techniques that reduce the burden of working with generated data, such as climate models, and with observed data, such as that gathered by scientific instruments. We capture the essence of the large dataset in a representation that is a fraction of the size. These smaller, visual representations can be analyzed and compared with other datasets more quickly and easily.

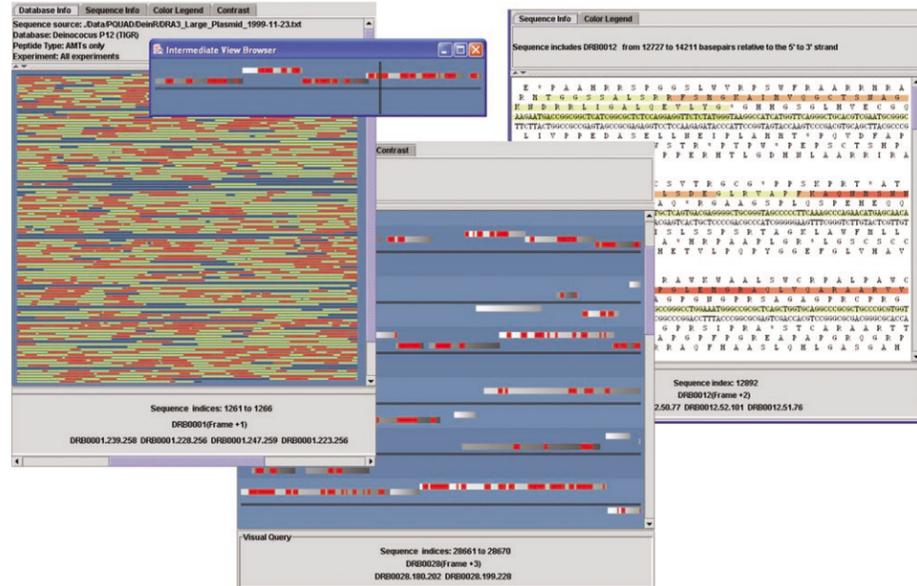


Traditionally, DNA data is displayed as strings of letters. By assigning color-coded pixels to represent the letters and arranging them along curves, images of multiple data strings can be created. Applying image processing techniques filters noise, allowing quick visual comparisons of data such as these two strains of the same bacteria.

## PQuad Simplifies Complex Proteins

New high-throughput proteomic techniques generate data faster than biologists can analyze it. Hidden within this massive and complex data are answers about how cells function. The data afford an opportunity to take a global or systems approach, studying whole proteomes. However, existing tools for studying a few proteins at a time are not suitable for global analyses.

We developed PQuad for the visual analysis of proteins and peptides identified from high-throughput data on biological samples. PQuad depicts the peptides in the context of their source protein and DNA, integrating proteomic and genomic information. A wrapped-line metaphor is applied across key resolutions of the data, from a compressed view of an entire chromosome to the actual nucleotide sequence. PQuad provides a difference visualization for comparing peptides from samples prepared under different conditions.



PQuad, short for Peptide Permutation and Protein Prediction, provides a wealth of visualizations for making sense of proteomic and genomic information.

## About PNNL

From the scientific research and development of cutting-edge concepts to engineering those concepts into visual analytic tools, PNNL is ready to meet your information analysis needs. PNNL delivers breakthrough science and technology in the areas of national security, environment, energy, health, and fundamental sciences. Battelle has operated the Laboratory for the U.S. Department of Energy since 1965.

### For commercial information, contact:

Dennis McQuerry  
Email: mcq@pnl.gov  
Phone: (509) 375-2953  
P.O. Box 999 MS: K7-28  
Richland, WA 99352

### PNNL's Information Visualization Website:

[www.pnl.gov/infoviz](http://www.pnl.gov/infoviz)

PNNL-SA-37167 September 2004

### For technical information, contact:

Kris Cook  
Email: kris.cook@pnl.gov  
Phone: (509) 375-6347  
P.O. Box 999 MS: K7-22  
Richland, WA 99352

Science.  
Technology.  
Innovation.

Pacific Northwest National Laboratory's interactive visual analytics provide new ways to ask questions and get answers, an essential element of new discoveries in:

- Technology Scanning
- Intellectual Property Management
- Information Assurance
- Scientific Research
- Legal and Regulatory Analysis
- Acquisition Analysis
- National Security and Law Enforcement
- Medical and Pharmaceutical Research
- Cyber Security
- Web Space Analysis

**Pacific Northwest National Laboratory**

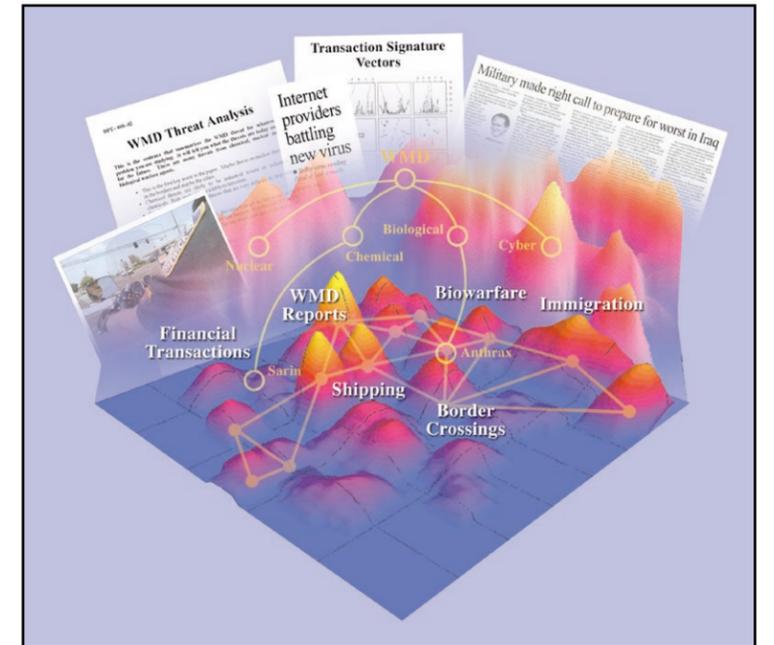
Operated by Battelle for the U.S. Department of Energy



## Interactive Visual Analytics

### A Look at Innovative Solutions at PNNL

So much data, so little meaningful information. Whether gathering market research, assessing terrorist threats, or determining how to treat a medical condition, the ability to uncover relationships, trends, and themes hidden within data can lead to new knowledge. Pacific Northwest National Laboratory's visual analytics research and technologies give people insight—the ability to see something different in the data they already have.



PNNL has created advanced visually based analytical tools that represent patterns, associations, and other connections in large, complex data collections. These tools bring analysts new insights into hidden and unexpected relationships among documents, people, concepts, and events.

### More Than Just a Pretty Picture

People have the natural ability to rapidly comprehend visual images. To make the most of this ability, PNNL's researchers develop creative and powerful approaches to present diverse data types in compelling visual displays. Statistics and complex mathematical algorithms behind the scenes transform abstract information into insight-triggering images. Furthermore, our tools allow people to interact with the visualizations and explore the information they reveal—without having to understand the sophisticated underlying mathematics.

### Working Solutions

With broad experience in research, engineering, and product development, PNNL's efforts in visual information analysis range from conducting basic research to developing innovative tools. Our interdisciplinary teams apply their diverse capabilities in software engineering, computer science, cognitive psychology, mathematics, and human-computer interactions to solve problems for industry and government clients. We look forward to further customizing these tools, as well as exploring new approaches, to meet your specific analytical needs.

## Discovering the Unexpected

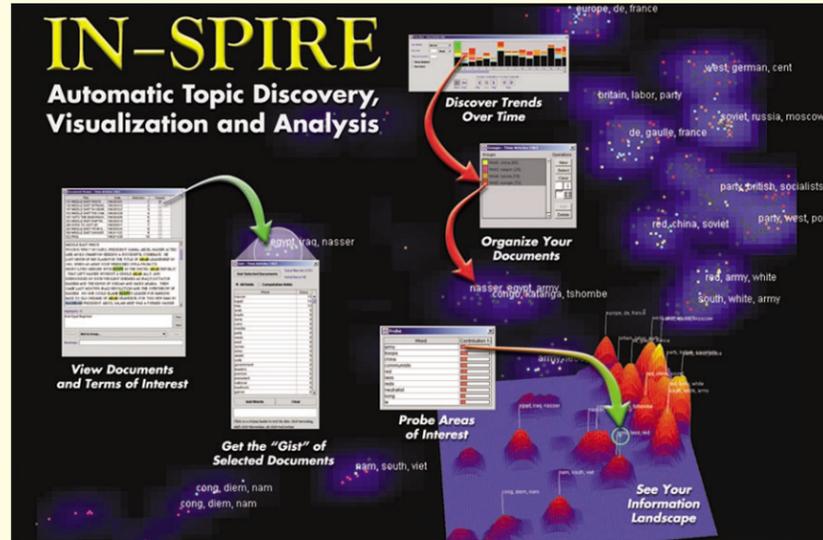
Our research helps people identify trends, temporal and cognitive relationships, and key themes within large collections of data. These collections might include a large number of documents, information on the web, images, video, or audio data.

### IN-SPIRE™ Brings Insight from Text

By clustering similar documents together, IN-SPIRE quickly and automatically conveys the gist of large sets of unformatted text documents, such as marketing and patent literature, web data, accident reports, newswire feeds, and message traffic. This Windows-based software unveils common themes and reveals hidden relationships within the collection. IN-SPIRE's powerful combination of conventional query tools and visualization techniques not only finds documents of known interest, it also uncovers their relationship to other documents in the collection.

Building upon PNNL's research for the U.S. Government, IN-SPIRE allows analysts to spend more time exploring the information they find most relevant and less time sifting through masses of irrelevant documents. IN-SPIRE supports analysts by

- Easily switching between high-level abstraction and original text to explore details
- Seeing new concepts develop over time using our trend analysis tool
- Supporting exploration and analysis of non-English datasets
- Integrating newly harvested documents and removing "aging" documents



The IN-SPIRE discovery tool integrates information visualization with interaction and query capabilities. In the Galaxy visualization, dots represent documents and cluster around center points that represent central topics or themes, like constellations in the sky. The ThemeView visualization provides an even faster way to get a visual overview of a collection of data. Users see a relief map where the highest peaks represent the most prevalent topics in the collection.

### Mozart—A Genius at Assessing Your Website

Because websites could give adversaries strategic information, government agencies and private companies are combing through their sites in search of sensitive, proprietary, or classified information. With Mozart, an Internet assessment tool being developed at PNNL, a thorough analysis can be done in hours, rather than the days, weeks, or months that manual searches would take.

Mozart archives and analyzes an organization's entire website based on search terms provided by the user and built-in search libraries. Then, Mozart compiles data on links to the web pages and determines if the site could expose sensitive information. Finally, Mozart creates a prioritized list of web pages requiring further examination. This information is also presented in a 3-D graphic representation.

The Mozart software tool assesses website information to identify and prioritize web pages that have the greatest potential to contain information that should not be published on the open Internet.



### Seeing Concepts with Knowledge Signatures

There are a lot of ways to express the concept of mortality, from "mass casualties" to "kicking the bucket." Analysts examining emails, documents, and other text know this, so they study the meaning or semantics of the text. The problem, of course, is that there is often more data than analysts can handle. This is where PNNL's knowledge signatures come in.

### Exploring Transaction Data

By combining the art and science of transactional analysis—analyzing how information, money, or other items pass from point A to point B—PNNL has found additional ways to help analysts. Transactions can be examined individually or aggregated to reveal broader patterns. They can be combined to describe, compare, and search alternative scenarios. They can also be statistically analyzed to identify hidden implicit relationships.

### Fusing Data Types

While PNNL's information analytics can provide significant insight, even more discoveries are possible when multiple data types are synthesized and analyzed together. We are developing approaches to data fusion that reveal a broader scope of relationships than could be identified when analyzing data of a single type. For example, data from multiple sensors and instruments, test results, and doctors' reports could be integrated to provide a better picture of a person's medical condition than any one of those pieces could on its own.



### Starlight Sheds Light on Data

The Starlight software analyzes large databases and displays information in ways that help users recognize subtle relationships. For example, it could be used to identify correlations in records about chemical spills by giving users the ability to sort by cause, then again by company responsible, age of equipment, or geographic location. Maps and photographs can be integrated into the display to assist in recognizing connections among the multiple variables. Potential applications for this tool include law enforcement, patent analysis, and medical and legal research.

Starlight reveals the relationships among items in a large data collection, enabling exciting and powerful new forms of information access, exploitation, and control. Starlight couples advanced information modeling and management with a visualization-oriented interface.

Knowledge signatures identify the semantic characteristics of large data sets. They also identify the absence of relationships where they were expected. They create a mapping from the information entity into a defined semantic space and provide new opportunities for visualization, classification, search and retrieval, and knowledge discovery.

We used our data exploration tools to identify computer attack instances with no prior knowledge of attack signatures, using anomaly and example-based analysis approaches.

