



# Canopy

## Interactive Multimedia Analytics

Canopy is a visual analytic software suite, developed by researchers at the Pacific Northwest National Laboratory (PNNL), to support intelligence analysts' ability to efficiently analyze massive amounts of multimedia data. Canopy incorporates cutting-edge extraction techniques, state-of-the-art content analysis algorithms and interactive information visualizations so that analysts can comprehend and articulate the big picture. Canopy helps to reduce analysts' workload and ultimately the effort of identifying critical intelligence for decision makers.

### CHALLENGE

Analysts who work with collections of multimedia understand how difficult it is to see relationships across diverse sets of mixed media, especially at today's rate of information creation and dissemination. For example, according to YouTube statistics, 72 hours of video are uploaded every minute. Flickr has 1.54 million images uploaded every day. There are 55.7 million WordPress sites in the world, contributing a total of 2.5 billion pages a month. Each day the amount of information available to consumers increases and all of it contains multimedia content. To some, this may seem like nirvana, but for those who must address policy issues or handle complex emerging events, analysis tasks quickly become overwhelming. Analysts who need to make sense of multimedia collections need tools that will help them discover associations among all types of data from multiple sources.

The ability to comprehend the big picture in multimedia information necessitates constructing and representing its context, content, and relationships. To address this challenge, we have



# Canopy

The Canopy suite is a multimedia analytics software platform that combines cutting-edge extraction techniques, state-of-the-art content analysis algorithms and interactive information visualizations to understand, explore and analyze large and diverse sets of image, video, and text.

created a visual analytic tool suite that incorporates state-of-the-art multimedia characterization, retrieval, filtering, and classification technology to provide end-users with a holistic approach to discovery and detection of relationships across the multimedia landscape. This suite provides insight into the relationships in the multimedia collection, minimizing the user's workload and ultimately the effort of identifying the critical intelligence for decision makers. This approach addresses the critical need for a next-generation content analysis system for modern and emerging media.



Multiple visualizations in Canopy provide insight into multimedia collections.

## SOLUTION

Canopy is an analytic software suite of interactive visualizations that supports overview, analysis, and discovery in mixed media data. Canopy normalizes and decomposes mixed media content into the base types of text, image, and video. Canopy applies type-specific analysis while maintaining context to the original document. A series of algorithms for analyzing text, images, and video are applied to identify features within the base types. From these features, trained classifiers help populate categories for the purpose of filtering and sorting. Canopy links the analysis of multiple base types together when the types were originally extracted from the same document so analysts can understand the inter-document and cross document relationships. Canopy has been developed as an enterprise system supporting scalable and federated data management and indexing systems, while enabling third parties to plug in services that will enhance the analytic platform.

The Canopy suite was developed with user-centered design practices that support the analytic and decision-making process through innovative user interfaces and interactive visualizations. One of the strengths of the software is its ability to provide a high-level understanding and summarization of the data through visual metaphors. This understanding is further enhanced by allowing the user to pivot, arrange, and relate the data by various attributes and metadata, thereby revealing relationships that previously were not possible to detect. These

visualization and interaction techniques allow users to gain insight and understanding of large, diverse sets of multimedia.

## IMPACT

The Canopy suite is positioned to fundamentally change the landscape for how multimedia analysis is performed. The combination of its enterprise platform, its powerful extraction and analytic algorithms, and its user interface and interactive visualizations provide a broad picture of multimedia information and its relationships. This allows the user to gain the insight needed for productive and efficient decision making.

## ABOUT PNNL

Interdisciplinary teams at PNNL address many of America's most pressing issues in energy, the environment, and national security through advances in basic and applied science. Founded in 1965, PNNL employs 4,400 staff and has an annual budget of nearly \$1 billion. It is managed by Battelle for the U.S. Department of Energy's Office of Science.

### For more information, contact:

#### Deborah Payne

Project Manager, Data Sciences & Analytics  
National Security Directorate  
debbie.payne@pnnl.gov  
(509) 375-2904

#### Dave Thurman

Manager, Computing Programs  
National Security Directorate  
dave@pnnl.gov  
(206) 528-3221

Visit us online at [vis.pnnl.gov](http://vis.pnnl.gov)



Proudly Operated by **Battelle** Since 1965