Sunfall: A Collaborative Visual Analytics System for Astrophysics

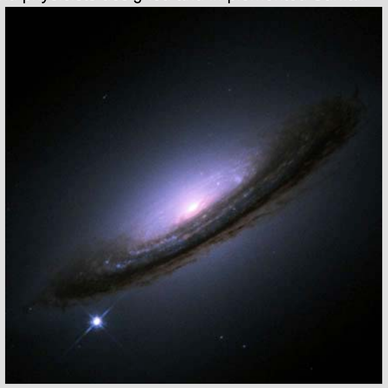
Cecilia R. Aragon, Stephen J. Bailey, Sarah Poon, Karl J. Runge, Rollin C. Thomas

aragon@hpcrd.lbl.gov, {sjbailey,sspoon,kjrunge,rcthomas}@lbl.gov

Lawrence Berkeley National Laboratory, One Cyclotron Road MS 50F-1650, Berkeley, CA 94720

ABSTRACT

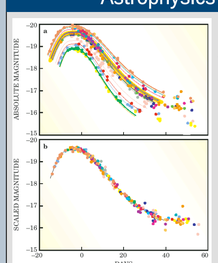
Sunfall is a production analytics system that accelerates science discovery by utilizing novel interactive visualization and analysis techniques to facilitate insight into complex, noisy, high-dimensional, high-volume, time-critical scientific data. A multi-disciplinary team of computer scientists and physicists designed and implemented Sunfall for the Nearby Supernova Factory, an international astrophysics experiment and the largest data volume supernova search currently in operation. The system combines novel image processing algorithms, statistical analysis, and machine learning with highly interactive visual interfaces to enable collaborative, user-driven scientific exploration of supernova image and spectral data.



When supernova 1994D in galaxy NGC 4526 exploded, it shone almost as brightly as the entire galaxy.

Sunfall is currently (Oct 2007) in operation at the Nearby Supernova Factory; it is the first visual analytics system in production use at a major astrophysics project. Project scientists estimate that Sunfall has reduced the data collection, processing, and management effort from 6 FTE to .75 FTE overall, freeing scientists to focus on science discovery. Additionally, by making previously unworkable scientific tasks possible, Sunfall has enabled scientists to attain new insights into their scientific data.

Astrophysics Grand Challenge: Dark Energy

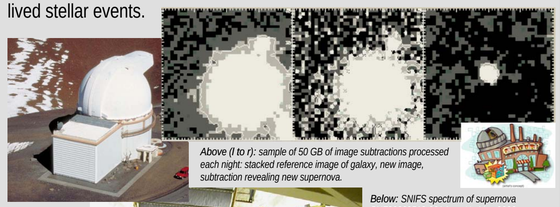


Calibrated type Ia SNe are standard candles; peak brightness is correlated with light curve shape.

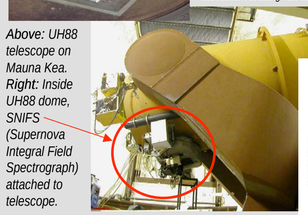
One of the grand challenges in astrophysics today is the effort to comprehend the mysterious "dark energy," which accounts for three-quarters of the matter/energy budget of the universe. The existence of dark energy may well require the development of new theories of physics and cosmology. Dark energy acts to accelerate the expansion of the universe (as opposed to gravity, which acts to decelerate the expansion). Our current understanding of dark energy comes primarily from the study of supernovae.

The Nearby Supernova Factory

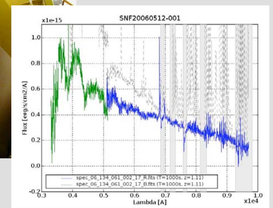
The Nearby Supernova Factory (SNfactory) is an international astrophysics experiment designed to discover and measure Type Ia supernovae in greater number and detail than has ever been done before. These supernovae are stellar explosions that have a consistent maximum brightness, allowing them to be used as "standard candles" to measure distances to other galaxies and to trace the rate of expansion of the universe and how dark energy affects the structure of the cosmos. The SNfactory receives 50-80 GB of image data per night, which must be processed and examined by teams of domain experts within 12-24 hours to obtain maximum scientific benefit from the study of these rare and short-lived stellar events.



Above (l to r): sample of 50 GB of image subtractions processed each night: stacked reference image of galaxy, new image, subtraction revealing new supernova.



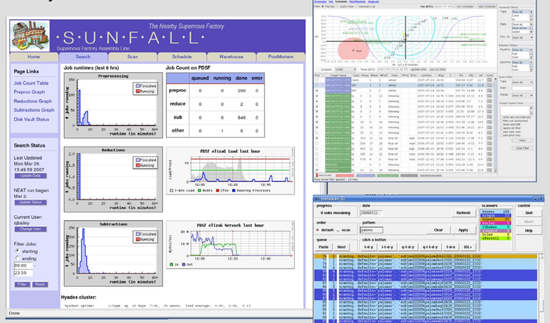
Above: UH88 telescope on Mauna Kea. Right: Inside UH88 dome, SNIFS (Supernova Integral Field Spectrograph) attached to telescope.



Below: SNIFS spectrum of supernova SNF20060512-001 found in May 2006

Sunfall

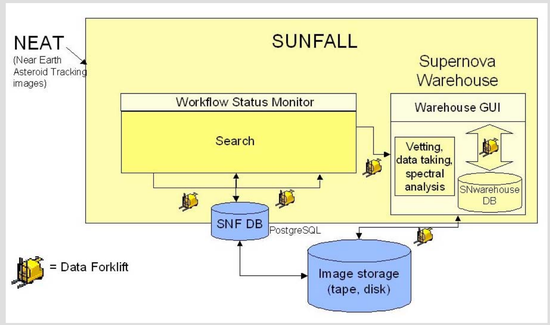
Sunfall (SuperNova Factory AssembLy Line), a collaborative visual analytics system for the Nearby Supernova Factory, has been in production use for over a year.



Above: Sunfall Status Monitor; upper right: SNwarehouse overview; lower right: metascan, a search tool.

Sunfall Architecture

Sunfall contains four major components: Search, Workflow Status Monitor, Data Forklift, and Supernova Warehouse.

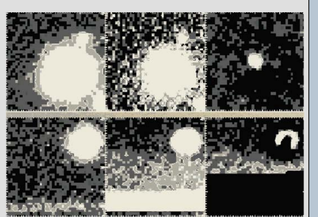


Sunfall architecture diagram, depicting the four components (Search, Workflow Status Monitor, Data Forklift, and Supernova Warehouse) and data flow between the components.

Sunfall incorporates sophisticated astrophysics image processing algorithms, machine learning capabilities including boosted trees and support vector machines, and astronomical data analysis with a usable, highly interactive visual interface designed to facilitate collaborative decision making. An interdisciplinary group of physicists, astronomers, and computer scientists (with specialties in machine learning, visualization, and user interface design) were involved in all aspects of Sunfall design and implementation.

Sunfall Component: Search

Search incorporates SNfactory legacy software for starfield image processing and subtraction, and includes machine learning algorithms and novel Fourier contour descriptor algorithms to reduce the number of false positive supernova candidates.



Sunfall Component: Workflow Status Monitor

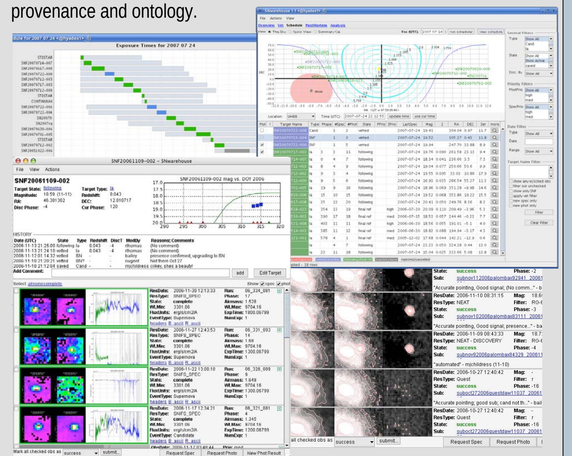
The Workflow Status Monitor is a web-based monitor to facilitate collaboration and improve project scientists' situational awareness of the data flow by displaying all relevant workflow (search pipeline) status data on a single site. Scientific process automation systems such as this have demonstrated benefits in managing data flow and visualization in complex scientific projects. (See image at left.)

Sunfall Component: Data Forklift

The Data Forklift is a middleware mechanism consisting of a coordinator and a suite of services to automate astronomical data transfers in a secure, reliable, extensible, and fault-tolerant manner. The Data Forklift supports "different place, asynchronous" collaborative scientific work by facilitating data and information transfer amongst a geographically separated team. Due to the time-critical nature of the data collection (telescopes must be operated at night and are located in different time zones), tasks must take place at distinct, specified times.

Sunfall Component: Supernova Warehouse

The most visible component of Sunfall is the Supernova Warehouse (SNwarehouse), a comprehensive supernova data management, workflow visualization, and collaborative scientific analysis tool. The SNwarehouse contains a PostgreSQL database, middleware consisting of Forklift mechanisms, and a graphical user interface (GUI) implemented in Java. SNwarehouse supports collaborative remote asynchronous work in several different ways. SNwarehouse centralizes data from multiple sources (including telescopes, spectrographs, star catalogs and web-based asteroid databases) and supports task-oriented workflow while maintaining data provenance and ontology.



Clockwise from upper left: 1. Visual scheduling tool; 2. SNwarehouse overview; 3. Details view with photometric images; 4. Details view with lightcurve and spectral data, including data provenance and scientists' annotations.

Contact information

Cecilia R. Aragon, Ph.D.

Staff Scientist, Computational Research Division

Lawrence Berkeley National Lab
One Cyclotron Rd., MS 50F-1650
Berkeley, CA 94720 USA

Tel: 510-486-4106
Email: aragon@hpcrd.lbl.gov
http://vis.lbl.gov/~aragon