VisTrails

Erik Anderson, Steven Callahan, Juliana Freire*, Emanuele Santos, Carlos Scheidegger, Claudio Silva*, Nathan Smith, Huy Vo *PI

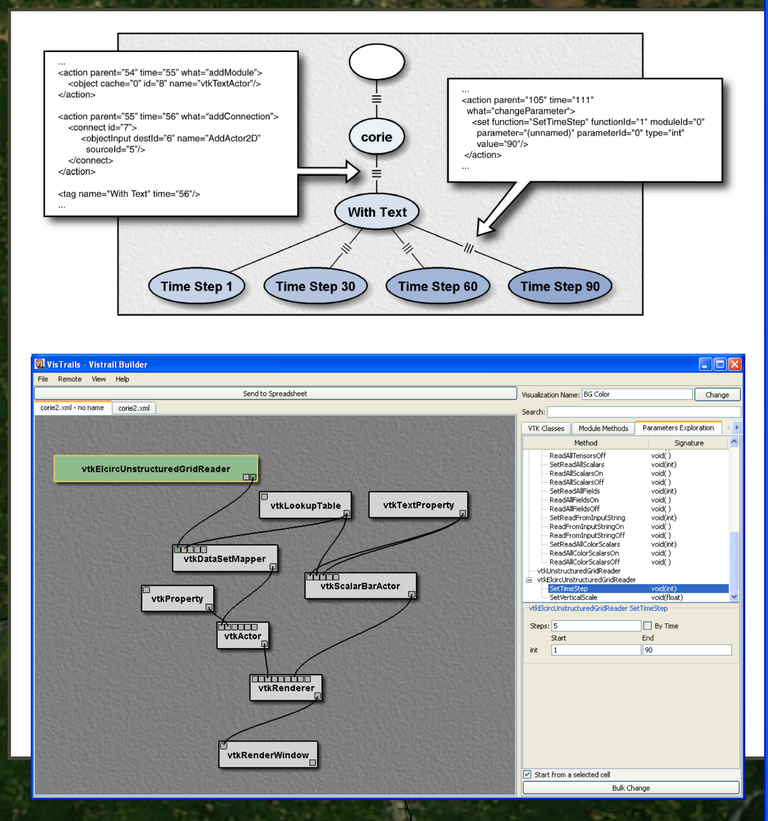
THE UNIVERSITY OF UTAH

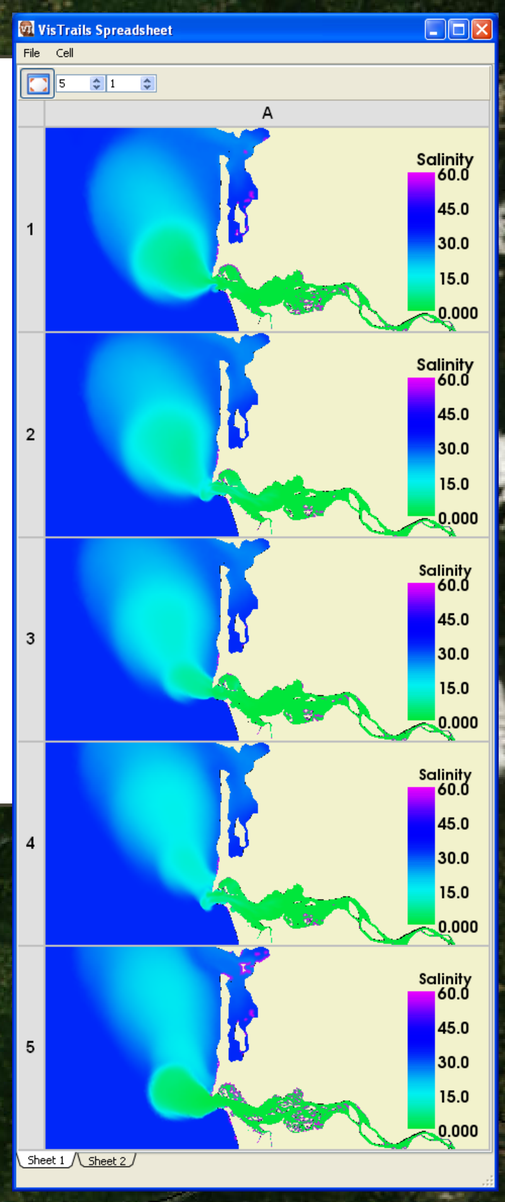
SCI www.sci.utah.edu

CORIE

CORIE is an environmental observation and forecasting system (EOFS) for the Columbia River. The goal of this multi-decade project is to predict complex circulation and mixing processes in a system encompassing the lower river, the estuary, and the near-ocean. Paradigms for modeling and visualization of complex ecosystems like the Columbia River are changing quickly, creating enormous opportunities for scientists and society. EOFSs seek to generate and deliver quantifiably reliable information about the environment in a timely manner. This can be a problem when real-time forecasts and simulations lead to tens of thousands of visualization products manually generated using a series of scripts maintained by the CORIE staff. This process is both error-prone and time consuming. In particular, it makes comparing the results of different simulations, time-steps, or rendering algorithms difficult.

VisTrails manages the data, metadata and the exploration process, while scientists can focus on science. By capturing the provenance of both the visualization process and data scientists manipulate, VisTrails enables reproducibility and simplifies the complex problem of creating and maintaining visualizations. The scientists can explore their visualizations by returning to previous versions of a workflow (or visualization pipeline); apply a workflow instance to different data; explore the parameter space of the workflow; query the visualization history; and comparatively visualize different results.

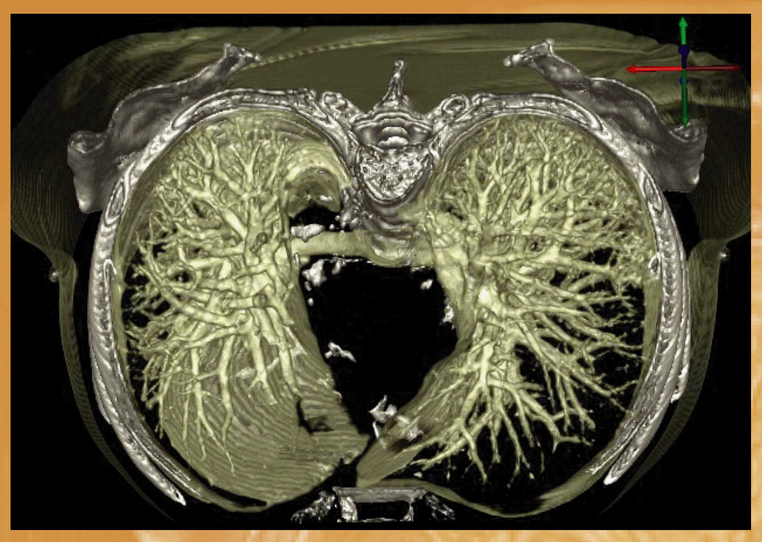




Top: A vistrail captures the evolution of workflows by recording the actions performed by the user to create a series of related visualizations. It consists of a collection of workflows---several versions of a workflow and its instances. We represent vistrails using XML, which provides the ability to easily query, share, and publish vistrails.

Bottom: Powerful operations are possible through direct manipulation of the version tree. These operations greatly simplify data exploration and enable the rapid creation of a large number of visualizations. Here, through the bulk-update interface, a visualization workflow is applied to multiple time steps of CORIE data.

Right: The VisTrails Spreadsheet is used to comparatively visualize workflows and data sets. In this case, visualizations of multiple time-steps of the CORIE data were generated with a bulk update and are displayed in synchronized views that can be manipulated interactively.

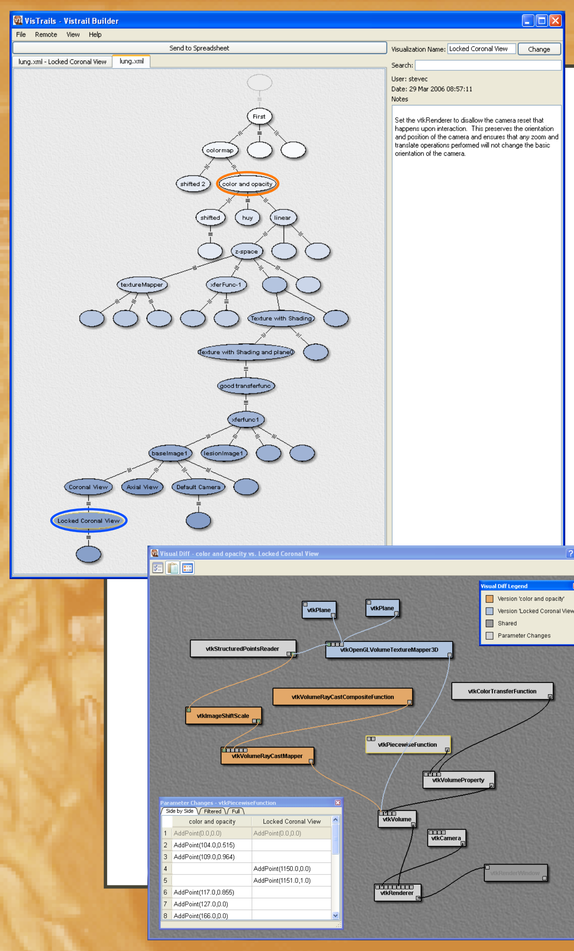


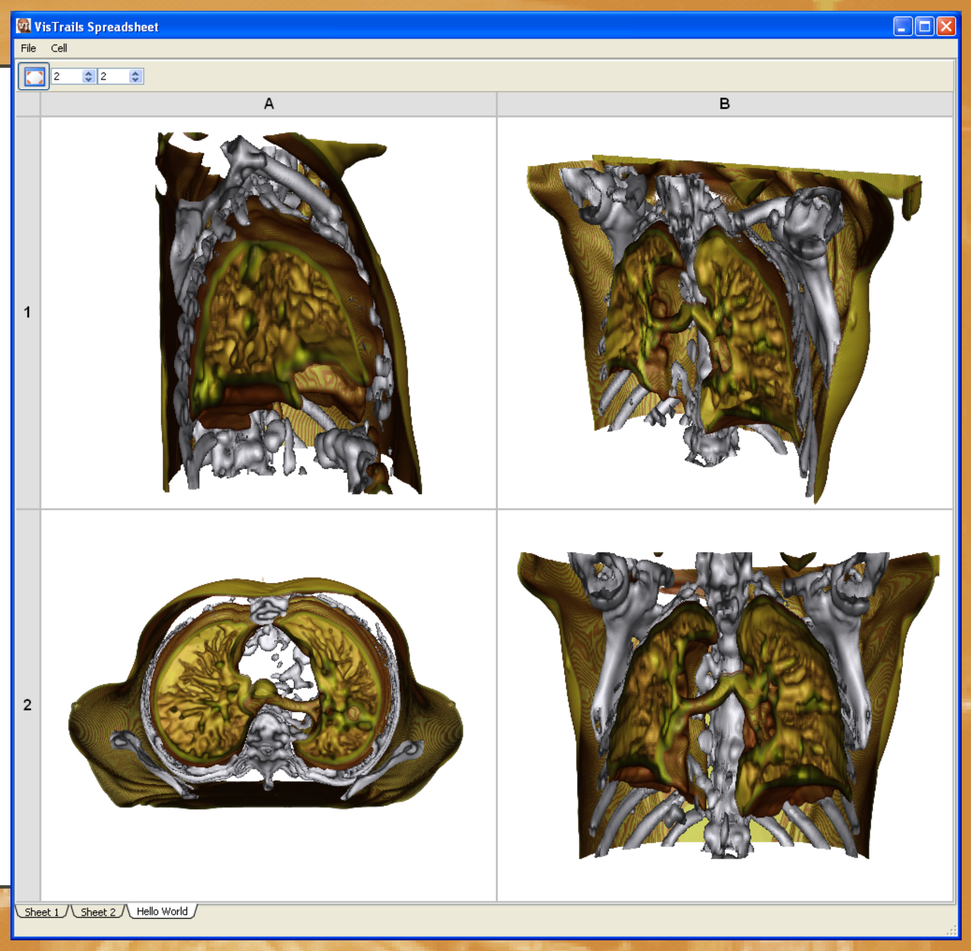
RADIATION ONCOLOGY

The Radiation Physics Division of the Department of Radiation Oncology at the Massachusetts General Hospital uses advanced visualization algorithms to facilitate the diagnosis process and to locate pathological tissue in preparation for radiation therapy treatment. The visualization process as currently deployed is very complex and time-consuming. Whereas a scanner can create a new data set in minutes, with advanced tools it takes from several hours to days to create appropriate visualizations. In addition, to allow reproducibility, many pages of hand-written notes and a large number of files need to be maintained. This greatly limits the ability to explore the large volumes of data, hinders collaboration, and hampers the useful lifetime of data.

VisTrails streamlines the process of data exploration through visualization by keeping detailed provenance of both visualization workflows and associated data. The system allows a user to return to previous versions of a workflow, query the visualization history, annotate workflows and images, and visually compare workflows. In essence, VisTrails replaces the laboratory notebook. Because VisTrails unobtrusively captures detailed provenance information and also manages visualization data, it removes the need for manually creating and maintaining large directory structures of resulting datasets and images as well as the need for writing detailed notes.

www.sci.utah.edu/~vgc/vistrails





Top: A snapshot of a vistrail showing the history of the visualization process for one patient. The complete tree has over 1,500 nodes. Here, only nodes tagged by the user are displayed and the presence of hidden nodes is indicated by edges crossed with short perpendicular lines. VisTrails allows an easy return to a previous workflow by selecting the corresponding node in the history tree. Detailed metadata for each workflow is maintained by the system, including the creator, date of creation, and additional notes. This information can be queried through the VisTrails search interface.

Bottom: By dragging one node onto another in a vistrail tree, the difference between two corresponding workflows can be easily computed and visualized. Modules that are shared between the two workflows are shown in dark gray. Modules unique to each workflow are shown in orange and blue. Parameter changes are displayed as light-gray modules, and can be enumerated in a pop-up window with a spreadsheet by selecting the changed module.

Right: The VisTrails Spreadsheet is shown with a series of visualizations, generated from a Computed Tomography (CT) scan from, for planning the radiation treatment of a lung-cancer patient.

Funding provided by