The Canopy Database Project

Component-Driven Database Design & Visualization

http://canopy.evergreen.edu

Judith Bayard Cushing, Nalini Nadkarni, Michael Finch, Youngmi Kim and Emerson Murphy-Hill

The Evergreen State College

Collaborating Ecologists & NSF LTER Information Managers:
Bond, Dial, Ishii, Parker, Shaw, Sillett, Sumida, Brunt, Kaplan, Menendez, Ramsey, Stafford, Vanderbilt, Walsh

Collaborating Computer Scientists:
Dave Maier, Lois Delcambre, Travis Brooks, Jim Thomas

OGI SCHOOL OF SCIENCE & ENGINEERING
OREGON HEALTH & SCIENCE UNIVERSITY

Project Vision

Use database tools to increase individual researcher productivity & enable documented data & integrated ecology research

- Focus on Research Productivity
- Validate data early
- Capture metadata close to the source
- Make visualization, etc., easy
- Enable cross-study queries
- Focus on the forest canopy

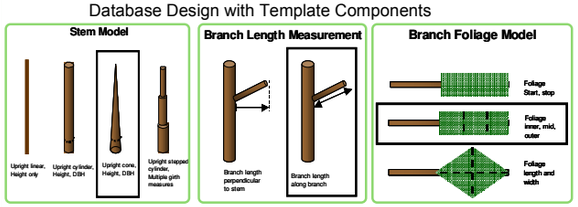
Project Objectives

Use database tools & visualization to increase individual researcher productivity & enable documented data & integrated ecology research

- Define database components (templates) based on structural aspects of the forest
- Implement a database design tool that uses those templates
- Provide visualization & other tools
- Enable cross-study queries

Canopy *Databank*

Database Design with Template Components



What is it
- End-user database design with templates
- Variable & table level metadata inherent
- Study-level metadata available from the BCD

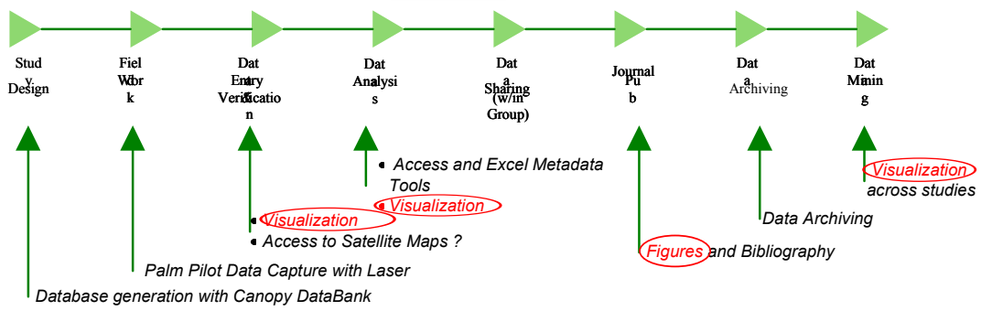
Technology
- HTML, Java, Enhydra, SQLServer, Access, JTK
- Aim to produce XML/EML for exchange and archive

Status
- Some templates (mostly spatial tree structure)
- About 5 field studies
- Some visualization

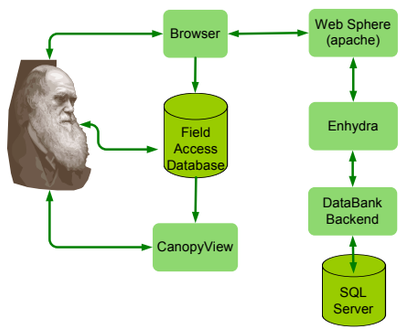
This Project Funded by
NSF CISE 0131952, BIR 9975510, 9630316

Please see our Vis 2003 Demo

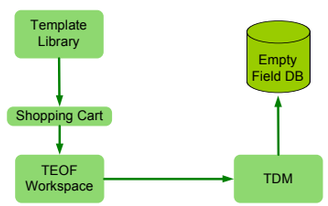
Easing Research Bottlenecks



Databank and *CanopyView* Architecture

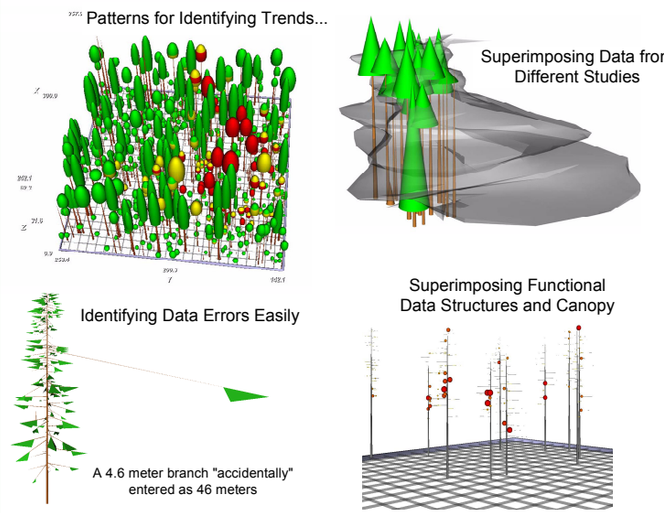


System Workflow



CanopyView

Visualization from Components



Abstract

Solving ecology problems such as global warming, decreasing biodiversity, and depletion of natural resources will require increased data sharing and data mining. This in turn will require better data infrastructure, informatics and analysis tools than are now available. Investments are being made in needed data warehouses for ecology, though a major bottleneck remains attaining adequate data documentation. Integrating database technology early in the research process would make metadata provision easier, but barriers to database use by ecologists are numerous. The Canopy Database Project is experimenting with database components for commonly used spatial (structural) data descriptions in one ecology discipline (forest canopy research), and we have two prototype systems, DataBank and CanopyView that exercise our vision.

Using domain-specific components for designing databases would make using databases easier, but explicit productivity gains must be evident before researchers would change their modus operandi. We have identified easier data visualization as a possibly effective reward, and our visualization program CanopyView, developed with VTK, takes as input databases designed from those components and produces visualization specific to structural aspects of the ecology study.

To conceptualizes how researchers might use conceptual components to design field databases with DataBank, consider three real-world canopy entities (stem, branch and branch foliage), and several spatial or structural conceptualizations that correspond to commonly measured variables for each. Imagine a researcher selecting the conceptualization that best matches his or her research objectives. DataBank then uses the selected components to generate a database design, validates it, and then generates SQL for any particular database system. (We currently use MS Access). Additional observations can be added to the generated database.

CanopyView is a visualization application that generates interactive scenes of ecological entities at the tree-level and plot-level using the same predefined data structures (aka database components or templates) used by DataBank to generate field databases. CanopyView uses an ecological field database (generated by DataBank and usually in MS Access) as its primary data source. The following figure shows scenes generated by CanopyView for several of our sample field data sets.

To the best of our knowledge, CanopyView is unique in that it produces visualizations directly of field data. Other visualization aids we have seen are either map-based or are essentially visual representations of statistical analyses. While those are essential, sometimes the scale of an ecological study such as for within-tree structure does not lend itself to a map-based first-cut visualization. Furthermore, our researchers have found that visualization of raw data contributes to their understanding of the data for data validation and discovery. CanopyView is implemented using the Visualization Toolkit (VTK) and Java.

We conclude that using components for field database design is feasible. Furthermore, databases thus developed can be used with a companion visualization application to generate scenes easily by end users. However, conceptualization of the components requires time and collaboration between ecologists and computer scientists; we are considering cost-benefit tradeoffs. VTK was a significant productivity aid in developing the visualization application.

For further information, see http://canopy.evergreen.edu , http://www.evergreen.edu/bdel , http://biernet.edu/ , http://www.ildeans.com and

J.J. Helly, Visualization of Ecological and Environmental Data, in W.K. Michener, J.H. Porter and S.G. Stafford, eds., Data and Information Management in the Ecological Sciences, LTER Network Office, University of New Mexico, Albuquerque, New Mexico, 1998, pp. 89-94.

Poster prepared for Visualization 2003