Curating and Preserving the Big Canopy Database System: An Active Curation Approach using SEAD

James D. Myers1, Judith B. Cushing2, Peter Lynott2, Noah Weinert2, Anna Ochchinnikova2, Nalini Nadkarni3, Anne McIntosh4

Preserving At-Risk Scientific Data and Cyberinfrastructure

Modern research is increasingly dependent upon highly heterogeneous data and on the associated cyberinfrastructure required to support its creation and visualization. However, it is the particular need for ensuring that data is seen and preserved in the long term at reasonable cost and in a way that retains its scientific value. In this presentation, we describe how this challenge is being met in an active risk assessment of data and tools — Canopy Science Data and Applications (CanopyDB; Project OAI), and the databases created using this software, The Big Canopy Database (BDCS), and CanopyView — using an active approach and leveraging the Sustainable Environment - Accessible Data [SEAD] project’s data curation services. The CanopyDB Project applications were developed over more than a decade and have been host to immeasurable callables on data from research on a broad range of forests.

Canopy DB

The Big Canopy database was designed to support structured and heterogeneous data repositories for a number of forest studies. The most extensive was a major forest project, the forest chronosequence dubbed the Thousand Year Forest Chronosequence. Objectives of this study were to characterize the composition, density, surface area, biomass, and spatial distribution of trees, and understory vegetation in a chronosequence of eight Douglas-fir western hemlock stands ranging in age from 50 to approximately 950 years, all located in the Western Washington Cascades: Wind River Experimental Forest, Carson, WA, Mount Baker National Forest, and Cedar Flats Research Natural Area. The study took place from September 1, 1999, to September 1, 2000.

In order to help forest managers of Washington State Public Lands use results from forest canopies, we developed a new research tool to help determine which trees to leave when harvesting a stand of trees, exploring how tree features such as canopy structure, broken tops, large branches, crown gap, and continuous crown could be related to policy rules such as “leave tree citrus,” and others. The tool was designed to work with unique CanopyView was used to generate visualizations of 100 trees for which detailed structure had been recorded.

Although the level of documentation by value study, the project website includes a wide range of detailed information including section description, data schemas, dictionaries, publications, and presentations, executable software manuals, photo galleries, guides describing how to navigate to field sites, and descriptions of experimental procedures and calibrations.

Canopy Cybertechnology

The CanopyDB applications were developed, with support from the U.S. National Science Foundation, over more than a decade at The Evergreen State College to address the needs of forest canopy researchers. CanopyDB is an early yet sophisticated exemplar of the type of cyberinfrastructure that has become common in biological research and science in general, with multiple relational databases for different experiments, a custom database generation tool used to create the databases, a video repository, a bibliographic and research reference tool, and desktop and web tools to visualize the data.

Components:

- **Database Database Generator**: Graphical tool to generate ecological databases from standard and custom database templates. Generated databases include data entry forms, dictionary, and Ecological Metadata Language (EML).

- **CanopyView**: An interactive visualization tool designed to view tree structure, using various models to display canopy structure. CanopyView is a web-based tree visualization tool that allows the user to view tree structure and to query the database.