An information infrastructure for vegetation science: A model and prototype database for storing and integrating vegetation data

John Harris[1], Marilyn Walker[2,3], Robert Peet[4], Dennis Grossman[5], Michael Jennings[1].

National Center for Ecological Analysis and Synthesis *1*,
USDA Forest Service Boreal Ecology Cooperative Research Unit *2*,
University of Alaska Fairbanks *3*, University of North Carolina *4*, Association for Biodiversity Information *5*

ABSTRACT

North American ecologists, including government agencies, academics, and private organizations, have collected an enormous legacy of vegetation data over the past century. The value of these data increases greatly over time, and an emerging classification requires combining new data with existing data and using "legacy data" in new ways. The ESA panel on North American Vegetation Classification has been developing standards for collecting, analyzing, and classifying vegetation data that will meet a variety of needs and will be compliant with federal standards. A critical need identified by this panel is a set of electronic tools for storing, managing, and updating data. This represents a significant technological challenge because of the continuous updating of taxonomic and syntaxonomic elements central to it. We are developing a prototype database system that is both perfectly archived (data can be retrieved as they were at any point in time) and continuously updated (data can be retrieved to be consistent with new concepts and changes). The model is based on three linked databases: one containing the plot data, one containing the taxonomic information that is the basis of the plots, and one containing the vegetation classification based on the plots. We present the data model for the entire system and a working prototype plots database.

This project is funded by NSF grant: DBI-9905838

Project supported by:
National Center for Ecological Analysis & Synthesis
U.S. National Science Foundation
USGS-BRD Gap Analysis Program
ABI / The Nature Conservancy

Project Design

Typical Vegetation Related Project

Vegetation Information Infrastructure Impact

PUBLICATION

For Plot-Related Publications:
- Ref erence to Accession Number in Plots Database .
- P otential to Annotate Plots Stored in the Plots Database .

ANALYSIS

Vegetation Data Export To Analytical Tools Including:
- Ar cView
- Canoco

EXTRACTION

Web-accessible Query Mechanism.

Vegetation Database Workbench Software to Query Either the Central Vegetation Database or the Local Embedded Database.

Complex Nested Queries Leveraging Logical and Boolean Statements.

ARCHIVAL

Plots Database Based on a Comprehensive Data Model which Includes Hundreds of Commonly Collected Plot Attributes.

Databases Designed with Scalabilty in Mind.

Vegetation Community and Plant Taxonomy Databases Based on Emerging Concept-Based Taxonomic Data Models.

INTEGRATION

Integration of Legacy Plots Data From:
- Go vernment Agencies
- Academic Institutions

'Concept-Based Taxonomic Normalization'
- USD A Plant List
- ITIS

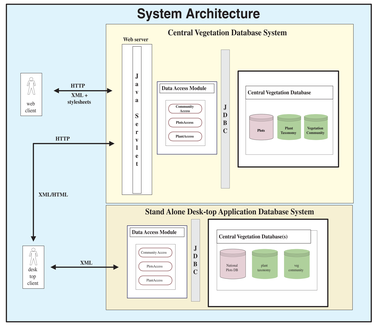
Dynamic Linking Between Plots and Vegetation Community Assignment.

Uses Extensible Markup Language (XML) to Automate the Import of Data from Legacy Data Sources to the Vegetation Databases.

COLLECTION

Information System Design

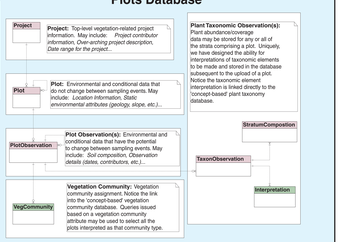
System Architecture



Plots Database

The vegetation plots database structure is comprised of many entities which, collectively, will be used to store and integrate plots that have been collected according to a wide range of protocols and stored in disparate data formats. Paramount to our plots database design exercise was the capability to:

1.) Support the National Vegetation Classification.
2.) Provide a comprehensive facility to store the most commonly collected vegetation plot data attributes.
3.) Provide the user with a large number of user-defined attributes to store not-so-commonly collected data.
4.) Integrate plots with the dynamic plant taxonomy and vegetation community data.

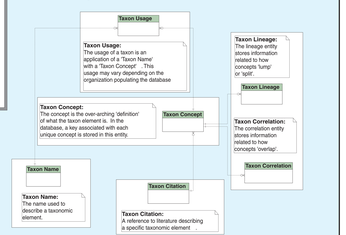


Concept-Based Taxonomic Database

The 'Concept-Based' taxonomic database module is used to store both the **Vegetation Community** and the **Plant Taxonomy** data". Key to the design of the Taxonomic database were the following:

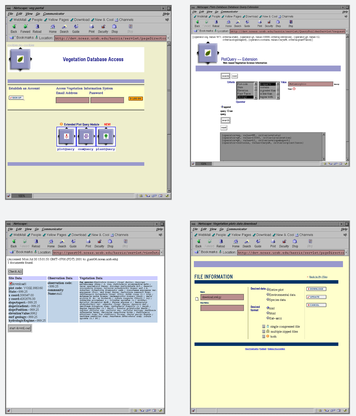
1.) Logical separation of a "taxonomic name" with the "taxonomic concept", so that taxonomic data can be stored at the most 'atomic' level without ambiguity.
2.) The ability to incorporate multiple organizations' 'views' of how a taxonomic name is applied to a taxonomic concept
3.) The ability to link a taxonomic name used in the Plots database with a 'name - concept' pair in the taxonomic database.

*Although one can store vegetation community data in the same database table-structure as the plant taxonomy database, we have implemented two separate table structures and have created two separate data sets.



Desktop Software

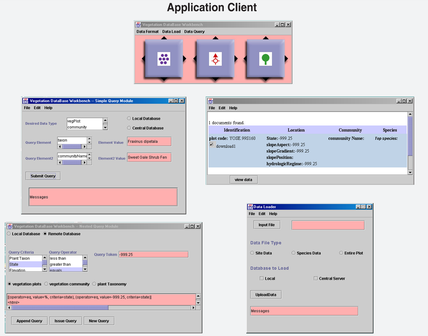
Web-Based Client



The web-based client facilitates the interaction with the National Plots Database through a standard web-browser. With these tools the user may:

1.) Upload plots data directly for review and database ingestion.
2.) Query plots, vegetation community and plant taxonomy data.
3.) View data summaries based on the query results.
4.) Extract data in a variety of formats.

Application Client



Using the desk-top application, a user may interact with the National Plots Database or manage their own plots data locally on a built-in database which is consistent in structure with the national database. Specific functionality for this application includes:

1.) Data format transformation from legacy format to the format used b y the plots database.
2.) Query plots, vegetation community and plant taxonomy data stored in both the national database as w ell as the local database.
3.) View data summaries based on the query results.
4.) Perform detailed plant taxonomy and vegetation searches f or use in plot annotation.
5.) Export data in a variety of formats.
6.) Populate local database directly with results from queries issued against the national database .
7.) Upload data from the local filesystem to the local database and to the national database .

Plots Data

Plant Data

Community Data