Integration of Bio-Geo spatial database for selected watersheds in Himalayan region

A. K. Gosain[1,]*, Sandhya Rao[2], Puja Singh[2] and Anamika Arora[2]

[1]Department of Civil Engineering, Indian Institute of Technology, Delhi 110 016, India
[2]INRM Consultants Pvt Ltd, New Delhi 110 016, India

An integrated database has been designed to cater to the data collected by different agencies on various aspects of biological and geophysical data, as part of the coordinated programme on 'Bio-Geo database and ecological modeling for Himalayas' initiated by NRDMS division of the Department of Science and Technology. The spatial database conforms to the NSDI standards. In order to facilitate the unified database system, an attempt has been made to design a geodatabase for Bio-Geo data using the state-of-art design techniques. The model is specified in an industry-standard modelling notation called the Unified Modeling Language and conforms to the same. The web interface is developed to evolve a system of information sharing, access and use of the Bio-Geo data collected for different watersheds. GIS-based query system developed has user-defined query and custom report facilities. The present article reports the detailed design and integration of Bio-Geo database. Need assessment carried out has also been discussed.

Keywords: Bio-Geo databse, geodatabase, need assessment, Unified Modelling Language, web interface.

Introduction

THE Bio-Geo database project was initiated by the NRDMS. Pilot watersheds from Upper, Middle and Lower Himalayas were identified in Himachal Pradesh and Uttarakhand. A large number of groups/institutions were involved in the collection of data under different sectors representing biological and geophysical domain. The collation of this diversified spatial and non-spatial data requires standardization and a detailed database design. This article presents the various aspects of the effort starting from conceptualization to implementation of the Bio-Geo database. This component is intended to be implemented in the National Spatial Data Infrastructure (NSDI) framework of the country.

*For correspondence. (e-mail: gosain@civil.iitd.ac.in)

Methodology

To standardize, design and integrate the spatial and non-spatial databases for Bio-Geo data, the following steps were performed: (i) Need assessment to identify the objectives and the scope, (ii) Database design on the basis of identified objectives, (iii) Standardize the data collected and create metadata and integrate spatial–non-spatial databases and (iv) Web application development.

Need assessment

Need assessment is the first step in implementing a successful Geographic Information System (GIS). Need assessment is a systematic look at how departments function and the spatial data needed to do their work. This activity itself serves as a learning tool where potential users in each participating department learn about GIS and how it can serve the department.

As the first component of the project the requirement analysis (need assessment) was to be performed. The principal investigators of various projects have carried out user need assessment with the local authorities (comprising of district administration, agriculture, irrigation, soil, fisheries and other user departments). A list of applications was identified based on the reports. The designing paradigm was to be based on the end-user specification, but in the absence of proper understanding of the different aspects of this application, the designers came out with a preliminary module with their perception of the requirement. It was also felt that this would facilitate the process of need assessment with the users who are not computer savvy. Table 1 gives salient applications and their use for various end-user departments.

A detailed outline addressing each of these identified applications was drawn. Each application is divided into five templates. The first template gives a brief description of the application with the department which has identified the application, the task to be performed and the data needed. The second template outlines the GIS data model in terms of the Entity Relationship diagram. The purpose of the data model is to ensure that the data has been identified and described in a completely rigorous manner. The

Table 1. Salient applications after need assessment exercise

Applications	Departments
Micro-watershed map	All departments
Drainage pattern map	Agriculture Dept, Rural Development Dept, Horticulture Dept, Forest Dept, Irrigation Dept
Diversion gate/water tank location	Rural Development Dept, Forest Dept, Agriculture Dept, Horticulture Dept
Water availability	Rural Development Dept, Agriculture Dept, Horticulture Dept, District authorities, ZSI
Water source for inhabited villages	Rural Development Dept
Landuse and land capability maps	Rural Development Dept, Agriculture Dept, Horticulture Dept, Forest Dept, District authorities
Wasteland availability/development/ treatment map	Rural Development Dept, Agriculture Dept, Horticulture Dept, Forest Dept, District authorities
Soil conservation map	Agriculture Dept, District authorities
Soil series map	Rural Development Dept, Agriculture Dept, Horticulture Dept, Forest Dept, District authorities, Irrigation Dept
Irrigated area map	Rural Development Dept, Agriculture Dept, Horticulture Dept, Irrigation Dept
Cropping pattern	Rural Development Dept, Agriculture Dept., District authorities, Irrigation Dept
Crop intensity map	Rural Development Dept, District authorities, Agriculture Dept
Diversity (flora/fauna) map	ZSI, Horticulture Dept, Agriculture Dept, Forest Dept
Biofertilizers map	Horticulture Dept, Agriculture Dept
Meteorological map	All departments
Landslide information maps	District authorities
Potential growth centre map and land suitability and availability map	District authorities

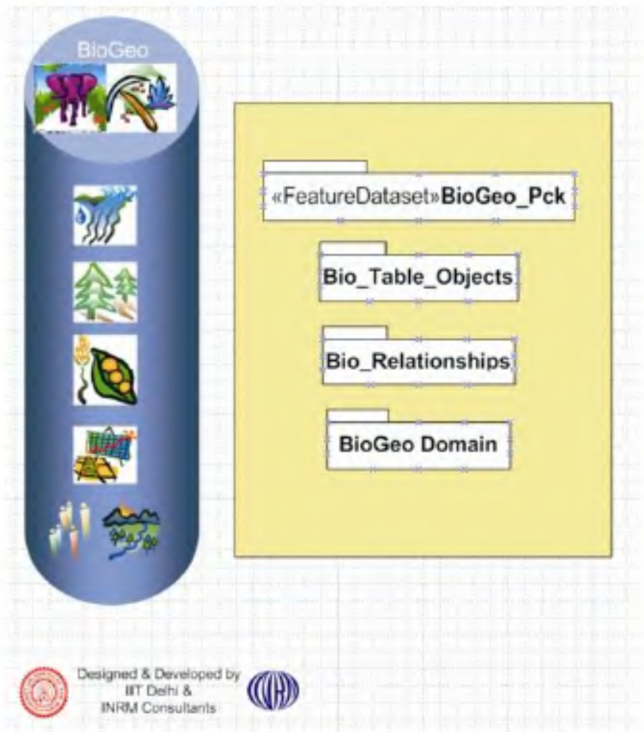

Figure 1. The core model packages of Bio-Geo database.

data model is the formal specification for the entities, their attributes and all relationships between the entities for the GIS. The information needed to develop the E–R diagram representing the spatial data model comes from the need assessment. The third template contains the table design to keep all the identified attributes of the spatial entities. The attributes of the entities may be held in a relational database management system linked to the GIS. The fourth template provides the user with query and map output option. The fifth template caters to the customized reports.

UML design of Bio-Geo data model

Data models are a major component of the GIS, as they determine the ways in which real-world phenomena may best be represented in digital form. Data model provides a basic template for implementing GIS projects (i.e. inputting, formatting, geo-processing, sharing data, creating maps, performing analyses, etc.). It provides a basic framework for writing program code and maintaining applications. A data model for Bio-Geo applications is complex because of the multidisciplinary and interconnected data. To facilitate the unified database system, an attempt has been made to design a geodatabase for Bio-Geo data using the state-of-art design techniques. The model is specified in an industry-standard modelling notation called the Unified Modeling Language (UML)[1]. This conceptual design is translated into core packages.

Core model packages: The core model packages have been organized into spatial objects, non-spatial objects, relationship objects and domain objects (Figure 1).

Bio-Geo feature dataset: Figure 2 shows the objects of the feature dataset. A feature dataset is simply a collection of feature classes that share a common spatial reference. A spatial reference is part of the definition of the

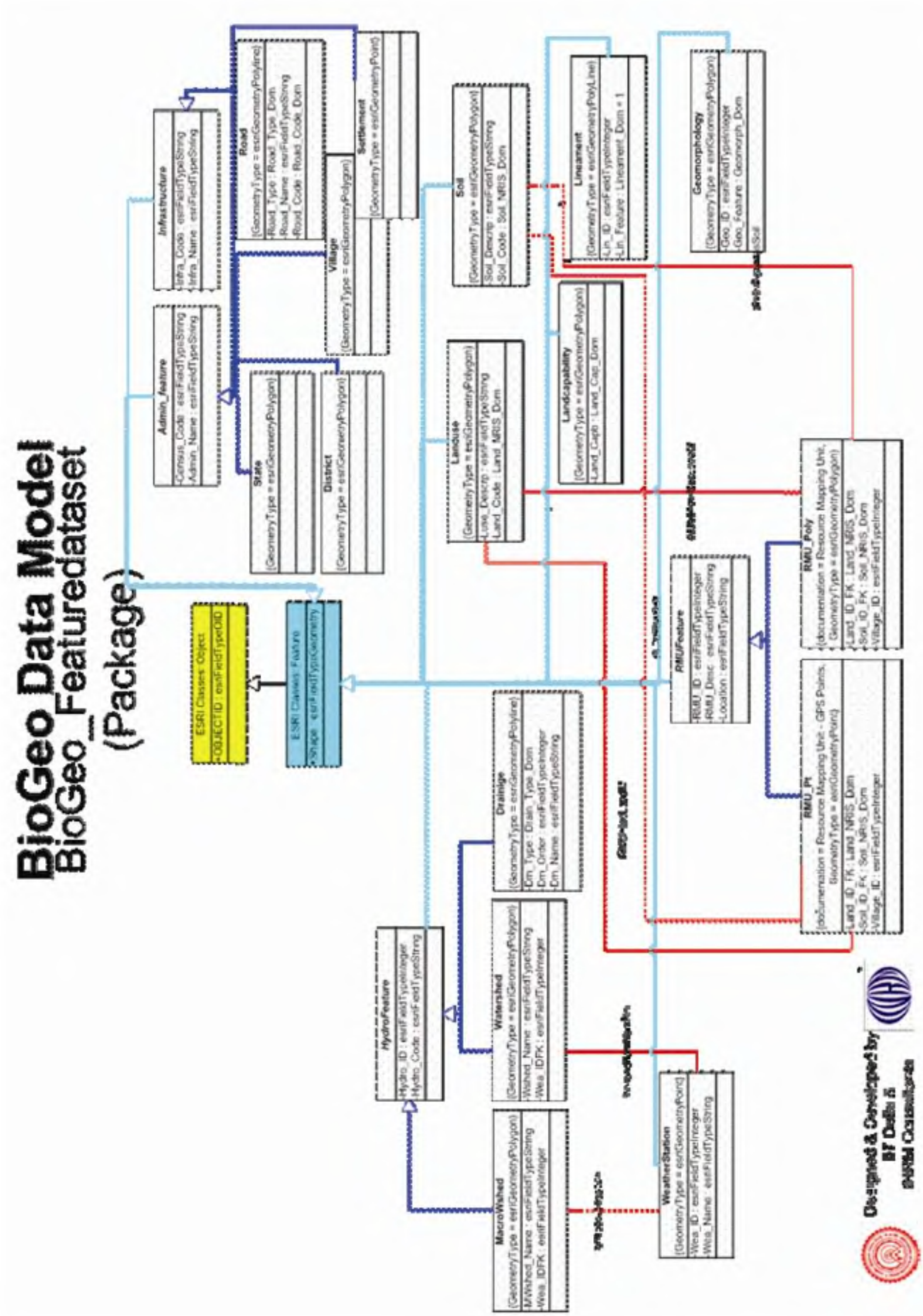

Figure 2. Bio-Geo feature dataset.

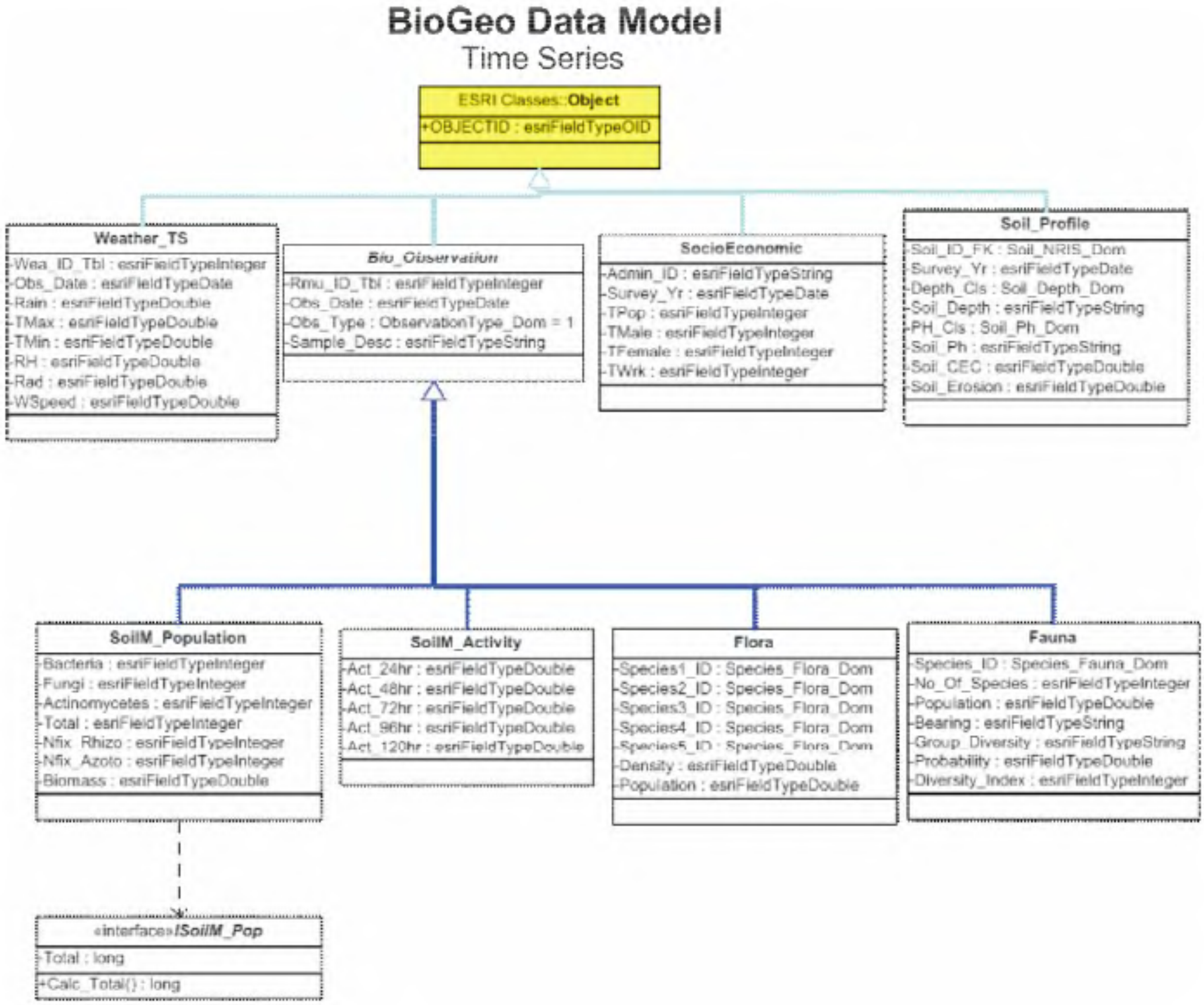

Figure 3. Bio-Geo table objects.

geometry field in the database. Features are geographic objects that have a spatial location defined. The major features include: watershed, land-use, soil, resource mapping unit, administrative unit, settlements, roads and weather stations.

Bio-Geo table objects: The objects which do not have a geographic representation, such as observations on flora are stored as table objects. Figure 3 shows the structure of the objects which include the time series data along with the other static data. Some of the objects include: weather observations, flora and fauna observations, soil microbial, socio-economic data and soil profile data. These tables are related to the feature dataset (spatial objects) through relationships.

Relationship objects: Spatial or explicit relationships can exist between objects or features. Spatial relationships are merely the spatial coincidence of features. Explicit relationships are created between objects as a relationship class (Figure 4). One object serves as the origin class and the other object is referred to as the destination class.

Domain objects: Domain objects define a set of legal values for a field's attribute. These are the standard codes or the ranges which are used to define an object, for example, biological codes which are standard codes for a particular flora. Figure 5 shows some of the coded value domains.

Implementation of geodatabase

Geodatabase[2] is a type of geographic data format (GDF) based on object-oriented model. Users can add behaviour,

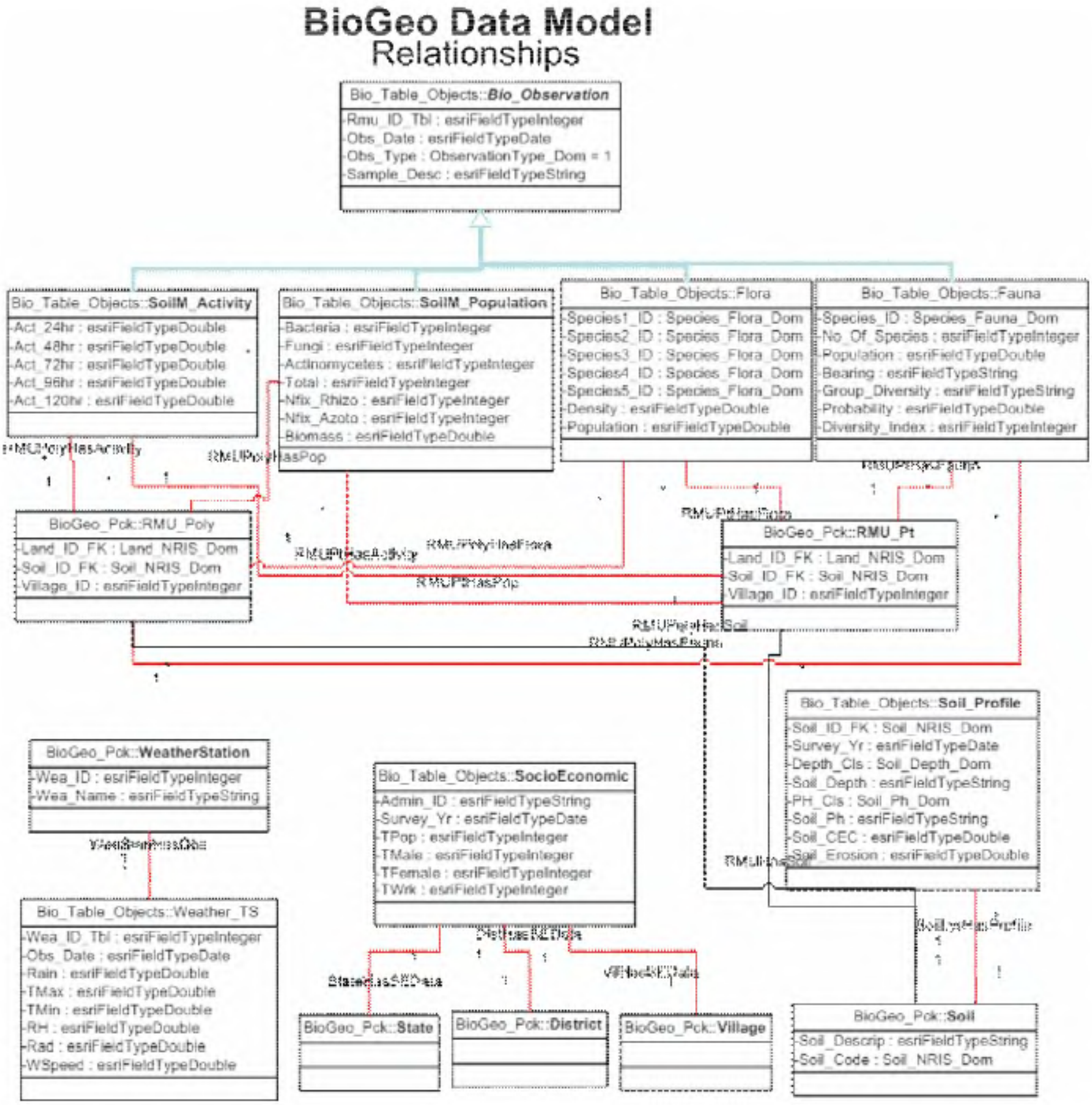

Figure 4. Relationship objects.

properties, rules and relationships to data and this is implemented as extension to standard relational database technology. This supports topologically integrated feature classes. Geodatabases organize geographic data into a hierarchy of data objects. Features classes are collections of features with the same type of feature geometry and attributes. CASE tools are used to create new custom objects and generate a geodatabase schema from UML. The procedure to convert UML schema to geodatabase involves first exporting UML schema from the case tools to XML and then generating the geodatabase schema from the XML using Schema Wizard. Figure 6 shows the geodatabase generated from the UML schema for Bio-Geo data model.

Web application development

The web interface is developed to evolve a system of information viewing, access and analysis of the Bio-Geo data collected for different watersheds. An interface has

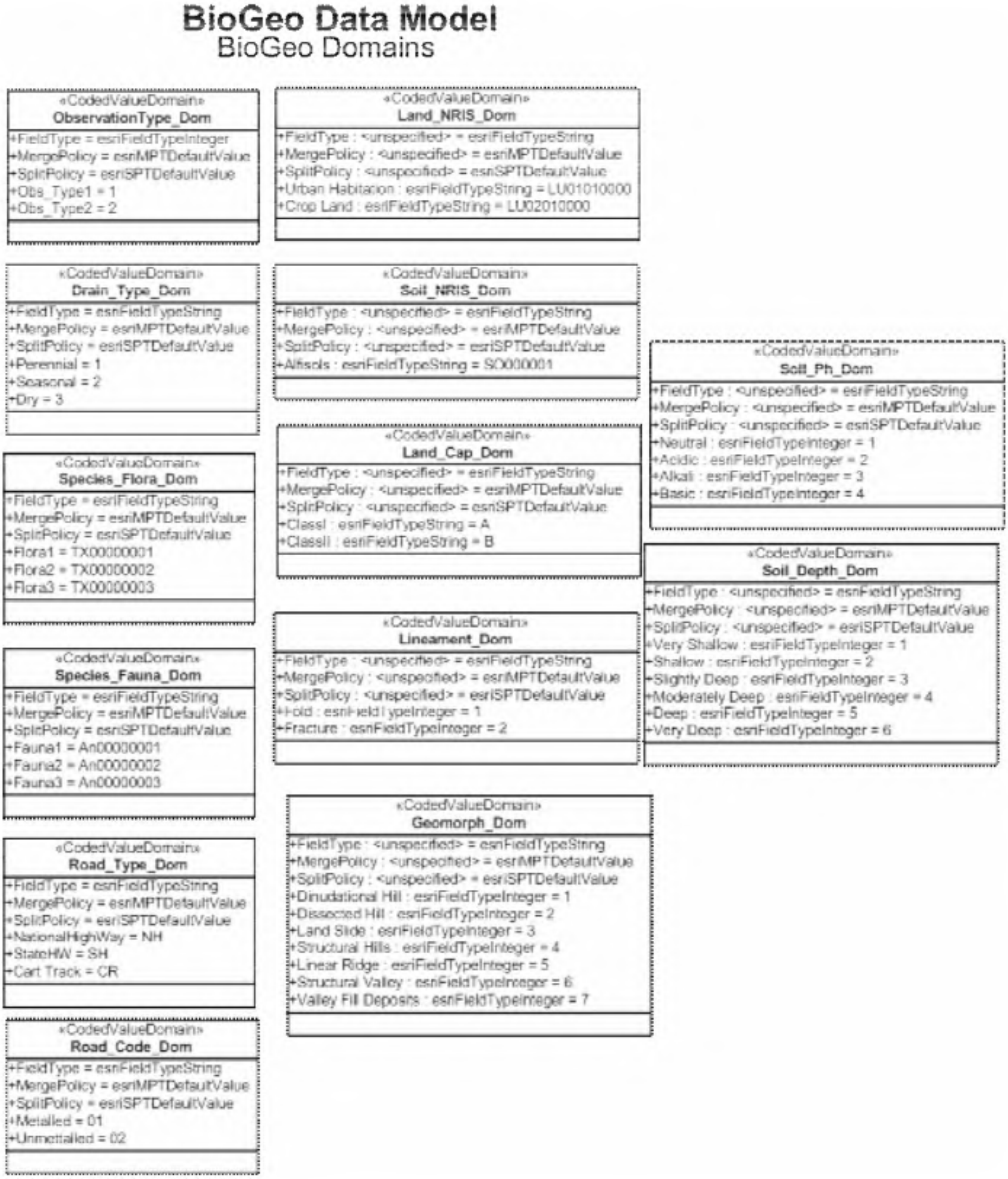

Figure 5. Major domain objects.

been developed to showcase the kind of applications and query which can be performed on the Bio-Geo database. This interface caters to querying the data collected by different PIs. Care has been taken to make it user friendly and at the same time serve the user community with functionalities to retrieve the relevant information.

The salient features include navigation to a general description, site locations, Bio-Geo data model template

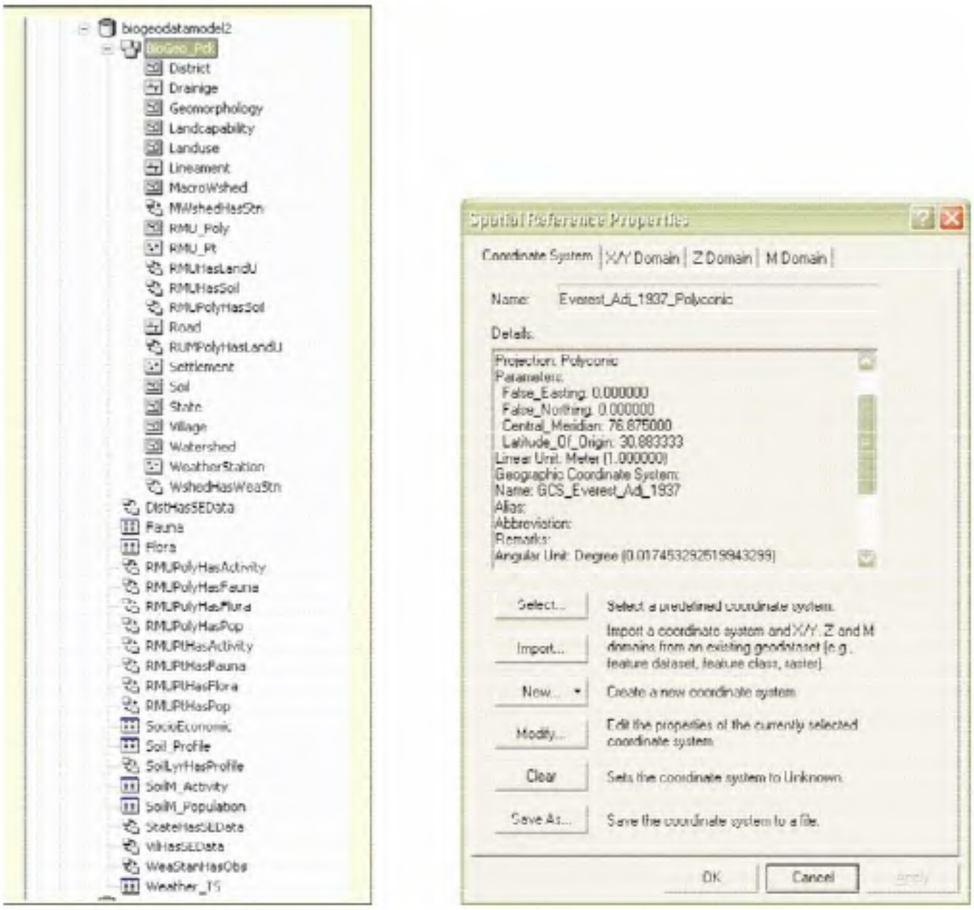

Figure 6. Bio-Geo geodatabase template.

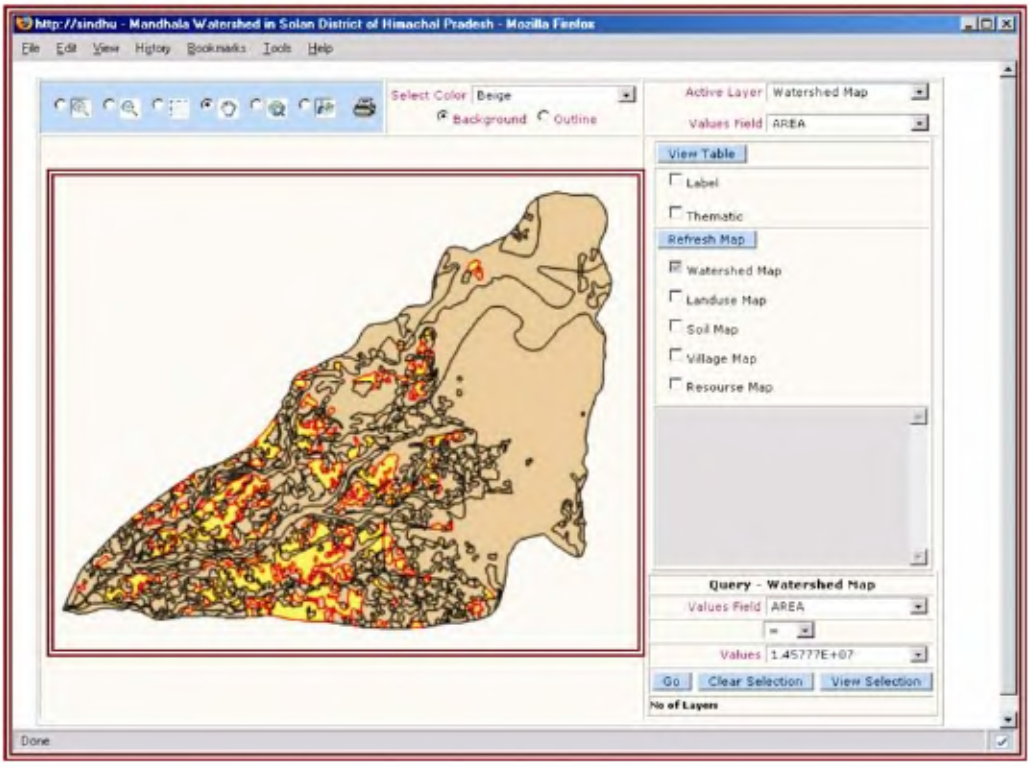

Figure 7. Mandhala watershed (Shimla, Himachal Pradesh) showing the landuse.

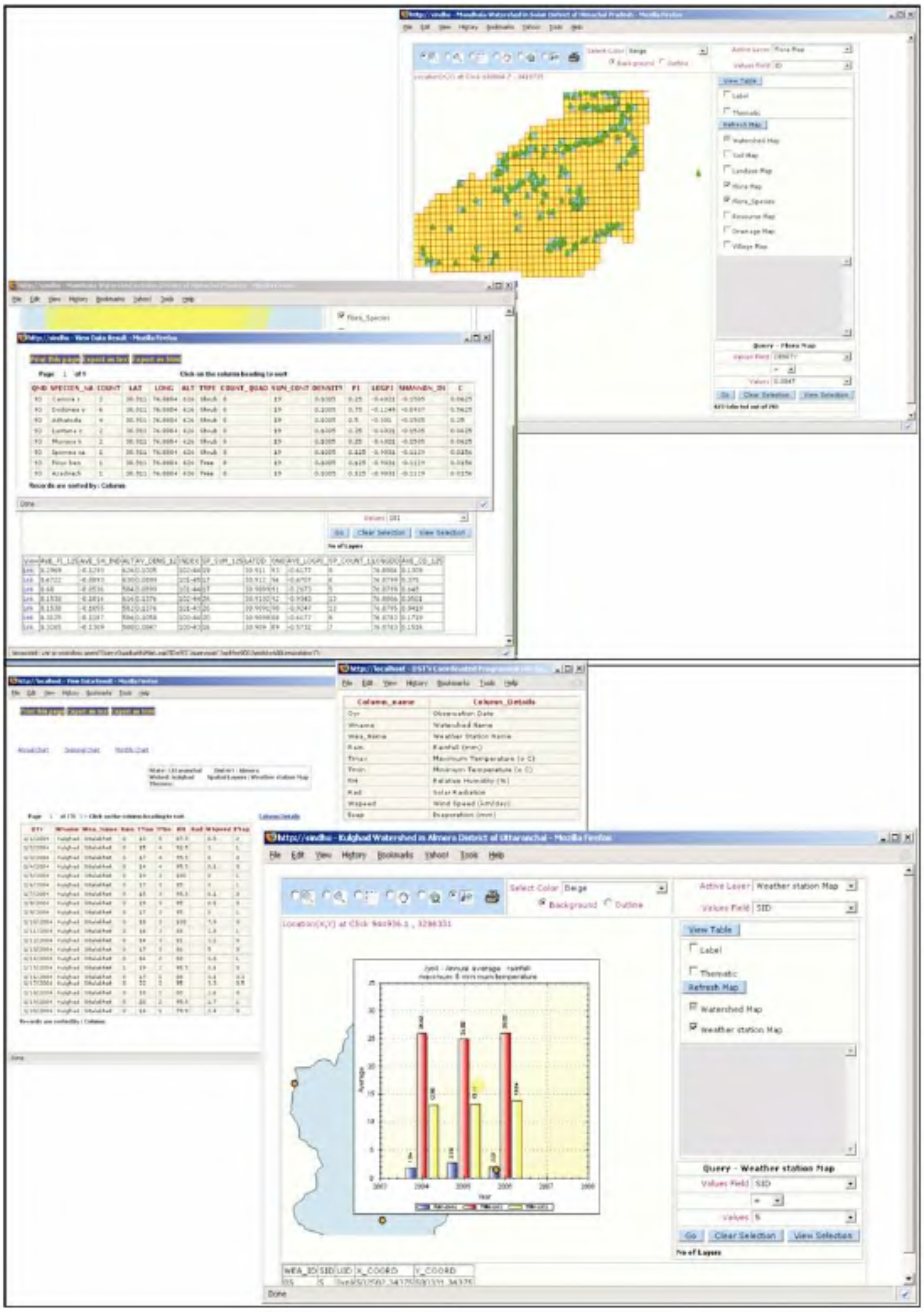

Figure 8. Kulghad watershed (Uttarakhand) showing floral biodiversity and weather data.

download, download other documents and presentations and access to Bio-Geo database. A brief description of the query module, i.e. access Bio-Geo database module is given below.

Access Bio-Geo database module: This module facilitates users with map-based query and generates custom reports. Map-based query has been divided into two sections, viz. map-based query and table-based query. Drop down box selection to choose the watershed belonging to a particular state or district is provided. Choice of spatial layer can also be made. The data can be viewed in the map and tabular format. Two such screenshots are shown in Figures 7 and 8.

The data obtained from different groups (Himachal Pradesh and Uttarakhand (Garhwal and Kumaon transects)) for their representative micro-watersheds (size between 500 and 5000 ha) covering lower, middle and upper Himalayan altitudinal transects has been integrated. The watersheds include, Me Gad in Lahaul and Spiti, Mandhala in Solan, Moolbari in Shimla in Himachal Pradesh, Dabaka in Nainital, Dhulgarh-Uttari Kosi in Almora, Saryu in Bageshwar belonging to Kumaon region, and Pathri Rao in Haridwar, Fakot in Tehri, Gomukh in Uttarkashi of Gharwal region of Uttarakhand. The data integration for the three transects of Himachal Pradesh and two transects of Uttarakhand has been completed for the water, land, soil, flora and microbial sectors. Figure 8 shows the spatial floral species diversity and the time series weather data in table and chart form.

Conclusions

An integrated database has been designed to cater to the data collected by different agencies on the various aspects of biological and geophysical data. A geodatabase for Bio-Geo data using the state-of-art design techniques has been designed using UML. The data integration for the three transects of Himachal Pradesh and two transects of Uttarakhand has been completed for the water, land, soil, flora and microbial sectors. A web interface is developed to access and use the Bio-Geo data collected for different watersheds which caters to user-defined GIS-based query and generates custom reports.

1. A Quick Guide to the Unified Modeling Language (UML), http://www.csci.csusb.edu/dick/samples/uml0.html
2. Arctur, D. and Zeiler, M., *Designing Geodatabases: Case Studies in GIS Data Modelling*, ESRI Press, 2004.

ACKNOWLEDGEMENTS. We thank the Department of Science and Technology, New Delhi and Dr R. Sivakumar, Head, NRDMS for entrusting this study. We also thank Ms Nisha Mendiratta, Mr P. S. Acharya for their help and constructive suggestions. We also thank Dr K. S. Rao, Prof. J. S. Rawat and all the PIs of the Himachal Pradesh and Uttarakhand transects for help.