Data mining and mapping of herbarium specimens using geographic information systems: A look at the biodiversity informatics project of the W. P. Fraser Herbarium (SASK)

Catherine Peters, David Peters and J. Hugo Cota-Sánchez
Department of Biology and W. P. Fraser Herbarium (SASK), University of Saskatchewan, Saskatoon, SK

INTRODUCTION

The W. P. Fraser Herbarium (SASK) is devoted to the study of the Saskatchewan flora. With nearly 179,000 specimens, SASK has a comprehensive chronological, historical, taxonomic, and geographic coverage of provincial plants, except those from inaccessible areas. The significance of the W. P. Fraser Herbarium collection to Saskatchewan biodiversity is evident: the flora of Saskatchewan is essential for scientific research in plant taxonomy, systematics, conservation, and collection management and represents the province’s past and present legacy. The combination of traditional technologies of historical specimen record preservation with modern computer technology will facilitate the mobility of existing data stored in SASK’s collection via the World Wide Web. Our ongoing biodiversity informatics program will provide information and tools relevant to conservation management and modeling systems and other biodiversity programs dealing with the flora of Saskatchewan and Canadian biodiversity. The first phase and scope of this project integrates the Specify software with an on-line geographical information system to allow queries of herbaria records, generate specimen geolocation maps, and access data for analysis.

OBJECTIVES

• To develop a data export system using the Specify (ver. 5) to MySQL 5.1 to store and retrieve SASK data on-line.
• To implement a webpage linked to SASK website to allow specimen queries based on scientific name, collector, date collected, locality, ecoregion, and geographic coordinates.
• To develop a GIS using Google™ Maps API (application programming interface) to graphically display the geolocation of voucher specimen retrieved from the MySQL database at SASK.

METHODS

Phase I of this project is limited to 11,685 Saskatchewan specimens collected from 1905 to 2005 and available in SASK's Specify database. Specimens in endangered categories are excluded to protect sensitive information. Phase I is accessible at: http://herbarium.usask.ca/html/MapDevelopment/maps.html. It operates using the Mozilla Firefox (ver. 3.0.8) browser for Windows® and Macintosh®, and Safari (ver. 3.2.1) browser for Macintosh®. This program is not yet fully functional for Windows® Internet Explorer.

The data set includes the following voucher information: familial, generic, and specific names, locality, province, longitude and latitude coordinates (decimal degrees), ecoregion (Environment Canada), collector’s name and number, date collected, and SASK’s accession number. Missing fields on the label were entered as “Unknown”.

RESULTS

Retrieving Data

Phase I of this project allows the retrieval of data in tabular format (Fig. 1) by choosing the “Add to Table” option, plotted graphically on a Saskatchewan map (Fig. 2) by choosing the “Add Pointers to Map” option, or by choosing both options. The search option can be narrowed to “genus”, “species”, “locality”, and/or “collector name”.

Other options featured include:

• The user can add information to the table or map by choosing the “Add to Table” and “Add Pointers to Map”. Also, the data can be replaced on the map or the table with new data using the “Replace Table” and “Replace Pointers”.
• The data can be sorted on any field criteria in ascending or descending order.

Mapping Options

Our working model of the GIS shows a map with the geographic boundaries of Canada, sized to 1000px (width) by 500px (height). The “termin” display is the default view, but the Google™ software allows changing the map display from a political road map, a satellite display, or a hybrid map, which shows both the satellite option and political boundaries, in addition to interactive zoom features and positioning of the map within the screen view. The physical coordinates of the specimens are indicated on the map with green pointers (Fig.2).

DISCUSSION

Planning and implementing effective conservation efforts without the necessary tools required to determine areas of priority is challenging. Hence, GIS has been incorporated to assist in the assessment biodiversity and the development of environmental models to monitor and evaluate changes using herbarium specimens.[1,2,3] In our experience, maps API are very useful, but our work is limited to the guidelines established by the Google™ on-line documentation and use of map application tools and codes provided. As a result, we are unable to map trends nor compare data sets with this system. Upgrading to a more robust GIS software program such as ArcGIS is necessary, as it will provide a platform for creating, editing, and analyzing biodiversity data more effectively. With ArcGIS the data can be viewed on a map and analyzed in different ways to unveil patterns, relationships, and trends that are not apparent in tabular format. In conclusion, the digitization and sharing of biodiversity data via the Internet is a key towards more efficient policy-making decisions in environmental and biodiversity management issues at the local and global levels. The combination of herbarium data with a GIS provides information in a form that is more coherent, easy to understand and communicate, and applicable to decision makers.

FUTURE DIRECTIONS

Several changes are planned to improve the functionality of our website, including:

1) To add a date field to display range dates to study historical trends in the data, e.g., distribution/changes of native and invasive species over time.
2) To create a tab to download tabular data in CSV (Comma-Separated-Value) format, which is used to import data from one software application to another.
3) To implement a user button to create a downloadable file of the map display.
4) To expand the data set and include specimens from the rest of Canada and the world, SASK’s ultimate goal.
5) To include a hyperlink to the voucher digital image in JPEG format.
6) Refine the source dataset to minimize error and improve quality of the data used in the mapping program.
7) Upgrade to ArcGIS software to allow layering options and enhance programming and reporting features.

LITERATURE CITED