Mapping Biogeographic Patterns from Species Occurrence Data – the Biodiversity Analysis Tool

Dan Rosauer - Natural Heritage Assessment - Australian Department of the Environment & Heritage
Bryn Honeyman - National Pollutant Inventory - Australian Department of the Environment & Heritage

Distributed Biological Collections

Biological collections such as museums and herbaria hold many millions of specimens of organisms collected from where they naturally occur. These collections are the authority and information source for taxonomy of the worlds' species, but access to the data in these collections has been limited.

Digitisation of these collections has vastly increased their value and usability, and spawned a range of new applications. Free, open and efficient exchange of this data has been championed by the Global Biodiversity Information Facility (GBIF) an international organisation, with 26 countries as members. GBIF participants maintain data in their own databases, but adopt standards to facilitate access to the data. The GBIF approach includes standard data structures, and standard web service protocols for automated querying of their data. The GBIF website provides an index to data providers and to the data itself. There are now 159 data providers serving over 86 million records.

So far, the emphasis of this developing network has been to maximise the amount of available data, but, opportunities for direct online application of this distributed data are now being explored.
GBIF funds projects to demonstrate applications of the GBIF data network. In 2004 the Australian Government Department of the Environment & Heritage was funded to develop an online tool to use GBIF data sources for biogeographic analysis and mapping.

Biodiversity Analysis Tool

The Biodiversity Analysis Tool (BAT) uses data drawn automatically from biological databases, to analyse and map the biogeographic patterns in the distribution of groups of species. For example, it can identify areas where many species are concentrated, or where a number of species are limited to a small distribution range (centres of endemism). These analyses can help to prioritise areas for conservation, or to shed light on evolutionary processes.

Specifically BAT enables any web user to easily generate maps of species richness, endemism or taxonomic diversity, for a taxonomic group and region of interest, for example, richness of Scarabaeidae (dung beetles) of New Caledonia, or endemism of Australian *Eucalyptus* trees. Users view the results as a map, or can obtain a species list for a particular area.

Software

BAT is a small scale open source application with a simple modular and extensible architecture. It runs as a web server application using a local web mapping service and a standard relational database for storage and analysis. Species occurrence data is retrieved directly from data providers via the DiGIR protocol commonly used in the GBIF network. The data may be routinely retrieved and cached in the database for speed of service.

BAT modules can be grouped into four categories: user interface, mapping, data maintenance and retrieval, and analysis. The components are loosely coupled to allow other developers to add, improve or replace components.

A challenge faced by the project was the implementation of a robust and flexible DiGIR client to request and receive species records. A number of implementations existed for DiGIR providers, but there were no flexible stand-alone DiGIR clients when BAT was developed, so a large part of the project was devoted to writing Net::DiGIR – a Perl module designed for this purpose. DiGIR providers offer query functions which are analogous to a subset of standard SQL. Net::DiGIR takes advantage of this feature by providing a programmer friendly SQL interface, allowing programmers and scientists to focus more on what data they want rather than how to retrieve it. Experiences with various providers lead to the inclusion of tolerance for network failures and for providers with inconsistent implementation of the DiGIR protocol.

It is hoped that the lessons learned from the BAT project will be useful in other projects to develop online analysis tools based on distributed data.