A web service based approach for integrating statistics tools into an information system for experiment data

Dennis Heimann, Jens Nieschulze
Max-Planck-Institute for Biogeochemistry
Jena, Germany
{dheimann, jniesch}@bgc-jena.mpg.de

Birgitta König-Ries
Friedrich-Schiller-Universität Jena
Jena, Germany
birgitta.koenig-ries@uni-jena.de

Data management in the life sciences has evolved from simple storage of data to complex information systems providing additional functionalities like analysis and visualization capabilities, demanding the integration of statistical tools. The Biodiversity Exploratories, a large-scale and long-term biodiversity research project in Germany, are an adequate example in need of such a system. Within this umbrella project, a large number of individual, independent projects from a diverse set of communities investigate different aspects of biodiversity. One of the expectations towards the umbrella project is to make data available beyond individual projects to allow for analysis of data across disciplines and over time. As a technical basis for this task, we develop the web-based Biodiversity Exploratories Information System (BExIS) which offers central storage and management of all project data.

One of the main non-technical challenges faced by any data management system is acceptance by the user community. It is generally acknowledged that such systems should offer added value, so users have a direct benefit of their usage. We believe that seamless access to analysis tools and the ability to plug in new tools as needed is one way to provide such added value. In the long run, seamless integration of statistical methods will also enable common analyses across projects directly within the system. Statistical methods in need range from simple summary of data sets to complex analysis comprising a chain of models and include also graphical analysis. Over the last few years, a number of attempts to solve at least similar requirements have been proposed. These use different approaches to integrate their tools. Some solutions use a generic approach to integrate external tools by using configuration files for providing definitions for tools and data types, and physical descriptions of resource locations. While this approach is very promising, the integration of such a generic workflow environment into a project data management is difficult. In this paper we describe our, more lightweight, approach that combines all statistical methods within one web service. By abstraction from the underlying applications, the web service will be easy to integrate in basically any information system. It provides only three operations to list all available statistical methods, to describe a method more specifically, and to invoke a method. The service architecture as well as the data exchange process between client and service and the adding of analysis applications to the underlying service provider are described. Furthermore a practical example demonstrates the functionality of the service.