

Web services in the contexts of EEA and GBIF

Miruna Badescu

Abstract

The requirements for data collection applications grew significantly over the last few years so nowadays they are oriented towards having the data indexed and catalogued by all kinds of Web applications and engines, regardless on the platform or technology used. Since many systems are reengineered or even replaced after some time, as result of the emerging tools and standards, it is important not to embrace dedicated ways of storing information or communicating with other applications, but to rely on harmonised data description, protocols and data exchanging standards.

1. GBIF Web services

The Global Biodiversity Information Facility (GBIF) is concerned with making primary biodiversity data available through the Web. Further processing, analysis and integration will be possible once the primary point data is query-able in a standardised manor.

The information model encompasses a Registry which holds metadata for GBIF Nodes and the services these expose, through which records of biodiversity data (taxonomic names, specimen observations, links and images about those, etc.) become available for portals, search engines, and other applications. This architecture is implemented according to the Distributed Generic Information Retrieval (DiGIR) protocol that makes location and technical characteristics of the native resource transparent to end users. Closely paired with the services registry, a global indexing component retrieves metadata about the providers' datasets.

Given the different possibilities the data providers have, this software was build using platform-independent software. Following the initial PHP implementation, a data repository tool was developed in Python for the Zope Web and application server, with extendable support for data storage in various formats (XML, plain text files, databases), easy customisation through a friendly Web interface, data validation and logging facilities.¹

¹ Author: Miruna Badescu, Finsiel Romania, <http://www.finsiel.ro>, Ficusului 44A, Bucharest, Romania, email: m.badescu@finsiel.ro, speaking on behalf of GBIF and EEA

2. Reportnet Web services

Reportnet is a suite of IT tools optimised to support the businesses processes of the European Environmental Information System and build a shared information infrastructure. It is imperative that the various components of Reportnet functioning either at European or local level are able to interoperate in a harmonised way in order to streamline the reporting process, which comprises collecting, validating, processing, transmitting, aggregating and disseminating data.

The focus of this presentation will be the interaction between the Data Repositories – built in Python for Zope – and the Generic Data Exchange Modules – Java – along with their communication with the other Reportnet tools. The Data Repository software was reused and adapted in the Python version of GBIF's Data Provider, proving its flexibility and wide area of applicability.

Using simple exchange protocols – HTTP, XML/RPC – and XML-based system of communication, Data Repositories are able to contact the Data Exchange Modules for validation and conversion services, get metadata from the Data Dictionary, query the Reporting Obligations Database, integrate with EIONET LDAP, share and export deliveries, be indexed and searched by Content Registry and Web dashboards, etc. In other words, this component is build with modern Web services technologies which make it fit into the complex Reportnet architecture.

3. GEMET

The GEneral Multilingual Environment Thesaurus (GEMET) was originally developed as an indexing, retrieval and control tool for the Catalogue of Data Sources (CDS) of the European Environmental Agency (EEA). Later, due to its usefulness in facilitating the global dissemination and exchanging of environmental information, the need appeared for its content to be queried by all sorts of applications via HTTP, XML/RPC and SOAP.

Subsets of GEMET are used by networks of portals for defining metadata in a unified way, indexing and searching content or just displaying definitions of environmental terms that appear on Web pages.

Once again, Zope, powered by Python scripting language, was found a suitable platform for interfacing the relational database structure of the multilingual thesaurus and exposing an API to remote applications that need displaying, storing, caching, updating or translating terms and their definitions in various languages. Also, a console for maintaining – add/update/delete – GEMET will be available for non-technical administrators. By calling this API from platforms based on Java, Python, ASP, PHP, etc. via the preferred protocol it will be possible not only to build Web-based interfaces using GEMET's terms and concepts, but also to provide meaningful and accurate results to the search engines.