

eCUDO.pl – towards the Polish Oceanographic Data Committee

Marcin Wichorowski, Institute of Oceanology Polish Academy of Sciences, wichor@iopan.pl

Michał Piotrowski, Maritime Institute in Gdańsk, Michal.Piotrowski@im.gda.pl

Lena Szymanek, National Marine Fisheries Research Institute, Iszymanek@mir.gdynia.pl

Urszula Pączek, Polish Geological Institute National Research Institute, upac@pig.gov.pl

Mirosława Ostrowska, Institute of Oceanology, Polish Academy of Sciences, ostra@iopan.pl

Michał Wójcik, Maritime Institute in Gdańsk, Michal.Wojcik@im.gda.pl

General overview

Institute of Oceanology, Polish Academy of Sciences, Institute of Meteorology and Water Management National Research Institute, Maritime Institute in Gdańsk, Polish Geological Institute National Research Institute, National Marine Fisheries Research Institute and University of Technology in Gdańsk - consortium of organizations engaged in research and exploitation of marine resources, established as POLMAR, for many years lead actions targeting harmonization, integration and coordinated provisioning of environmental data resources. Sharing of data is focusing on transfer of information and enabling access to detailed oceanographic data. This is fundamental activity for planning and performing of activities and investments in coastal and offshore areas. The consortium is leading activities towards deployment operational state of the system delivering these demanded data and products to the users. The consortium consists of most of the organizations involved in marine research and continuous acquisition of oceanographic data. For this reason the consortium is perspective Polish Oceanographic Data Committee.

System design

The project assumes obtaining data openness at the level of 5 stars according to the "5 Star Open Data" scale, because the project develops a system that gives users not only the ability to access current, open and searched data, but also integrates the resources of other open collections. The data will be described with metadata compliant with the standards applicable to geospatial data. The data will also be marked with unique URIs and in all appropriate places the metadata will contain relevant links to other resources available on the network.

The following design and architectural assumptions have been adopted:

- The central element of the platform is infrastructure based on a virtualization server connected to the data storage system. The use of such a solution allows for easy implementation of individual modules, as well as their replacement in the event of a need to conduct a complex software update or to remove errors in the operation of the system.
- The platform's infrastructure consists of a main server that provides required functionalities to external users, a single-signal server allowing one-time login to the platform and use of all its resources, spatial data server performing WMS, WFS and CSW services and a server processing data from non-federated external systems .
- The platform's environment (users, federated systems and non-federated external systems) can be distributed to any geographical location with communication via the Internet. For the sake of simplifying the model, it is assumed that each external system (both federated and un-federated) will consist of a server providing data, combined with a data warehouse.
- Users can access the system via any device (computer, laptop, tablet, mobile phone) with installed software supporting the display of websites using the http protocol or supporting WFS, WMS and CSW type services.

- All virtualized system components work on the CentOS operating system.
- All applications running under the virtualization server operate as part of the virtual Java machine provided by the Java runtime environment for the CentOS system. In the case of web applications and enterprise-class applications, they will be launched as part of the WildFly application server running in the Java virtual machine.
- The PostgreSQL server is used as the database server with the PostGIS extension providing support for geographic data processing (GDAL).
- The central element is eCUDOApp, an enterprise class application consisting of the EJB eCUDOEngine module and the eCUDOWeb web module communicating with each other using local EJB interfaces. All operations on the database will be performed from the level of the EJB module through the JDBC protocol (Java Data Base Connectivity).
- A GeoServer application embedded in a Jetty container running in a Java virtual machine is used as a GIS data server. Both the GeoServer application and the Jetty container are free software available under the Open Source license. The application through the eCUDOWeb module will issue WMS, WFS and CSW services in the Representational State Transfer (REST) architecture. As in the case of eCUDOEngine, communication with the database will be implemented through the JDBC protocol.
- The eCUDOLinkedDataParser application is launched directly in the Java virtual machine and in specified time intervals it will synchronize metadata from non-federated external systems using the http protocol (Hypertext Transfer Protocol) with metadata on the eCUDO platform via the eCUDOWeb module using the published services in REST and SOAP architecture (Simple Object Access Protocol). The application will be able to process data provided by external systems in the Linked Data format.
- The SSO server will contain the Keycloak web application acting as a central login point that allows the use of all platform elements after a single login. Keycloak is a free application available under the Open Source license. Login will be possible via the website by the http protocol as well as through authorization connected to REST and SOAP service calls.

The above elements, the CentOS system, the WildFly application server and the PostgreSQL database server together with the PostGIS extension, are free software available under the Open Source license.

The data presented by the eCUDOWeb module will come from external systems, especially from federated systems. Each of the federated systems will already have an enterprise application for storing and sharing data. A module or a separate application serving as a proxy for the central eCUDO system should be prepared for each of these systems.

Summary

The eCUDO.pl project, when finished is intended to be well integrated with SeaDataCloud and EMODNet systems. Advanced services provided by clients (including data analysis services) extend availability of oceanographic data both to Polish and European organisations.