

GEOSPATIAL DATA SHARING, ONLINE SPATIAL ANALYSIS AND PROCESSING OF INDIAN BIODIVERSITY DATA IN INTERNET GIS DOMAIN- A CASE STUDY FOR RASTER BASED ONLINE GEO-PROCESSING

Harish Karnatak*, Kamal Pandey, Kapil Oberai, Arijit Roy, Deepti Joshi, Hariom Singh, PLN Raju and YVN Krishna Murthy
Indian Institute of Remote Sensing, Indian Space Research Organization, Department of Space, Govt. of India, Dehradun, harish@iirs.gov.in

Commission VIII, WG VI/4

KEYWORDS: Biodiversity, Services Oriented Architecture, OGC web services, GDAL/OGR, and Geo-processing

ABSTRACT:

National Biodiversity Characterization at Landscape Level, a project jointly sponsored by Department of Biotechnology and Department of Space, was implemented to identify and map the potential biodiversity rich areas in India. This project has generated spatial information at three levels viz. Satellite based primary information (Vegetation Type map, spatial locations of road & village, Fire occurrence); geospatially derived or modelled information (Disturbance Index, Fragmentation, Biological Richness) and geospatially referenced field samples plots. The study provides information of high disturbance and high biological richness areas suggesting future management strategies and formulating action plans. The study has generated for the first time baseline database in India which will be a valuable input towards climate change study in the Indian Subcontinent.

The spatial data generated during the study is organized as central data repository in Geo-RDBMS environment using PostgreSQL and POSTGIS. The raster and vector data is published as OGC WMS and WFS standard for development of web base geo-information system using Service Oriented Architecture (SOA). The WMS and WFS based system allows geo-visualization, online query and map outputs generation based on user request and response. This is a typical mashup architecture based geo-information system which allows access to remote web services like ISRO Bhuvan, Openstreet map, Google map etc., with overlay on Biodiversity data for effective study on Bio-resources.

The spatial queries and analysis with vector data is achieved through SQL queries on POSTGIS and WFS-T operations. But the most important challenge is to develop a system for online raster based geo-spatial analysis and processing based on user defined Area of Interest (AOI) for large raster data sets. The map data of this study contains approximately 20 GB of size for each data layer which are five in number. An attempt has been to develop system using python, PostGIS and PHP for raster data analysis over the web for Biodiversity conservation and prioritization. The developed system takes inputs from users as WKT, Openlayer based Polygon geometry and Shape file upload as AOI to perform raster based operation using Python and GDAL/OGR. The intermediate products are stored in temporary files and tables which generate XML outputs for web representation. The raster operations like clip-zip-ship, class wise area statistics, single to multi-layer operations, diagrammatic representation and other geo-statistical analysis are performed. This is indigenous geospatial data processing engine developed using Open system architecture for spatial analysis of Biodiversity data sets in Internet GIS environment. The performance of this applications in multi-user environment like Internet domain is another challenging task which is addressed by fine tuning the source code, server hardening, spatial indexing and running the process in load balance mode. The developed system is hosted in Internet domain (<http://bis.iirs.gov.in>) for user access.

1. INTRODUCTION

1.1 Biodiversity Characterization at landscape level

National Biodiversity Characterization at Landscape Level, a project jointly sponsored by Department of Biotechnology and Department of Space, was implemented to identify and map the potential biodiversity rich areas in India (Roy et al, 2012). The project has contributed to the scientific understanding, characterizing and deciphering spatial patterns of Indian natural vegetated ecosystems, their disturbance regimes and biological richness. It provides spatial information on 120 vegetation types consisting of natural, semi-natural and managed formations. The database preparation was supported with phyto-sociological data collected from geospatially located ~16,000 field sample plots for ~7500 plant species (Roy et al, 2012). The geospatial analysis of the data was done for different landscape metrics in conjunction with disturbance, ecological uniqueness, species diversity and economic value of different forest types assessed using field plot data. The analysis resulted in geospatial products providing spatial patterns on disturbance and biological richness facilitating conservation prioritization.

According to Roy et al, 2012, the project has resulted in the generation of information with the following end-use:

- Digital database on vegetation type distribution, the first of its kind of systematically organized databases developed in India, a basic input for identifying the species habitats and would serve as benchmark for further biodiversity related ecological studies.
- Disturbance Regimes assessed across the landscape flag as the 'stressed' eco-systems and may highlight the causative factors.
- Biological Richness maps (BR) lay emphasis on the areas, which should be treated on priority while formulating the strategies for conservation of biodiversity.
- National inventory of 7500 unique plant species recorded during field sampling.

1.2 Biodiversity Information System

The spatial data generated in national project on Biodiversity Characterization at landscape level is organized as central data repository and published as Web GIS based Biodiversity Information System (BIS) (<http://bis.iirs.gov.in>). The BIS is developed based on Service Oriented Architecture (SOA) where OGC web services are published for Geo-visualization and spatial queries in Web GIS environment. The WMS and WFS based system allows geo-visualization, online query and map outputs generation based on user request and response. This is a typical mashup architecture based geo-information system which allows access to remote web services like ISRO Bhuvan, Openstreet map, Google map etc., with overlay on Biodiversity data for effective study on Bio-resources.

The BIS Geo-portal provides various GIS based services for Biodiversity data sets. It has major two components for spatial data sets and information services i.e. Biodiversity spatial viewer and Data download facility. These GIS viewers are developed using OpenLayer API and PHP. The rich GUI is developed using GeoEXT. The major features of BIS include:

- Adoption of open system architecture for development and deployment of the web services and database management.
- The national data spatial and non-spatial is organized as a central repository for biodiversity of India.
- The data download utility using online clip-ship-zip facility. The user can define AOI and can download original GIS layer of vegetation type, Disturbance Index, Fragmentation and Biological Richness map.
- Biodiversity Spatial Viewer for advance online GIS utilities like: effective geo-visualization using map cache technique, overlay of Bhuvan satellite imagery (56mt to 5.8 mt), connection to remote GIS servers, adding user defined servers, basic GIS navigation tools, area and distance measurement, spatial filter, overlay of sample plot grid data (with advanced filtering utility), layer swapping for effective visualization, layer transparency, search places and search species etc.

2. LITRATURE REVIEW

2.1 Online Geo-processing and Web service standards

Geoprocessing is a GIS operation used to manipulate spatial data. A typical geoprocessing operation takes an input dataset, performs an operation on that dataset, and returns the result of the operation as an output. For online geospatial data processing, Web Processing Service specification was released in 2005 by Open Geospatial Consortium, that incorporate complex spatial process through a standardized service interface based on the Hypertext Transfer Protocol (HTTP) (Foerster and Stoter, 2006). Karnatak et al., 2007 has developed Web GIS based multicriteria decision analysis system for biodiversity conservation and prioritization. The most popular multicriteria technique AHP is used to select optimal alternative. The developed SDSS is a typical example of web based GIS data processing for decision analysis.

Some notable and updated OGC WPS specification implementations are PyWPS, a project with GRASS-based processing and R (Geostatistical computing) available to web client (Cepicky and Becchi, 2007), 52 North WPS, ZOO open WPS platform (Fenoy et al., 2013), the Deegree project is a Java based framework for processing services and other striking commercial solutions have been implemented such as ESRI

ArcGIS Server, ERDAS APOLLO 2010 etc (Evangelidis et al., 2014). All the above said frameworks are based on SOA implementation in which the developed services can be accessed over a network and can be combined and reused in web application. These services used for communication with each other by exchanging data in a shared and well defined format or managing activities between two or more services (Bruin et al., 2014).

The WPS specification defines a mechanism by which a client may submit a processing task to a server to be completed. The service defines a server instance, or server, as an entity which may provide one or more processes, or individual processing tasks (e.g., adding two raster datasets together could be one process) (Christopher et al., 2009). WPS basic properties are: 1) Inputs can be web-accessible URLs or embedded in the request. 2) Outputs can be stored as web-accessible URLs or embedded in the response. It supports multiple inputs and output formats (Wikipedia WPS, 2013). It is a service on the web following a SOA architecture where pre-programmed geospatial computation models are available over a network. The SOA model permits processes to be available in a transparent way, independent of programming language and operating system. The services describes a mechanism by which a client computer may submit a job to be processed on a server computer, using data provided as a WFS (Vector) or WCS (Raster). It is client/server architecture, meaning that both a client component and a server component are mandatory (Anthony et al., 2013 ; Dubois et al., 2013) For implementation and testing purposes it is useful to build the client-side application on a GIS to take advantage of existing services or develop their own services (Christopher et al., 2009).

Some important works have been done by using online geoprocessing such as the shortest path between two points (PML: Shortest path), systematic planning and cultivation of agricultural fields (Bruin et al., 2014), ecological modeling and forecasting, forecast the impact of climate change on protected areas (Dubois et al., 2013) ,process geospatial data such as NDVI calculation on different computing backend (Giuliani et al., 2012), Geostatistical computing and INTAMAP: The design and implementation of an interoperable automated interpolation web service (Pebesma et al., 2011), thematic map generation (Rautenbach et al., 2013) and Forecasting biomes of protected areas (Skoien et al.,2011) etc.

3. APPROACH

The Geospatial data sharing, online spatial analysis and processing of Indian Biodiversity data in web GIS environment is achieved based on SOA architecture. The web service specifications from OGC are implemented for centralized data repository. The spatial queries and analysis with vector data is achieved through Structured Query Language (SQL) using POSTGIS and WFS-T operations. The vector data stored in PostgreSQL is accessed through POSTGIS and presented as WMS and WFS using Geoserver. But the development of geo-processing engine for raster data analysis in Web GIS environment is a major objective of this study. The methodology adopted in the study is shown in Figure 1.

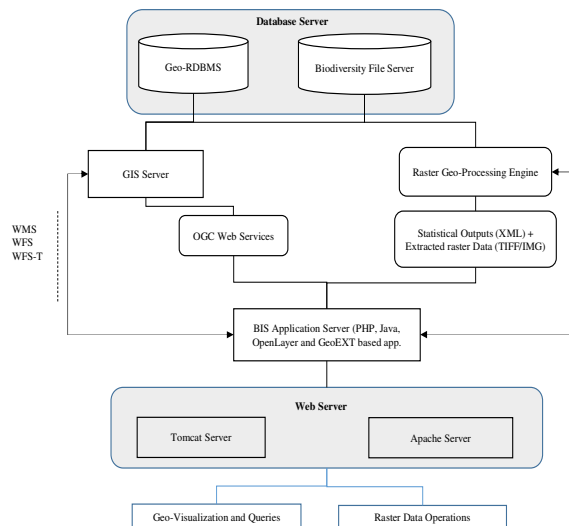


Figure 1. Methodology of the study

The complete information flow from user input to presenting output is shown in Figure 2.

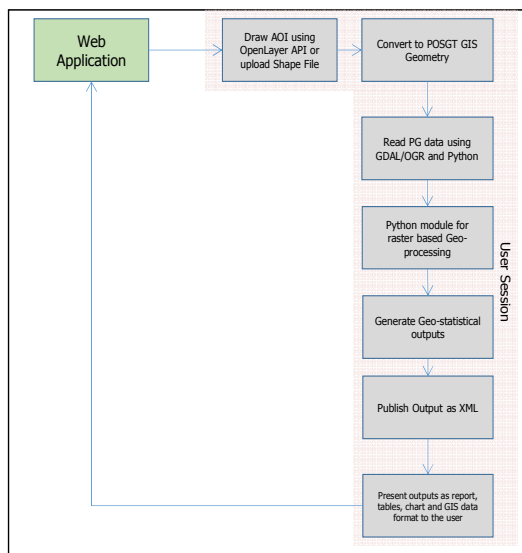


Figure 2. Information Flow

The study is based on three basic principles i.e. users input as WKT and Shape File, GIS raster operations using GDAL/OGR and presenting outputs as XML. The major steps involves:

3.1 Web Application:

The web application is developed using PHP, OpenLayer API and GeoEXT. The web application has two components for GIS related activities i.e. Biodiversity spatial viewer and Data download facility. The biodiversity spatial viewer provides WMS and WFS based web services, GIS data visualizations, basic and advance queries with spatial filter (<http://bis.iirs.gov.in/maps.php>).

3.2 User's inputs:

This module is developed under data download category. The user provides its inputs either as simple drawing over available map viewer for Area of Interest (AOI) or uploading the boundary of study area as simple Shape file. The system creates

a user session for each request which will be live till end of the process.

3.3 Convert to POSTGIS Geometry:

The user's inputs are converted as PostGIS geometry and stored as a table in PostgreSQL. This will be the inputs to geo-processing engine of the study.

3.4 Python module for raster based Geo-processing:

During this study a Python based software product for raster data analysis is developed which is core of the geo-processing engine. The GDAL/OGR library for python is used for reading and writing of raster and vector data.

3.5 Generate Geo-statistical outputs as XML:

The geo-processing engine developed for this study provides the facility to do geo-statistical analysis in multi-user environment. The intermediate products are stored in temporary files and tables which generate XML outputs for web representation. The raster operations like clip-zip-ship, class wise area statistics, single to multi-layer operations, diagrammatic representation and other geo-statistical analysis are performed. This is indigenous geospatial data processing engine developed using Open system architecture for spatial analysis of Biodiversity data sets in Internet GIS environment.

3.6 Present outputs as report, tables, chart and GIS data format to the user:

The module generates the outputs in the form of HTML reports, tables, charts and subset of raster data as .tiff and .img files.

4. RESULTS AND DISCUSSION

The geo-spatial data generated under national level project on Biodiversity Characterization at landscape level using RS&GIS is disseminated through web GIS based Biodiversity Information System (<http://bis.iirs.gov.in>). The system provides various facilities to access and analysed the biodiversity data for effective conservation planning. The system provides data visualization and query system under Biodiversity spatial viewer (Figure 3)



Figure 3- Biodiversity Spatial viewer.

The Biodiversity spatial viewer provides access to GIS layers from central and distributed servers. For example the satellite data from ISRO Bhuvan portal as WMS service is accessible. The basic layers are available in layer tree and additional layers can be added by the user. WFS based query builder is developed for sample plot data set which allows to perform various queries on vector data sets in web browser environment (Figure 4). The Layer swiping tool for WMS is also developed as part of geo-

visualization. The implementation for mashup based architecture for geo-information services provides unique capability to accessing multiple data and information services from remote servers. This also allows an interoperable solution for information overlay in GIS environment.

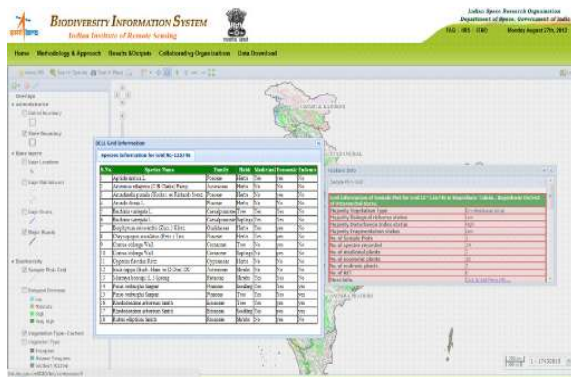


Figure 4- Species queries and geo-tagging

The data download section of BIS portal provides various raster based analysis tools based on geoprocessing engine developed during this study. The major objective is to perform clip-ship-zip operation for sharing of original raster data based on user defined AOI (Figure 5).

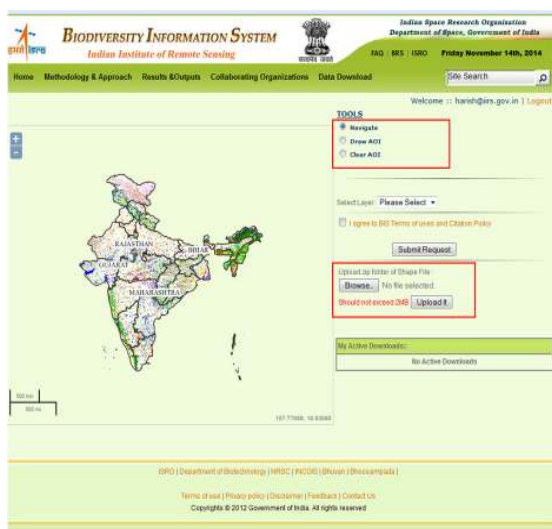


Figure 5- Data download and analysis

The user inputs are accepted as user defined drawing in map viewer or existing shape file. The AOI and shape file are uploaded to the server and converted as PostGIS geometry. Further raster analyses are performed based on this geometry using GDAL/OGR library for Python. The map data of this study contains approximately 20 GB of size for each data layer which are five in number. An attempt has been made to develop system using python, PostGIS and PHP for raster data analysis over the web for Biodiversity conservation and prioritization. The intermediate products are stored in temporary files and tables which generate XML outputs for web representation. The raster operations like clip-ship-zip, class wise area statistics, single to multi-layer operations, diagrammatic representation and other geo-statistical analysis are performed (Figure 6).

The approach adopted in this study for raster based GIS data analysis and processing in Web GIS environment is important for many studies related to raster operations. Similar operations using PostGIS for Vector data sets are possible by performing overlay operation using PostGIS operators. However for raster data sets the RDBMS approach is not very effective and is still in development stage. The performance of raster operations in Web GIS environment is a critical and important aspect for success of the system. In this study the performance enhancement is achieved by splitting processes as multiple threads. The presentation of outputs is purely based on XML standard which is an interoperable approach applicable in any application environment. One of the important challenge in file based raster operation is to maintain user sessions for web application. The temporary files for each session are stored in server and presented to the user. Once user session is expired, all the temporary data is automatically deleted from the server. In this scenario the availability of sufficient space at server end is also important.

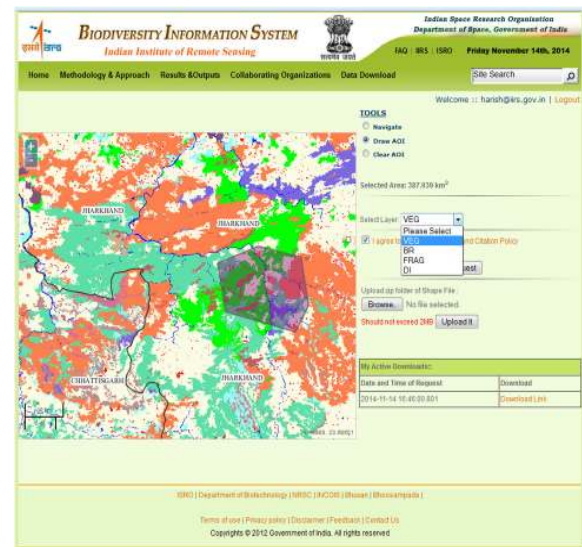


Figure 6- Raster analysis and data download facility

The system is developed using open source GIS solution as open system architecture. The performance of the system is

tested for 100 concurrent virtual users performing raster operations for ~10,000 square km for different geographical locations. In LAN environment (100 Mbps) the output was generated within 12 second and in internet environment (10 Mbps) the output was generated within 20 second.

5. CONCLUSION

The geo-spatial data generated under national level project on Biodiversity Characterization at landscape level using RS&GIS for Indian is disseminated through Web GIS based Biodiversity Information System (<http://bis.iirs.gov.in>). The BIS portal provides geo-visualization with various basics and advanced tools, data download and analysis for raster data using clip-ship-zip operation as open system architecture. The study demonstrates use of Python, GDAL/OGR and PostGIS for online raster based geo-processing in multi-user environment. The same approach can be used for development of online spatial analysis and modelling system for any theme. Further, the performance of the system can be improved by including cluster computing environment like HPC for larger areas.

ACKNOWLEDGEMENT

Authors express their gratitude to Chairman, Indian Space Research Organisations (ISRO) for motivation and encouragements. Thanks to Dr. P.S. Roy Former Director IIRS, Dr. SPS Kushwaha Head FED and GD PPEG, Dr. Sarnam Singh PC CSSTEAP for their guidance and support. Authors also express their thanks to Dr. S.K. Srivastav Head Geoinformatics Department IIRS for his critical observations and suggestions.

REFERENCES

- Anthony, M., Castronova, Jonathan, L., Goodall, Mostafa, M. and Elag. (2013). Model as web services using the Open Geospatial Consortium (OGC) Web Processing Service (WPS) standard. *Environmental Modelling & Software*, Vol. 41, pp.72-83.
- Bruin, S.D., Lerink, P., La Riviere, I.J. and Vanmeulebrouk, B. (2014). Systematic planning and cultivation of agricultural fields using a geo-spatial arable field optimization services: Opportunities and obstacles. *Biosystems engineering*, I 20, pp.15-24.
- Cepicky, J. and Becchi, L. (2007). Geospatial processing via Internet on remote servers–PyWPS. *OSGeo Journal*, Vol. 1, pp. 39-42. <http://osgeo.org/journal>. (Accessed 17.01.14).
- Christopher D. Michaelis and Daniel P. A. (2009). Evaluation and Implementation of the OGC Web Processing Service for Use in Client-Side GIS. *Geoinformatica*, Vol. 13, pp.109–120.
- Dubois, G., Schulz, M., Skoien, J., Bastin, L. and Peedell, S. (2013). eHabitat, a multi-purpose Web Processing Services for ecological modeling. *Environmental Modelling & Software*, Vol. 41, pp.123-133.
- Evangelidis, K., Ntouros, K. and Makridis, S. (2014). Geospatial services in the cloud. *Computers & Geosciences*, Vol. 63, pp.116-122.
- Fenoy, G., Bozon, R. and Raghavan, V. (2013). ZOO-project: The open WPS platform. *Applied Geomatics*, Vol. 5, pp.19-24.
- Foerster, T. and Stoter, J. (2006). Establishing an OGC Web Processing Service for generalization processes. In: Workshop of the ICA Commission on Map Generalisation and Multiple Representation.
- Giuliani, G., Nativi, S., Lehmann, A. and Ray, N. (2012). WPS mediation: An approach to process geospatial data on different computing backends. *Computers & Geosciences*, Vol. 47, pp.20-33
- Karnatak Harish Chandra, Sameer Saran, Karamjit Bhatia and P.S. Roy, (2007) "Multicriteria Decision Analysis in web GIS environment", *Geoinformatica* (2007) 11: pp 407-429: DOI 10.1007/s10707-006-0014-8.
- Pebesma, E., et al. (2011). INTAMAP: The design and implementation of an interoperable automated interpolation web service. *Computer & Geosciences*, Vol. 3, pp. 343-352.
- Roy PS, Karnatak Harish, Kushwaha SPS, Roy A, Saran S (2012), " India's plant diversity database at landscape level on geospatial platform: prospects and utility in today's changing climate", *Current Science*, Vol 102, Issue 8, pp 1136-1142.
- Rautenbach, V., Coetzee, S. and Iwaniak, A. (2013). Orchestrating OGC web services to produce thematic maps in a spatial information infrastructure. *Computers, Environment and Urban Systems*, Vol. 37, pp.107-120.
- Skoien, J.O., Dubois, G. and Jesus, J.D. (2011). Forecasting biomes of protected areas, *Procedia Environmental Sciences*, Vol. 4, pp. 44-49.
- Wikipedia, WPS,
http://en.wikipedia.org/wiki/Web_Processing_Service
(Accessed, 12.04.14)