THE GEOBRAIN DATA-INTENSIVE ON-LINE ENVIRONMENT FOR ENABLING EASY ACCESS AND INTEGRATION OF DISTRIBUTED GEOSPATIAL RESOURCES

Meixia Deng (mdeng@gmu.edu)

Abstract

Access and integration of distributed geospatial resources (e.g. data, services, and computing resources) plays critical roles in solving many science and application problems today. There are many difficulties and challenges in access and integration of distributed geospatial resources due to the genetic complexities involved in the process. For example, integration of geospatial data from multiple sources need many efforts even for expert data users and almost impossible for non-expert users since it requires substantial computing resources, adequate data preprocessing skills such as subsetting, reformatting, reprojecting and resampling, and thorough understanding about the data syntax and semantics. Enabling easy access and integration of distributed geospatial resources with advanced information and computing technologies will help science and application studies significantly in many ways, e.g. time and efforts can be saved largely by concentrating on solving scientific problems instead of tedious data acquiring, preprocessing and integration, individuals who lack adequate computing resources or data processing skills may still conduct desired scientific researches, and some tasks needing (near) real-time analysis or quick response may be achieved in an efficient way. Many researches have been conducted and significant progresses have been made for the purpose. This paper addresses a data-intensive online environment, established by the GeoBrain project (funded by NASA), for enabling easy access and integration of distributed geospatial resources. Based on latest Web service, interoperability and distributed computing technologies, this environment implements innovative capabilities in providing easy, open and seamless discovery, access, integration and analysis of geospatial data, information, services, and models through any Internet connected computers. Specifically, the GeoBrain data-intensive online environment is powered by interoperable, personalized, and on-demand data access services (IPODAS), catalog federation service, large amounts of geospatial Web services, and workflow-based data and information infusion technologies. Users are able to use intuitive Web portals provided by the environment to get data of interest from multiple resources (e.g. distributed NASA online data pools searched through ECHO), process them into desired forms automatically with on-the-fly subsetting, reformatting, reprojecting, resampling and integrating services, and perform further analysis on any data or information products acquired through large amounts of geospatial Web services. This will have some significant impact on how scientific education and research can be conducted since any person who has an Internet-connected desktop/laptop computer can access vast amount of geospatial resources that used to be only available to a few scientists or researchers.