Harvesting Metadata from THREDDS and OPeNDAP

Valentijn V. (1), Skidmore A.K. (1), Lemmens R. (1), Toxopeus A. (1), Retsios B. (1), Davis E.R. (2), de Bie C.J.A.M. (1), and Nieuwenhuis W. (1)

(1) ITC, NRS, Enschede, Netherlands (venus@itc.nl), (2) UCAR Unidata Program Center, P.O. Box 3000 Boulder, United States of America

Abstract

Large and time varying data collections are cumbersome to store, analyze, and view, especially when using traditional hierarchical structures. In this paper, we present enhancements to a framework to bridge the gap between data providers and data users, with the aim to make scientific data 'discoverable' and 'usable' in a variety of models and in a distributed data setting. Within this context a number of additional applications and services are developed and demonstrated.

Traditionally, catalogs are used to communicate the data content description (metadata) of a data set to users. However, a catalog is also the foundation on which additional layers of services can be built. For instance, the combination of a GIS storage system and interoperable software services can supply external discovery systems with needed information, and they can supply information to improve data display and analysis of geo-location information.

A new HarvestThredds service is demonstrated, offering capability to "harvest" metadata from a UNIDATA THREDDS catalog server, and use the 'harvest' on another server or for another purpose. This new service is fully compliant with the Open Geospatial Consortium (OGC, http://www.opengeospatial.org/) standards for storing metadata (ISO19115 and ISO 19139). This helps to overcome interoperability issues between the Earth/Atmosphere Sciences Information Community and the GIS (Geospatial) Information Community who, because of their diverse backgrounds, tend to use different server technologies for generating and publishing data catalogs, leading to data 'isolation' and 'redundancy'.

Another new service, entitled CrawlableDatasetDods, finds datasets on remote data servers from which to build catalogs and aggregations of data sources. This ability to crawl, catalog, and aggregate data (i.e. satellite images) despite being stored on a remote server, helps to restore a sense of "ownership". This aids in the user-adaptation of technologies for distributed scientific data access and also eases the maintenance of distributed data bases that show high dynamics over time.