Geophysical Research Abstracts
Vol. 13, EGU2011-3356, 2011
EGU General Assembly 2011

Extensions to the GI-cat geospatial discovery broker

Enrico Boldrini, Fabrizio Papeschi, Alessio Baldini, and Valerio Angelini
Italian National Research Council (CNR), Institute of Methodologies for Environmental Analysis (IMAA), Potenza, Italy
{boldrini, papeschi, baldini, angelini}@imaa.cnr.it

GI-cat implements a discovery broker making possible to query heterogeneous data sources through a common discovery interface; both distributed queries and harvesting approaches are implemented.

GI-cat realizes a flexible framework to interconnect heterogeneous resources (i.e. data repositories and services) by means of a mediation and adaptation approach.
Indeed, for each resource type, the mediation functionalities are implemented by a specific software component called "Accessor". This approach allows to interconnect, in a loosely-coupled way, existing and even future resources.

Currently GI-cat federates several resources that are compliant with international and community standards. They include: OGC Web Services (e.g. CSW, WCS, WMS, WFS, WPS), THREDDS Data Service, GBIF services, SeaDataNet CDI, GeoRSS, OpenSearch, OAI-PMH. The federated resources can be queried through one of the discovery interfaces published by the broker. Significant SOA and Web 2.0 standard interfaces are supported, such as OGC Catalogue Service for the Web (CSW) in its ISO and ebRIM profiles, OpenSearch, OAI-PMH.

Several extensions and optimizations have been lately developed for the broker. New Accessors were added to federate new services and data types.

A new Accessor was implemented for THREDDS-ncISO services: the THREDDS Data Service (TDS) allows to inventory on a web service large collections of netCDF files (other formats are as well supported); ncISO is a TDS plugin that enrich the TDS information content by generating also ISO 19115-2 metadata from the underlying netCDF datasets.

A new Accessor was also implemented to enable queries to local file systems; thus, the broker is able to access local file collections by harvesting metadata elements from XML documents, like: Dublin Core, ISO 19139 (encoding of ISO 19115 and ISO 19115-2 data models), Directory Interchange Format (DIF), SeaDataNet CDI, ncML (CF and OD attribute conventions). Binary files are also supported, such as the NetCDF.

Moreover, it was accomplished a significant optimization to the harvesting technology. A test with large THREDDS catalogs (belonging to the Unidata Motherlode service) achieved the harvesting and querying of about 400,000 records. The optimizations involved the creation of database indexes and more efficient XPath queries, as well as speed increase in XML formatting of the results.

Web 2.0 technologies have also been experimented, providing the GI-cat framework with an internal HTTP server publishing two main web applications named GI-portal and GI-conf, accessible through common browsers. GI-portal allows to query the broker by simple constraints (e.g. area, time, free text); pagination of results and overlay on a map are also available. GI-conf allows to manage the currently federated resources and published interfaces, as well as customize the broker web portal and configure advanced settings (e.g. proxy parameters, authentication) to tailor the framework according to the desired use scenario.

GI-cat framework has been developed and deployed in the context of several international and national initiatives and projects, amongst them European Commission (EC) SeaDataNet, GENESI-DR and EuroGEOSS projects, ESA Heterogeneous Missions Accessibility Testbed (HMA-T), the CNR intiative GIIDA, GEOSS Interoperability Process Pilot Project (IP3) and Architectural Implementation Project Phase 2 (AIP-2).