



GENESI-DR Portal: a scientific gateway to distributed repositories

Pedro Goncalves (1), Fabrice Brito (1), Fabio D' Andria (1), Roberto Cossu (2), and Luigi Fusco (2)

(1) Terradue SRL, Roma, Italy, (2) ESA-ESRIN, EOP-S, Frascati (RM), Italy

GENESI-DR (Ground European Network for Earth Science Interoperations - Digital Repositories) is a European Commission (EC)-funded project, kicked-off early 2008 lead by ESA; partners include Space Agencies (DLR, ASI, CNES), both space and no-space data providers such as ENEA (I), Infoterra (UK), K-SAT (N), NILU (N), JRC (EU) and industry as Elsas Datamat (I), CS (F) and TERRADUE (I). GENESI-DR intends to meet the challenge of facilitating "time to science" from different Earth Science disciplines in discovery, access and use (combining, integrating, processing, ...) of historical and recent Earth-related data from space, airborne and in-situ sensors, which are archived in large distributed repositories.

"Discovering" which data are available on a "geospatial web" is one of the main challenges ES scientists have to face today. Some well-known data sets are referred to in many places, available from many sources. For core information with a common purpose many copies are distributed, e.g., VMap0, Landsat, and SRTM. Other data sets in low or local demand may only be found in a few places and niche communities. Relevant services, results of analysis, applications and tools are accessible in a very scattered and uncoordinated way, often through individual initiatives from Earth Observation mission operators, scientific institutes dealing with ground measurements, service companies or data catalogues. In the discourse of Spatial Data Infrastructures, there are "catalogue services" - directories containing information on where spatial data and services can be found. For metadata "records" describing spatial data and services, there are "registries". The Geospatial industry coins specifications for search interfaces, where it might do better to reach out to other information retrieval and Internet communities.

These considerations are the basis for the GENESI-DR scientific portal, which adopts a simple model allowing the geo-spatial classification and discovery of information as a loosely connected federation of nodes. This network had however to be resilient to node failures and able to scale with the growing addition of new information about data and services. The GENESI-DR scientific portal is still evolving as the project deploys the different components amongst the different partners, but the aim is to provide the connection to information, establish rights, access it and in some cases apply algorithms using the computer power available on the infrastructure with simple interfaces. As information is discovered in the network, it can be further exploited, filtered or enhanced according to the user goals.

To implement this vision two specialized graphical interfaces were designed on the portal. The first, concentrates on the text-based search of information, while the second is a command and control of submission and order status on a distributed processing environment. The text search uses natural language features that extract the spatial temporal components from the user query. This is then propagated to the nodes by mapping them to OpenSearch extensions, and then returned to the user as an aggregated list of the resources. These can either be access points to dataset series or services that can be further analysed and processed. At this stage, the user is presented with dedicated interfaces that correspond to context of the action that is performing. Be it a bulk data download, data processing or data mining, the different services offer specialized interfaces that are integrated on the portal.

In the overall, the GENESI-DR project identifies best practices and supporting context for the use of a minimal abstract model to loosely connect a federation of Digital Repositories. Surpassing the apparent lack of cost effectiveness of the Spatial Data Infrastructures effort in developing "catalogue services" is achieved by trimming the use cases to the most common and relevant. The GENESI-DR scientific portal is, as such, the visible front-end of a dedicated infrastructure providing transparent access to information and allowing Earth Science

communities to easily and quickly derive objective information and share knowledge based on all environmentally sensitive domains.