Ground European Network for Earth Science Interoperations

GENESI-DR

Digital Repositories

Issue 4. April 2010.

In this issue:

The GENESI-DR consortium informs its users that the infrastructure created during the GENESI-DR project will remain operational at the end of the project.

INFRA-2007-1.2.1 : Scientific Digital Repositories

Duration : January 1, 2008 – April 30, 2010

Contract n. 212073

Total EC funding : 4.4 M€

SEVENTH FRAMEWORK PROGRAMME

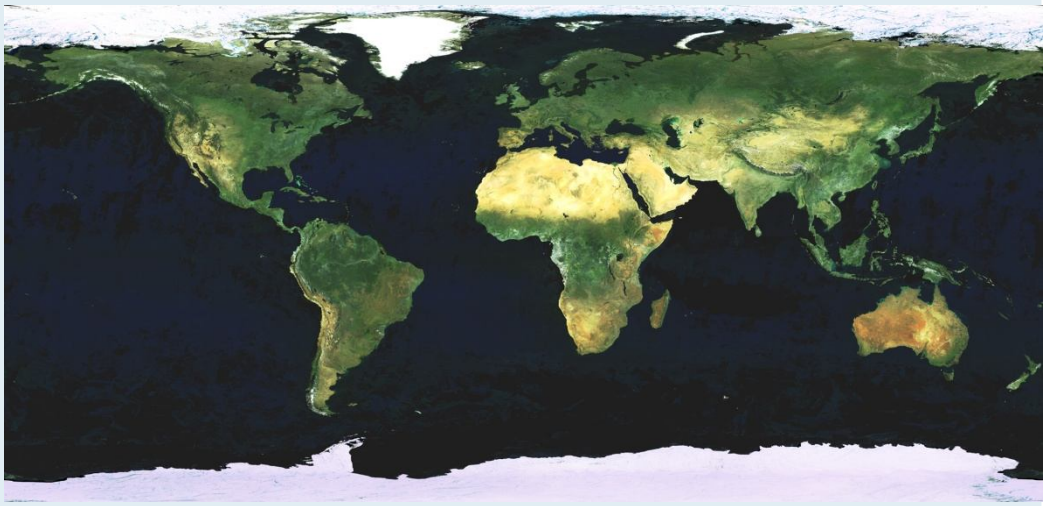
Using GENESI-DR
How do I add my data to the system?

Over the course of the GENESI-DR project we have seen the creation of a new data infrastructure providing search and retrieval of Earth science data held in diverse digital repositories across Europe. As the project draws to a close, this newsletter will provide further information on how this infrastructure, which will remain in place, can be used by the science community. This, either from the point of view of the data user who is looking for data, or from that of the data provider, looking to achieve broader use of their data.

How do I GENESI-fy my digital repository?

The GENESI-fication of a digital repository is quite simple: a few steps to follow and you're done! This short decription will walk you through the installation and configuration of a GENESI-DR Digital Repository. It uses as an example a real-life scenario: the GENESI-fication of the ESA MERIS Level 3 products.

The MERIS Level 3 products are a set of 17 different monthly aggregated products covering land (e.g. MERIS Global Vegetation Index), atmosphere (e.g. water vapour) and water (e.g. Algal-1) and are publicly available on a ESA web site. Each product is available in netCDF and HDF file formats, have a quick-look and an XML file containing the metatada associated. These are all published on a WMS server.



Let's start by the GENESI-fication software checklist; you'll need to install and configure:

- The GENESI-DR Discovery Agent is responsible for retrieving, storing and publishing the Digital Repository data locations and metadata obtained from the RDF files (these are XML files containing a number of metadata elements).

• The GENESI-DR Catalogue is a PostgreSQL/PostGIS database with a schema replicating the GENESI-DR core metadata schema (this schema can be possibly extended by the DR owner) for the series and datasets.
• The GENESI-DR Catalogue Access Service (CAS) is a mono C# Web service gateway allowing external clients to query (through the OpenSearch protocol) and publish metadata values and data resources locations held in the GENESI-DR Catalogue.
• The GENESI-DR Data Access Services are the set of dataset access protocols (HTTP(s), FTP(s), GRIDFTP, OGC WMS/WCS/WFS) that clients and applications will use to consume the data. These also expose the GENESI-DR Auth/N and Auth/Z security mechanisms.

Let's go back to our MERIS Level 3 products. The first step is to generate the RDF files from the XML metadata. Although all the required metadata in contained in the XML files, we need to transform it into RDF files. To achieve this a simple XSLT file can be used. These RDF files are then passed to the GENESI-DR Discovery Agent that ingests them into the GENESI-DR Catalogue and finally exposed by the GENESI-DR Catalogue Access Service.
The MERIS Level 3 products are now discoverable and ready to be used in Earth science applications!
More information can be found in GENESI-DR Wiki and in the GENESI-DR ESDR-Minimum-Requirements-and-GENESI-fication Guide. ***http://www.genesi-dr.eu***

Using a "Virtual Organisation" to manage user rights

GENESI-DR community is made up of digital repositories (DR) of Earth Science data, Earth scientists, users of this data, and processing services.

Each of the digital repositories that make up GENESI-DR can have their own data policy, with only a subset of users interested in consulting data from this DR.

GENESI-DR must therefore provide a mechanism to differentiate between the rights and responsibilities of each user and to associate restrictions to be applied to a given DR.

Using the Virtual Organisations concept

Continued from page 2

GENESI-DR answer these needs by implementing the concept of a "Virtual Organization" (VO). Virtual Organizations are in grid computing definition "a set of individuals and/or institutions having direct access to computers, software and data for collaborative problem solving"; resources are highly controlled and resource providers and consumers define clearly what is shared, who is allowed to share and the associated conditions.

GENESI-DR enforces the VO concept at two different levels:

•DR level by implementing means to strictly check access to resources.

•Central Site level by implementing means to subscribe/unsubscribe users to Virtual Organizations.

At the Central Site the Virtual Organization Management Systems (VOMS) assigns privilege attributes to users at VO level. Users subscribe to a specific VO and are assigned to specific groups and/or roles by the VO Administrator; in order to download data they must ask the VOMS server for their "enhanced" proxy. The VOMS server verifies user identity using GSI (Grid Security Infrastructure a specification for secret, tamper-proof communication between software used in Grid computing environments), then returns the user's attributes in the form of a signed "pseudo-certificate" similar to a SAML extension that is embedded in the proxy. This "pseudo-certificate" contains the VO defined attributes and/or roles of its subject that are specified using Fully Qualified Attribute Names (FQAN).

Access control is enforced only when users get data and/or use services from DRs while metadata access is free. Access control is implemented using an Apache HTTP server as a front end. Apache uses the OpenSSL library providing strong encryption using Secure Socket Layer and Transport Layer Security protocols. In GENESI-DR Apache is used in conjunction with GridSite that is a set of extensions for managing credentials, access control lists (GACL) and HTTPS protocol operations. GridSite intercepts some processing of the OpenSSL library to manage proxies (short lived credentials) and VOMS attributes certificates. In this way ACL (Access Control List) to directories, services and files can be based on VOMS extension attributes, basically Fully Qualified Attribute Names.

GENESI-DR Training

During the course of the project an extensive training package has been developed and delivered to potential users and data providers.

At the end of the first deployment cycle, in november 2008 (Frascati) a training workshop limited to project participants only was organised. At the ISPRA General Assembly in January 2009, this was extended to include some external participants. To allow quick growth of this GENESI-DR community, special care was taken to allow users to link their own data repositories to GENESI-DR in a way that is as straightforward and secure as possible. Training was focused on step-by-step guides describing : 1. how to provide other scientists with access to the data; 2. security configuration and access control set-up at the data repository; 3. data discovery and data access; 4. use by an 'expert user', showing how to run his own application. Two training events targeted at the broad Earth science community were conducted at major international conferences: OceanObs, Venice, September 2009 ; Fringe, Frascati, Nov/Dec 2009. These two events provided training to around 100 people.

The tutorials and demonstrations can be downloaded from the http://www.genesi-dr.eu Training menu.

Upcoming events where GENESI-DR will be present …

- ***European Geophysical Union (EGU), Vienna, Austria. 3rd – 7th May, 2010***
- ***INSPIRE conference, Krakow, Poland. 23rd– 25th June, 2010***

Web site:
www.genesi-dr.eu

Contact person:
Luigi Fusco

email: luigi.fusco@esa.int
tel.: +39 06 941 80 530
fax.: +39 06 941 80 532

Project participants:

ESA	FR
Elsag Datamat	IT
DLR	D
CNES	FR
KSAT	N
INFOTERRA	UK
ASI	IT
NILU	N
JRC	IT
Uni Reading	UK
ENEA	IT
TERRADUE	IT
CS SI	FR