

## EMSO-Grid

Pilot activity to study a Grid-based computing model for the EMSO ESFRI project

### Computing Model

It should be similar to the one adopted by the LHC community with a layered (tiered) structure. Tier0s will be created as close as possible to the experimental sites and will host raw data, Tier1s will host replicas of the data needed for the analysis and will run analysis jobs. In the future the need for a Tier2 layer will be evaluated, but we will start with just 2 layers, Tier0 and Tier1. Once the infrastructure will be set up at the European level the idea is to create a service which will act as a broker, it will receive user requests (for data and analysis) and will dispatch them to the appropriate grid resources.

For the pilot activity (EMSO/IGI) data will be generated by two offshore experimental sites and will reach harbors through fiber cables, one is located offshore Catania, the other offshore Porto Palo. The biggest part of data will come from hydrophones at the rate of 6Mb/s (\*), these data will be analyzed in real time by a first level of trigger, but EMSO would like to save all the RAW data. Data are firstly saved on a server located in the harbor and then transferred using rsync to LNS (INFN South Laboratories) in Catania where a grid site is already installed (INGV has C class IP addresses in LNS). The storage in the harbor is limited in space and organized as a round robin buffer.

(\*) About the data rate: 6Mb/s is not the full rate:

4 hydrophones with sampling rate of 96kHz are located in the CATANIA NORTH site ~ 3 Mb/s for each hydrophone (a bit more given to the overhead of the transmission protocol)

4 hydrophones with doubled sampling rate (192kHz) are located in the CATANIA SOUTH site with the rate doubled accordingly.

So 4 x 3 Mb/s from the north plus 4 x 6 Mb/s from the south ~36 Mb/s = 4,5 MB/s ~ 11 TB/month

Preprocessed data are also inserted into a dedicated DB (MOISTdb), we will investigate how to griddify this DB in a second phase. LNS (or INFN Catania) will be the first EMSO T0 (INFN-CATANIA was chosen). Data hosted on that site had to be accessed by the whole EMSO collaboration but in a first phase only by colleagues located in ROMA1, PAVIA, MILAN and LNS/CATANIA. The storage currently available to EMSO at LNS is 80TB that will host RAW data and should be griddified, making this space visible by a Grid SE. Currently only local account can access it. The griddification of these data is the first step towards the creation of the EMSO T0 in CATANIA (LNS) together with the deployment of ad hoc VOs LNS guarantees two logical volumes:

- One will contain raw data coming from the harbors (VL01) will be readonly
- One will contain data from users, replicas from analysis etc..(VL02)

How to move data from the harbor to LNS is under discussion, now is done through cron jobs and rsync, but more grid-oriented methods (glite clients and lcg-utils) should be analyzed in a second phase to have stored data atomically recorded into the grid catalogues. Also gridftp as an alternative was evaluated but the decision postponed. A requirement is that data at the harbor should not be accessible via grid and the access limited to a very small number of researchers. The access for the rest of the community will be done via Grid services once the data will be stored into the SE. The site for the first T1 is discussed probably will be located at INFN-CNAF using IGI resources but the decision is postponed. There is a brief discussion about another project that could have similarities in the computing model with EMSO. The project is KM3NET.

## Virtual Organisation:

Three possibilities are analyzed: 1) Using a general purpose Italian VO 2) Using the INGV VO 3) Creating a EMSO VO, something like EMSO.eu or EMSO.it We chose option 3) and probably EMSO.it will be created, but the name is still under discussion IGI is in favor of creating EMSO-eu.org which is also the name of the EMSO website and INGV has all the needed rights on that name.

## Digital Certificates:

We do not see any problem in obtaining Grid digital Certificates (from INFN CA) for the people involved in the activity. INGV has already three RA: ROME, BOLOGNA, MILAN/PAVIA.

## Grid Catalogues:

A Grid Catalogue is absolutely needed starting from now. INGV will decide the folders structure in the catalogue, IGI will provide the catalogue (probably LFC)

## Metadata:

A metadata catalogue is not urgently needed but it will be useful in the future (medium/long term) in particular to keep track of the characteristics and calibration curves of the instruments. IGI will provide it, currently investigating if AMGA is the right solution (need to know future maintenance plans for AMGA)

## Analisy Applications:

Possible application that will be needed on the grid to analyze data at the T1s. Many of them will be based on MATLAB and SCILAB. License problems for the first one are discussed, while porting the second one is much more easier. MATLAB seems anyway really needed. Porting already done by IGI of R and Octave are mentioned- they could be useful to EMSO. Scilab (v5.3.3) was already installed on the INFN-CATANIA site for testing.

## Mailing List:

an EMSO-Grid mailing list was created with the people involved from IGI, INFN, EMSO, INGV

## Mini Grid Tutorial:

A tutorial, about Grid topics of interest for EMSO (job submission and in particular data management), was held on the 4th of May and a seminar on introduction to Grid computing with EGI and IGI was given at INGV in Rome the day before). Agenda and material available here

## Presentation at conferences

Talk given at the EGI-Technical Forum 2012

-- DanieleCesini - 2012-10-09

---

This topic: UserSupport > EMSOGrid  
Topic revision: r2 - 2012-12-13 - EmidioGiorgio



Copyright © 2008-2023 by the contributing authors. All material on this collaboration platform is the property of the contributing authors.

Ideas, requests, problems regarding TWiki? Send feedback