

EGO External Computing Committee (ECC)

Draft report of the spring 2013 review (report to STAC)

7 May 2013

1 Introduction

The EGO External Computing Committee (see Annex) has made recommendations in its Preliminary report of the 2012 review of 18 April 2012. Since then it has followed the development of the Computing Model for Advanced Virgo (AdV) and has received a formal presentation and documentation at a videoconference meeting on 18 April 2013.

Substantial progress has been observed. The ECC continues to think, however, that many of the difficulties that prevent faster progress are of organizational and inter-institutional nature, and urges the highest levels of management of AdV and EGO to work together to solve them.

2 Remarks on the Computing Model for AdV

The ECC is pleased to see that substantial progress has been made towards an initial version of a Computing Model for AdV. In particular, document VIR-0129A-13 “The AdV Computing Model. Draft Version 0.1” already contains a base version of the documentation on Computing and Data Analysis Workflows with their corresponding datasets, which can be used for preliminary planning of the data resources needed at the interferometer, at CNAF and at CC-IN2P3. Much work remains to be done, however, to achieve a complete initial version of the AdV Computing Model which can start to be managed and used for complete resource estimates. In particular, a thorough description of the software involved (both for services and algorithms for analysis), including plans for development and management, is of extreme importance as it may have substantial manpower implications as well as timelines which are difficult to compress.

Recommendation 7: The schedule of work on the AdV Computing Model needs to be accelerated. In particular, the following should be completed by the fall 2013 committee meetings:

- The initial version of the complete AdV Computing Model, with all the components listed in section 4 of the ECC 2012 review.
- A short document detailing the management process for the AdV Computing Model, including the persons or bodies responsible for driving each part of the process, the chains of approval for considering each step completed, and the timing of each part of the process.
- A draft version of an implementation plan up to first AdV science operation, focusing on the methodological and technological choices and outlining the software development (both services and algorithms) which will be necessary.
- An early draft of an estimate of the budgets necessary to execute the implementation plan, including human resources and quantifying monetary as well as in-kind contributions.

The following sections contain remarks on specific aspects stemming from Draft 0.1 of the Computing Model which are sufficiently clear to make recommendations.

3 Remarks on the Cascina Disk and CPU farm

Recommendation 8: The importance of agile access to a variety of datasets by the experts on commissioning and characterization of the interferometer is being clarified, and good progress in documenting the details of these datasets is being made.

The method proposed for granting this access is to provide a disk buffer and associated CPU farm at Cascina. The size of the buffer is dominated by the requirement of keeping 0.5 years of Raw Data on site, with the corresponding impact on CPU. These data, however, will also be archived and accessible for analysis at CNAF and CC-IN2P3.

The ECC requests a thorough evaluation of an alternative solution where an agile mechanism for access and analysis of Raw Data at CNAF and CC-IN2P3 is made available in a transparent manner to the experts working on commissioning at the interferometer site. This would reduce very significantly the size of the disk buffer space and accompanying CPU in Cascina without necessarily increasing the resources needed at CNAF and CC-IN2P3.

The ECC does acknowledge that a modest amount of disk space and CPU is needed in Cascina, beyond what is strictly needed for the AdV Data Acquisition system, so that small datasets and, particularly, databases, can be quickly accessed for commissioning and characterization.

Recommendation 9: The ECC re-emphasizes the importance of its recommendations 1 and 2 and encourages the Disk and CPU farms in Cascina to be seamlessly integrated with the AdV Data Acquisition system, operationally as well as technically.

Responsibilities for funding, deployment, operation and technical support of the Cascina Disk and CPU farm should be clearly and urgently defined between the EGO Laboratory and the AdV Collaboration.

Recommendation 10: Various workflows involve databases at the interferometer site, and in many cases there are previous Virgo implementations based on specific database packages. In some cases, these databases need to be exported or replicated to other sites. One key database contains data from both Virgo and LIGO.

A process should be started for the evaluation of choices, risk assessment, testing and prototyping, in order to reach an agreement on a single underlying database platform which serves all workflows and fulfills the offsite export/replication requirements, as well as those of the Virgo/LIGO database.

This database platform should be implemented and maintained as a logical component of the Cascina Disk and CPU farm.

4 Remarks on the use of the CNAF and CC-IN2P3 centers

The yearly data volumes and the need for widespread data distribution fully justify the use of the large LHC-class data centers at Bologna and Lyon, in order to profit from their economies of scale. Multi-year planning is needed to ensure coverage of requirements for the whole AdV science run. This planning should be incorporated into the management of the AdV Computing Model.

It should be noted that neither CNAF nor CC-IN2P3 are part of the EGO Laboratory or the AdV Collaboration. Therefore, mechanisms should be found to regulate the relationship with these computing centers.

Recommendation 11: The management of the EGO Laboratory and of the AdV Collaboration should discuss and agree with the management of CNAF and CC-IN2P3 a framework agreement, either multi-lateral or multiple bi-lateral, to regulate the use of these computing centers for the data processing of AdV.

The Computing Model management plan should take into account these agreements.

Recommendation 12: Organizational and technical communication and cooperation between EGO, AdV, CNAF and CC-IN2P3 needs to be urgently streamlined.

The ECC recommends the immediate setup of an AdV Computing Coordination Committee (AdV-CCC) composed of one contact each from EGO, CNAF and CC-IN2P3 and chaired by the AdV Data Analysis Coordinator.

The existence of this coordination committee should be incorporated into the framework agreements of recommendation 11, but given the urgency it should be setup immediately even if the framework agreements are not finalized.

Recommendation 13: The needs for Remote Data Recording, at both CNAF and CC-IN2P3, of 0.9 PB per year of AdV science operation have been clearly established.

Early budget planning should incorporate the need for deployment of 0.3 PB of tape archive by the start of AdV commissioning and 0.9 PB per year of AdV science run at each center.

Efficient use of this tape archive will require a disk buffer, data access methods and access to the CPU farms at the centers. Requirements for these elements remain to be estimated.

Recommendation 14: A process for evaluation of choices, testing and prototyping, and reaching agreement on a solution for the Remote Data Recording should be started, ideally driven by the AdV-CCC.

In this context, the ECC re-emphasizes the importance of its recommendation 4.

The search for a solution should give maximum priority to choices which maximize the economies of scale of both CNAF and CC-IN2P3, presumably those which will be adopted for the second run of LHC.

In order to maximize reliability, ease operations at the interferometer, and minimize technical manpower, emphasis should be given to finding a single solution with a common implementation for sending AdV data from the EGO site to both CNAF and CC-IN2P3.

Recommendation 15: An adequate network connection of the EGO Laboratory to the European Academic Research Networks, presumably through the Italian NREN, should be ensured in order to implement Remote Data Recording and, possibly, Remote Data Access for commissioning and characterization (see Recommendation 8).

The EGO Laboratory, in coordination with the AdV Collaboration, should start dialogs with the relevant network providers and funding agencies in order to ensure that the technical planning and corresponding funding for the network is in place by the start of AdV commissioning and maintained throughout the AdV science run.

5 Additional remarks

Document VIR-0129A-13 “The AdV Computing Model. Draft Version 0.1” refers in several places to adoption of standards and packages of the European Grid Initiative (EGI). The ECC considers that this is a reasonable strategy, but warns that the future of EGI is unclear. The ECC recommends the following in parallel of the development of the Worldwide LHC Computing Grid (WLCG), in particular of its deployment at CNAF and CC-IN2P3.

An intense period of development of WLCG is taking place during 2013-2014, taking advantage of the pause in LHC data taking to introduce improvements and innovations. AdV should aim to profit from these developments.

The ECC has been informed that the global agreement between Virgo and LIGO is to be renewed in the timescale of one year. The ECC understands that joint operation and simultaneous data analysis of AdV and aLIGO is fundamental to their gravitational wave research. Nevertheless, the ECC emphasizes that a robust and efficient data processing environment for AdV is a necessary pre-requisite to produce the high-level datasets used in the joint interferometer analysis.

The ECC has received a note from the EGO Laboratory titled “Note on the planning of new installations and dismissal of the current systems”. The ECC considers that the information contained in this note shows that there is a good start on actions related to ECC recommendations 1, 2 and 3. In particular, the committee is pleased to learn that EGO and AdV have reached a number of agreements that will permit the common use through Ethernet of the renovated fiber network, including the fibers along the interferometer arms, as well as the elimination of most of the copper cables and related obsolete equipment.

The ECC will schedule a more in-depth review of the plans for the Cascina site now that additional information from the Computing Model is available.

6 Annex: The EGO External Computing Committee

In its 33rd meeting, the EGO Council formed an external committee to evaluate the current program and the future needs in order to give input to the STAC and the Council.

The External Computing Committee is composed by:

Manuel Delfino, Chair
Ed Porter
Frédéric Hemmer
Mauro Morandin
Dominique Boutigny

The mandate to the ECC is:

- To analyse and evaluate the present program and Computing Model for the Advanced Virgo project and the consequent requirements in terms of resources.
- To evaluate and provide guidelines about the tasks, the goals and the activities with the assignments to the involved entities, including the INFN-CNAF, CC-IN2P3, EGO computing centers, but also the participating laboratories and the LSC
- To evaluate the corresponding budget plan and the assignments of resources to all the entities, irrespective whether the resources are funded by EGO, IN2P3 or INFN

EGO and the Virgo Collaboration are available to provide any information to the ECC.