# USER FORUM Position Statement

Giovanni Lamanna
13.11.2013

## The view of an Astroparticle Physicist

APPEC is a consortium of national government agencies and institutes responsible for coordinating and funding national research efforts in Astroparticle Physics. Among its strategic objectives APPEC aims to provide a "discussion forum" for the coordination of European astroparticle physics. Such a forum is supported by a number of individual researchers representing the views of the community committed in all astroparticle projects. Periodical meetings allow those researchers and national agencies representatives to meet together and exchange about both users' expectations and governmental common plans about all aspects including the "e-Science" and "e-infrastructure". APPEC is used to support forum of astroparticle scientists/users. Three dedicated workshops on different aspects towards an exhaustive view of the "e-infrastructure" needs were organized in the past: https://indico.in2p3.fr/conferenceDisplay.py?confId=3845 (October 2010, Lyon), http://indico.cern.ch/conferenceDisplay.py?confId=134280 (May 2011, Barcelona), and http://indico.cern.ch/conferenceDisplay.py?ovw=True&confId=159120 (May 2012, Hanover) and recently completed by replies to a questionnaire distributed to all projects. (APPEC is going to finalize a comprehensive "white paper about Astroparticle and e-infrastructures" which will elaborate all information gathered from the concerned community[1]).

The global view which resulted is that the last decade witnessed the intensive construction of large astroparticle physics infrastructures. Towards the end of the decade the projects passed from the noise hunting regime to the generation of large sets of data, the production of which needs large computing resources, intensive simulation and large storage space. Furthermore, the data need to be cast in formats and be accompanied by software permitting wide accessibility as well as space and time correlation with information generated by other observatories of astrophysical or cosmological type.
The domain will experience in the coming decade exponential growth of data produced; participating in the overall forthcoming "data tsunami" and presenting major challenges in terms of storage, computation and long term preservation. A specificity of astroparticle physics is that the data produced by the distinct experiments are different and can be characterized as follows: -event-like structure; - signal-like structure; - image-like structure.

Some general preliminary and not exhaustive list of issues can be easily reported:
1. Data rates delivered by the experiments are quite diverse, ranging from 10-20 GBytes per day from satellite experiments, to 100 GBytes per day (after reduction) for the current ground cosmic ray experiments, to a few TBytes per day for advanced gravitational wave experiments and of the order of 15 TBytes per day for large surveys. The field of astroparticle physics therefore goes towards a data rate of tens of PBytes per year. The

[1] "Towards a computing model for astroparticle physics", I. Agrafioti, B. Allen, V. Beckmann, T. Berghöfer, M. Delfino, S.Hesping, S. Katsanevas, G. Lamanna, R. Lemrani (2013) *In preparation*

above will need at least GBps links from the remote sites to the data centres on ground and high bandwidth satellite links for the space experiments.
2. Different solutions for file distribution and storage are currently used in astroparticle physics experiments. To increase the efficiency in the use of resources and funding a standardization taking into account the developments of large data centers such as at CERN and ESO would be beneficiary for the astroparticle physics community. This would be an asset for interoperability of data and the scientific use of the data in e.g. a multimessenger analysis.
3. Simulations required for the data analysis can be more CPU-consuming than the capture and analysis of real data, in particular for the high energy cosmic ray experiments (charged particles, photons, neutrinos) as well as for gravitational wave antennas. Only recently the use of GPUs permits to consider a full and detailed simulation of the highest energy showers.
4. Concerning computing hardware, the gravitational wave antennas pioneered the use of GPU's for the analysis of the signal (with prospects to arrive to 80% of the computing power needed) and are followed by detailed simulations of cosmic ray showers (with prospect to arrive up to 50% of the computing power needed). The GPU use is a front that will expand in the coming years.
5. Several astroparticle projects have used the Grid for part of their simulations and analysis. The experience in using Cloud technology in astroparticle physics experiments computing is at the very beginning. Astroparticle computer specialists have plans to move in principle towards heterogeneous DCI (Distributed Computing Infrastructures).
6. Currently, relational databases like Oracle, MySQL, or PostgreSQL are used by the large majority of projects. Since the scaling of these database systems is limited future projects are considering new database technologies such as NoSQL. The astroparticle physics community should try to coordinate with nearby fields while keeping attention to its diversified needs.
7. The organization of an open access to the scientific data is demanded by future astroparticle Observatory projects and the user community is concerned on its implication on the computing models, the scientific analysis systems and the e-infrastructure to be adopted.
8. A detailed investigation of used data formats and possible standardization in view of an integration into the frame of the Virtual Observatory as well as the development of a data preservation strategy is of particular importance for the field of astroparticle physics.

Clearly in favor of any "user forum" approach aimed to gather scientists' expectations, my personal consideration is that astroparticle researchers would likely be concerned by: - a more efficient inter-communities shared development about data management issues; - a more in line with user expectations evolution of e-infrastructures and any e-Science cooperative action; - the organization of a shared sustainable model for large computing infrastructures deployment.