E-SCIENCE, DATA SCIENCE AND SCIENTIFIC COMPUTING P67

M.O. Menezes[a], R. Semmler, G.M. de Carvalho, E. Landulfo and M.S. Dias

[a] mo.menezes@gmail.com

Nuclear and Energy Research Institute, São Paulo, Brazil

The publication of papers in scientific journals or conference proceedings, has being the main way of summarization of experimental results obtained by the researchers over the time. However, the sharing of the experimental data in raw format or after some processing, is also equally important for the scientific community, as they provide the necessary input to reproducible experiments and also to independent validation of scientific results. Nowadays, the volume of scientific data production has increased to giant amounts, demanding new means of storage and curation as well as processes and technologies to make them available in durable ways. As a consequence, and at the same time a response, to those demands, a new scientific paradigm has emerged: the e-Science. This new paradigm distinguished itself from the traditional science, being characterized by intense computational activity, required to process the large volume of data that can be obtained from modern scientific experiments. e-Science, ultimately, is related to knowledge discovery and sharing not only as scientific publications, but also as experimental data, rich theoretic vocabularies, and several reusable services useful to the scientific community. The great availability of scientific data, both in raw or processed formats, leveraged by the adoption of transparency and accessibility politics by scientists all over the globe which publish their data on institutional or private repositories, are making possible also the reutilization of such data for new analysis by other scientists, who, employing new statistical approaches, such as machine learning algorithms suited to large amount of data, are also obtaining new results, not only from old data, but also, from the big amount of data originated from modern experimental facilities, doing what is known as "data science". The demand for intense computational utilization by e-Science related activities include not only the traditional simulation methods, but also the development of new tools that can operate in these new environments, such as, cloud based storage, cloud based access and analysis, mobile access to their research data, equipment monitoring and management, etc. All these activities are the scope of Scientific Computing being conducted at the Research Reactor Center - CRPq (IPEN-CNEN/SP).