

Short Abstract

The Global Scientific Data Infrastructures: The Big Data Challenges

Costantino Thanos¹

¹ Istituto di Scienza e Tecnologie dell'Informazione "A. Faedo", Consiglio Nazionale delle Ricerche (ISTI-CNR), via G. Moruzzi 1, 56100 Pisa, Italy, e-mail: costantino.thanos@isti.cnr.it

The current data tsunamis produced by the new advanced instruments and/or sensors and/or by running simulations and the progress of science is revolutionizing the way in which research is conducted and this poses new challenges to the existing e-infrastructures from both the data and the application sides.

Science is becoming data-dominated and a new data-centric way of thinking, organizing and carrying out research activities is gaining ground which needs to be supported by a new type of e-infrastructure: the Scientific Data Infrastructure.

Scientific Data Infrastructures can be defined as managed digital data-networked environments consisting of services and tools that support the full life cycle of data (capture, collection, curation, documentation, analysis, visualization, preservation, and publication) for the benefit of different communities of researchers involved in data-intensive activities.

A Scientific Data Infrastructure should, thus, add the capacity of effectively and efficiently handling and publishing the current huge volumes of data to the computational capacity provided by the e-infrastructures.

The next generation of scientific data infrastructures is facing two main challenges:

To effectively and efficiently support data-intensive Science

To effectively and efficiently support multidisciplinary/interdisciplinary Science

In order to develop such data infrastructures several data, application, system, and organization/ policy challenges must be successfully tackled.

The talk will address the main data challenges.