



Contribution ID : 20

Type : **not specified**

# BIG DATA in Science, Best Practices

*Friday, 30 August 2013 09:00 (40)*

Whilst Big Data is often characterised in terms of its volume in bytes: Tera, Peta, Zeta, there is also the crucial aspect regarding the degree of complexity within the data set to consider. Such complexity means that good data management is an essential element in the creation of high quality research data, without which researchers who collect the data will themselves be unable to realise the full scientific potential of the data set. Research data needs to be well organised, documented, preserved and accessible if their accuracy and validity is to be controlled.

The advent of Big Data by definition only makes these challenges harder, meaning that along with advances in data processing technology and applications, policies regarding data handling need to be evolved and adhered to. Increasingly funding agencies are now making demands for data management plans to be included within research grant applications. This becomes particularly prevalent where data sharing and open access to scientific data are made preconditions for access to research funding.

This presentation will discuss the main issues concerning best practice for handling large scientific data sets, as well as trying to look ahead to see how funding agencies are increasingly attempting to influence how scientists handle their data, by themselves defining such best practice.

## Summary

**Primary author(s)** : Dr. APLIN, Steve (DESY)

**Presenter(s)** : Dr. APLIN, Steve (DESY)

**Session Classification** : Plenary talks

**Track Classification** : Big Data and Large Storage Systems