

3: 13:40-13:50. ESGF, a Globally Distributed Software Infrastructure to Address the Challenges of Data Management, Access and Analysis for High Volume Climate Model Data

¹Philip Kershaw

¹Head Centre for Environmental Data Analysis

contact:

Abstract

The Earth System Grid Federation is an internationally federated archive providing access to Earth science data. Sustained through a collaboration effort between research institutions across the world it was first established over ten years ago to support the distribution of the outputs from CMIP5. In the last two years a major effort has been undertaken to reappraise the overall architecture and re-engineer the underlying software taking into account the challenges around rising data volumes, opportunities presented by the use of cloud computing and new paradigms which combine data access and analysis through the use of web-based platforms. This has resulted in the development of a more modular system suitable for deployment on-premise and using public cloud. A complete node has been in operation on Amazon Web Services operated by US partners GFDL. In the European context, hybrid models are being explored with on-premise hosting of large data volumes, together with services for sub-setting and data reduction to facilitate use in Climate4Impact a web platform which utilises Jupyter notebooks and a data cache hosted on cloud. It is expected that the new US funded ESGF2 project will utilise similar approaches and other new innovations to support user communities with both data access and in-situ data analysis capabilities.