

EGU2020-14241

<https://doi.org/10.5194/egusphere-egu2020-14241>

EGU General Assembly 2020

© Author(s) 2021. This work is distributed under the Creative Commons Attribution 4.0 License.



Data management and analysis of the high-resolution multi-model climate dataset from the PRIMAVERA project

Jon Seddon¹ and Ag Stephens²

¹Met Office, Exeter, UK (jon.seddon@metoffice.gov.uk)

²Centre for Environmental Data Analysis, STFC Rutherford Appleton Laboratory, Didcot, UK

The PRIMAVERA project aims to develop a new generation of advanced and well evaluated high-resolution global climate models. An integral component of PRIMAVERA is a new set of simulations at standard and high-resolution from seven different European climate models. The expected data volume is 1.6 petabytes, which is comparable to the total volume of data in CMIP5.

A comprehensive Data Management Plan (DMP) was developed to allow the distributed group of scientists to produce and analyse this volume of data during the project's limited time duration. The DMP uses the approach of taking the analysis to the data. The simulations were run on HPCs across Europe and the data was transferred to the JASMIN super-data-cluster at the Rutherford Appleton Laboratory. A Data Management Tool (DMT) was developed to catalogue the available data and allow users to search through it using an intuitive web-based interface. The DMT allows users to request that the data they require is restored from tape to disk. The users are then able to perform all their analyses at JASMIN. The DMT also controls the publication of the data to the Earth System Grid Federation, making it available to the global community.

Here we introduce JASMIN and the PRIMAVERA data management plan. We describe how the DMT allowed the project's scientists to analyse this multi-model dataset. We describe how the tools and techniques developed can help future projects.