



Legacy data center integration into distributed data federations: The World Data Center for Climate (WDCC) experience

Stephan Kindermann (1), Katharina Berger (2), and Frank Toussaint (1)

(1) German Climate Computing Centre (DKRZ), Hamburg, Germany (kindermann@dkrz.de), (2) German Climate Computing Centre (DKRZ), Hamburg, Germany (berger@dkrz.de)

The integration of well-established legacy data centers into newly developed data federation infrastructures is a key requirement to enhance climate data access based on widely agreed interfaces.

We present the approach taken to integrate the ICSU World Data Center for Climate (WDCC) located in Hamburg, Germany into the European ENES climate data Federation which is part of the international ESGF data federation. The ENES / ESGF data federation hosts petabytes of climate model data and provides scalable data search and access services across the worldwide distributed data centers. Parts of the data provided by the ENES / ESGF data federation is also long term archived and curated at the WDCC data archive, allowing e.g. for DOI based data citation.

An integration of the WDCC into the ENES / ESGF federation allows end users to search and access WDCC data using consistent interfaces worldwide. We will summarize the integration approach we have taken for WDCC legacy system and ESGF infrastructure integration.

On the technical side we describe the provisioning of ESGF consistent metadata and data interfaces as well as the security infrastructure adoption.

On the non-technical side we describe our experiences in integrating a long-term archival center with costly quality assurance procedures with an integrated distributed data federation putting emphasis on providing early and consistent data search and access services to scientists. The experiences were gained in the process of curating ESGF hosted CMIP5 data at the WDCC. Approximately one petabyte of CMIP5 data which was used for the IPCC climate report is being replicated and archived at the WDCC.