

EGU2020-4658

<https://doi.org/10.5194/egusphere-egu2020-4658>

EGU General Assembly 2020

© Author(s) 2022. This work is distributed under the Creative Commons Attribution 4.0 License.



Integrating e-infrastructures for remote climate data processing

Christian Pagé¹, Wim Som de Cerff², Maarten Pliieger³, Alessandro Spinuso⁴, Iraklis Klampanos⁵, Malcolm Atkinson⁶, and Vangelis Karkaletsis⁷

¹CECI, Université de Toulouse, CNRS, Cerfacs, Toulouse, France (christian.page@cerfacs.fr)

²KNMI, De Bilt, Netherlands (wim.som.de.cerff@knmi.nl)

³KNMI, De Bilt, Netherlands (maarten.plieger@knmi.nl)

⁴KNMI, De Bilt, Netherlands (alessandro.spinuso@knmi.nl)

⁵National Centre for Scientific Research "Demokritos", Agia Paraskevi, Greece (iaklampanos@iit.demokritos.gr)

⁶EPCC, The University of Edinburgh, Edinburgh, UK (Malcolm.Atkinson@ed.ac.uk)

⁷National Centre for Scientific Research "Demokritos", Agia Paraskevi, Greece (vangelis@iit.demokritos.gr)

Accessing and processing large climate data has nowadays become a particularly challenging task for end users, due to the rapidly increasing volumes being produced and made available. Access to climate data is crucial for sustaining research and performing climate change impact assessments. These activities have strong societal impact as climate change affects and requires that almost all economic and social sectors need adapting.

The whole climate data archive is expected to reach a volume of 30 PB in 2020 and up to 2000 PB in 2024 (estimated), evolving from 0.03 PB (30 TB) in 2007 and 2 PB in 2014. Data processing and analysis must now take place remotely for the users: users typically have to rely on heterogeneous infrastructures and services between the data and their physical location. Developers of Research Infrastructures have to provide services to those users, hence having to define standards and generic services to fulfil those requirements.

It will be shown how the DARE eScience Platform (<http://project-dare.eu>) will help developers to develop needed services more quickly and transparently for a large range of scientific researchers. The platform is designed for efficient and traceable development of complex experiments and domain-specific services. Most importantly, the DARE Platform integrates the following e-infrastructure services: the climate IS-ENES (<https://is.enes.org>) Research Infrastructure front-end climate4impact (C4I: <https://climate4impact.eu>), the EUDAT CDI (<https://www.eudat.eu/eudat-collaborative-data-infrastructure-cdi>) B2DROP Service, as well as the ESGF (<https://esgf.llnl.gov>). The DARE Platform itself can be deployed by research communities on local, public or commercial clouds, thanks to its containerized architecture.

More specifically, two distinct Use Cases for the climate science domain will be presented. The first will show how an open source software to compute climate indices and indicators ([iclim](https://github.com/cerfacs-globcc/iclim): <https://github.com/cerfacs-globcc/iclim>) is leveraged using the DARE Platform to enable users to build their own workflows. The second Use Case will demonstrate how more complex tools, such

as an extra-tropical and tropical cyclone tracking software (https://github.com/cerfacs-globc/cyclone_tracking), can be easily made available to end users by infrastructure and front-end software developers.