

Using the EGI FedCloud and the ESGF CWT-API in a WPS workflow to provide data analysis computations for the IS-ENES climate4impact platform

[Abstract]

Supporting data analytics in climate research with respect to data access is a challenge due to increasing data volumes. Several international and European initiatives have emerged and provide standalone solutions that offer potential for interoperability. In Europe, the IS-ENES (<https://is.enes.org>) consortium has developed a platform to ease access to climate data for the climate impact community (C4I: <https://climate4impact.eu>). It exposes data from ESGF data nodes as well as any OpenDAP server. It provides user interfaces, wizards and services for search and discovery, visualization, processing and downloading. This platform is very useful for many categories of users, and it will be presented. One of the important aspect of the C4I platform is that it enables users to perform data analysis calculations. In order to improve data analysis capabilities and performance, work has been done to enable C4I to delegate calculations onto other infrastructures. In Europe, an emerging e-infrastructure is being designed and built for several scientific domains, lead by EUDAT (<https://eudat.eu>) and EGI (<https://egi.eu>), which will form the basis of the future European Open Science Cloud (EOSC) to support scientific researchers. This e-infrastructure provides services within the EUDAT CDI (Collaborative Data Infrastructure). The ENES climate community is participating in the EUDAT CDI. The work presented here show how the EUDAT CDI will be used by the C4I platform. Within the EUDAT project, work has been done to integrate these existing e-infrastructures. The goal is to develop interoperable interfaces. 1. A first-level prototype that deploys the GEF Docker-backend onto the EGI FedCloud to perform computations and feeds the results into the EUDAT CDI has been completed. 2. The second-level prototype involves integrating the GEF backend and the ESGF CWT-API. The GEF backend pulls data from the ESGF infrastructure through the CWT-API so that data reduction is achieved through on-demand calculations. Further more complex calculations are then executed on the EGI FedCloud, and the results fed back into EUDAT B2 Services. This raises the authentication and authorization integration between ESGF and EUDAT/EGI. 3. The third-level prototype is the same as the second one, except that the GEF is executed by the C4I platform on-demand by the user, and the final results are fed back into the C4I user space for visualization, storage or download. A variant of this third-level prototype that could be implemented is to get the data directly from one or several ESGF Data Nodes, using the C4I Search WPS to locate the data files. It will be shown how those technologies, services and frameworks, will be integrated within the newly EU-funded H2020 project DARE and its BDE Platform, providing researchers and users with programmable methods and tools making use of data and infrastructures as transparently as possible.

[Bibtex]

```
@INPROCEEDINGS{2018EGUGA..2012280P,  
author = {{Pag{\`e}}, C. and {Pivan}, X. and {Rajapakse}, A. and {Som de Cerff}, W. and  
{Plieger}, M. and {de Vreede}, E. and {Spinuso}, A. and {B{\`a}rring}, L. and  
{Cofi{\`n}o}, A. and {d'Anca}, A. and {Fiore}, S.},  
title = "{Using the EGI FedCloud and the ESGF CWT-API in a WPS workflow to provide data analysis  
computations for the IS-ENES climate4impact platform}",  
booktitle = {EGU General Assembly Conference Abstracts},  
year = 2018,  
series = {EGU General Assembly Conference Abstracts},
```

volume = 20,
month = apr,
pages = {12280},
adsurl = {<http://adsabs.harvard.edu/abs/2018EGUGA..2012280P>},
adsnote = {Provided by the SAO/NASA Astrophysics Data System}
}